JYU DISSERTATIONS 131

Otto Loberg

Co-occurrence of Oculomotor Behaviour and Electrocortical Brain Activity during Naturalistic Reading and Word Recognition

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF EDUCATION AND
PSYCHOLOGY

JYU DISSERTATIONS 131

Otto Loberg

Co-occurrence of Oculomotor Behaviour and Electrocortical Brain Activity during Naturalistic Reading and Word Recognition

Esitetään Jyväskylän yliopiston kasvatustieteiden ja psykologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212
lokakuun 12. päivänä 2019 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Education and Psychology of the University of Jyväskylä,
in building Seminarium, auditorium S212, on October 12, 2019 at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2019

Editors
Noona Kiuru
Department of Psychology, University of Jyväskylä
Päivi Vuorio
Open Science Centre, University of Jyväskylä

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-7849-5

ISBN 978-951-39-7849-5 (PDF)
URN:ISBN:978-951-39-7849-5
ISSN 2489-9003

ABSTRACT

Loberg, Otto
Co-occurrence of oculomotor behaviour and electrocortical brain activity during naturalistic reading and word recognition.
Jyväskylä: University of Jyväskylä, 2019, 85 p.
(JYU Dissertations
ISSN 2489-9003; 131)
ISBN 978-951-39-7849-5 (PDF)

Reading is a process of converting symbols to meaning. Traditionally, eye movements and brain activity during reading have been investigated separately. This thesis examined the co-occurrence of brain activity and eye movement patterns during reading by utilising co-registered Eye Tracking & Electroencephalography (ET-EEG) methodology. The studies' focus was on how sublexical features of words—specifically, word length and identity of a single letter—affect the co-occurring eye movements and brain activity. The influence of reading proficiency was also examined. Study I aimed to disentangle the influences of two word length aspects, spatial width and the number of letters, on eye movements to find out whether dysfluent readers would be more sensitive to them. The number of letters had an impact on the temporal aspects of eye movements while spatial width was reflected in the spatial aspects of the eye movements. Overall, dysfluent readers exhibited longer duration measures than typical readers but no different effects. Study II examined the influence of word length and reading proficiency on Fixation Related Potentials (FRPs) during sentence reading. Word length had an influence on brain activity during additional fixations but not during the first fixations. In terms of eye movements, word length had an influence on first fixations but not on additional ones. Both typical and slow readers had different brain activity and eye movement behaviour. The word length effect was stronger in the eye movements of slow readers but the effect on brain activity did not differ between groups. Study III investigated how semantic anomaly detection influences FRPs and eye movements. It was found that, a deviation of a single letter from a plausible sentence was able to modulate eye movements and brain activity. When comparing one letter deviations from a plausible sentence to full word deviations, the brain activity for semantic processing was delayed, suggesting competition between parallel interpretations of the sentence meaning. Study IV examined the effects of inserting deviating letters to actual words on co-occurring brain activity and small saccades. The spatial location of the deviation attracted small saccades suggesting attempts to focus attention on important letters. The deviation type had a clearer influence on brain activity and reaction times, which suggested that if a deviation does not adhere to spelling rules, then no attempt for deciphering the semantics of the letter string is made. Overall, these studies show that sublexical features of words cause co-occurring brain activity and eye movement effects. These results help refine the understanding of the relationship between the eye and the mind.

Keywords: FRP, eye movements, reading, electroencephalography

Author's address
Otto Loberg
Department of Psychology
P.O. Box 35 FIN-40014 University of Jyväskylä
otto.h.loberg@jyu.fi

Supervisors
Professor Paavo H.T. Leppänen
Department of Psychology
University of Jyväskylä
Jyväskylä, Finland

Dr Jarkko Hautala
Niilo Mäki Institute
Jyväskylä, Finland

Professor Jarmo A. Hämäläinen
Department of Psychology
University of Jyväskylä
Jyväskylä, Finland

Reviewers
Dr Olaf Hauk
MRC Cognition and Brain Sciences Unit
University of Cambridge

Professor Adrian Staub
Department of Psychological and Brain Sciences
University of Massachusets Amherts

Opponent
Dr Olaf Hauk
MRC Cognition and Brain Sciences Unit
University of Cambridge

TIIVISTELMÄ (FINNISH ABSTRACT)

Loberg, Otto
Okulomotorisen käyttäytymisen ja elektrokortikaalisen aivoaktiivisuuden yhteisesiintyminen luonnollisen lukemisen ja sanantunnistamisen aikana.
Jyväskylä: University of Jyväskylä, 2019, 85 p.
(JYU Dissertations
ISSN 2489-9003; 131)
ISBN 978-951-39-7849-5 (PDF)

Lukeminen on kognitiivinen prosessi, jossa visuaaliset symbolit muunnetaan vastaaviksi merkityksiksi. Perinteisesti lukemisen aikaisia silmänliikkeitä ja aivoaktiivisuutta on tarkasteltu erillisinä ilmiöinä. Tämä väitöskirja tarkastelee niiden yhteisesiintymistä luonnollisen lukemisen ja sanantunnistuksen aikana käyttäen katseenseurannan ja aivosähkökäyrän rinnakkaisrekisteröintiä. Tutkimuksissa keskityttiin siihen, miten sanojen subleksikaaliset ominaisuudet, erityisesti sanan pituus ja yksittäisen kirjaimen identiteetti, vaikuttavat silmänliikekäyttäytymiseen ja aivoaktiivisuuteen ja näiden yhteisesiintymiseen. Tarkastelun kohteena oli myös lukutaidon tason vaikutus näihin ilmiöihin. Ensimmäinen osatutkimus pyrki erottelemaan sanan leveyden ja kirjainmäärän vaikutuksia silmänliikkeisiin lukemisen aikana. Näiden suhdetta heikkoon lukutaitoon tutkittiin myös. Sanan kirjainmäärä oli ensisijaisesti yhteydessä silmänliikkeiden ajallisiin piirteisiin, ja sanan leveys oli yhteydessä silmänliikkeiden spatiaalisiin piirteisiin. Hitaiden lukijoiden silmänliikkeiden kestomuuttujien arvot olivat tyypillisien lukijoiden vastaavia pidempiä, mutta tämä ei ollut yhteydessä sanan kirjainmäärän tai sanan leveyden vaikutukseen. Toinen osatutkimus tarkasteli sanan pituuden ja lukutaidon vaikutusta fiksaatiovasteeseen lauseen lukemisessa. Sanan pituus vaikutti aivotoimintaan ensimmäistä fiksaatiota seuraavien fiksaatioiden aikana, ei ensimmäisen fiksaation aikana. Silmänliikkeissä sanan pituudella oli vaikutus ensimmäiseen, mutta ei sitä seuraaviin fiksaatioihin. Hitaiden ja tyypillisten lukijoiden välillä oli eroja aivoaktiivisuudessa ja silmänliikkeissä. Sanan pituuden vaikutus silmänliikkeisiin oli suurempi hitailla lukijoilla, mutta sen vaikutuksessa aivotoimintaan ei havaittu eroja ryhmien välillä. Kolmas osatutkimus tarkasteli sitä, miten semanttisten anomalioiden havaitseminen vaikuttaa fiksaatiovasteisiin ja silmänliikkeisiin. Jo yhden kirjaimen poikkeama uskottavasta lauseesta muokkasi silmänliikkeitä ja aivotoimintaa. Verrattaessa koko sanan poikkeamiin, yhden kirjaimen poikkeamaan liittyvä aivoaktiivisuuden ero oli viivästynyt, mikä viittaa rinnakkaisten merkitystulkintojen keskinäiseen kilpailuun. Neljäs osatutkimus tarkasteli sanoista muodostettujen epä- ja pseudosanojen prosessoinnin aivotoiminnan ja mikrosilmänliikkeiden yhteisesiintymistä. Epä- ja pseudosanat muodostettiin korvaamalla yksi sanan kirjain siten, että kirjainjono oli merkityksetön. Pienet sakkadit suunnattiin kohti poikkeaman sijaintia eli tarkkaavuus kohdentui informatiivisiin kirjaimiin sanantunnistuksen aikana. Poikkeaman tyyppi vaikutti aivoaktiivisuuteen ja reaktioaikoihin. Tulokset osoittavat, että mikäli kirjainjono rikkoo kielensä kirjoitusasua, aivot eivät pyri tulkitsemaan kirjainjonon merkitystä. Kokonaisuutena osatutkimukset osoittavat, että sanojen subleksikaaliset ominaisuudet ovat riittäviä aiheuttamaan yhteisesiintyviä vaikutuksia aivotoimintaan ja katselukäyttäytymiseen. Tuloksia voidaan käyttää edistämään teoreettista ymmärrystä mielen ja silmän vuorovaikutuksesta.

Avainsanat: fiksaatiovaste, lukeminen, silmänliikkeet, elektroenkefalografia

ACKNOWLEDGEMENTS

I would like to offer my gratitude to my supervisors: Professor Paavo Leppänen, Dr Jarkko Hautala and Professor Jarmo Hämäläinen. Your support and insights in navigating the convoluted world of science have been invaluable. I would also like to express my appreciation for the contributions made by the following reviewers of this thesis: Dr Olaf Hauk and Professor Adrian Staub. Your well-justified criticisms and astute observations allowed me to improve the contents of this thesis.

Furthermore, I would like to thank all of my colleagues, as well as the technical support staff, for all the vibrant and insightful discussions, especially the ones occurring in the late hours of the Neurodrinks. I am grateful to my senior colleagues Dr Piia Astikainen, Dr Ina Tarkka, Dr Kaisa Tiippana and Professor Simon Liversedge for opportunities to work on and contribute to research outside this thesis. The insights acquired from these external projects have had a considerable and positive impact on the work presented here. A special mention goes to Dr Federica Degno for long discussions on the minutiae of the co-registered ET-EEG methodology. These discussions have proven to be an indispensable source of clarity on the methodology and awareness regarding rationalising why certain things need to be done the way they are done. Often, these discussions have acted as a “sanity check” for the procedures used and, as such, have led to improved solutions. In this, I recognise immense value.

As I have become older and, hopefully, a bit wiser, I have come to understand that a person, if he indeed intends to remain a person, cannot be sustained by academic activities alone. Reflecting on this, I must give my thanks and express my sincere appreciation to my wife, Maria, who has walked by my side throughout this journey—witnessing my darkest moods and comforting me in my moments of despair. Deep gratitude must also be given to my sister Johanna, godmother Leena and grandmother Rauha. In the precarious situation that I found myself in during early childhood and that persisted into my early adulthood, you three provided the safety net that helped me avoid worst outcomes and allowed me to set up the foundations from which I could push forward. For this, I am eternally grateful to you. My father Hannu and brother Juha have been invaluable sources of wisdom in helping me understand the often confusing facets of life and, for that, I am in your debt.

One must also not forget the company one keeps. Johannes, Kalle, Mikko, Jussi, Tero and numerous others have been a source of joy, camaraderie and occasional mischief. May there be many more decades of laughter and song to come.

This thesis and the studies contained within it have been financially supported by the Eino Jutikkala Foundation, the Academy of Finland (Grant: 274022) and the Department of Psychology at the University of Jyväskylä.

Jyväskylä 2.9.2019
Otto Loberg

LIST OF ORIGINAL PUBLICATIONS

The dissertation is based on the following original publications:

I. Hautala, J., & Loberg, O. (2015). Breaking down the word length effect on readers' eye movements. *Language, Cognition and Neuroscience, 30*(8), 993–1007.
II. Loberg, O., Hautala, J., Hämäläinen J.A., & Leppänen P. H. T. (in press). Influence of reading skill and word length on fixation-related brain activity in school-aged children during natural reading. *Vision Research.*
III. 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* .
IV. Loberg, O., Hautala, J., Hämäläinen J.A., & Leppänen P. H. T. (2019). Influence of spatial location of and type of deviation from real words on small saccades and brain activity—A co-registered eye-tracking and EEG study. Submitted manuscript.

The author contributed to the original publications listed above as follows.

For Study I, the author contributed to designing and implementing the experiment, gathered the data, contributed to the analysis, writing and review process of the publication.

For Studies II and III, the author, taking note of the instructions and suggestions of the co-authors, designed the experiment and the stimuli, implemented the experiment, supervised research assistants in data gathering, analysed the data and wrote the manuscripts.

For Study IV, the author, taking note of the instructions and comments of the co-authors, designed the experiment, gathered the data, analysed the data and wrote the manuscript.

FIGURES

TABLES

CONTENTS

1 INTRODUCTION PART I: BRIEF OVERVIEW ON READING AND RELATED PHENOMENA

In everyday experience, we perceive our visual world to be a unified percept for which we are constantly receiving visual information. However, this is a delusion. Reality is that we are blind for tens of milliseconds several times each second. The phenomenon known as saccadic suppression interrupts the flow of visual information from the eyes to the confines of the occipital lobe every time we make even the smallest (ballistic) eye movement (Bridgeman, Hendry, & Stark, 1975; Thiele, Henning, Kubischik, & Hoffmann, 2002). Thus, for example, as you are reading this thesis, you are moving your eyes in an alternating sequence of rapid jerks and relatively slow stops (Holmqvist, Nyström, & Andersson, 2011; Rayner, 1997). Overall, you are making a stop for most of the words you read—some you skip over while some you might come back to because the text is complex and you need to re-evaluate what it is that is actually suggested by it (Rayner, 1997).

During each jerk, *a saccade*, you are blind. At the start of each stop, *a fixation*, the flow of information from the eye to the cortex begins anew.

Saccadic suppression allows fixation onset to be treated as a starting point for visual processing. With a simple combination of an eye-tracker (ET) and an electroencephalograph (EEG) amplifier, it is possible to examine the brain activity related to fixations (Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011). This is important because most of the research on brain activity during reading is done using methods that limit ecological validity. In the traditional methodology framework for brain activity studies, you would be reading this thesis one single word at a time, presented at the centre of the screen for a second or so and you would not be allowed to go back in the text. This is an important distinction from natural reading because parafoveal processes and volition are instrumental for natural reading, thus traditional experimentation in brain studies places fundamental constraints on the generalisability of the findings in relation to natural reading.

The overarching goal of this thesis is to examine eye movements and brain activity as they co-occur during the natural reading and word recognition process. This is achieved by observing modulations of eye movements and brain activity to sublexical features of words. More specifically, the impact of word length and the impact of switching one letter are examined. The secondary overarching goal is to examine how reading ability affects eye movements and brain activity during natural reading. This thesis comprises an overview and four original studies. The term natural reading is used through the thesis when it is important to denote that what is read is read in the manner in which reading would occur in non-laboratory settings and is in contrast to other reading research procedures in which behaviour is constrained by methodological requirements.

1.1 Reading

Even though it is one of the core skills for functioning in modern society, reading is a curious cultural phenomenon. It is a process of converting fairly arbitrarily chosen visual symbols into assigned meanings. In its purest essence, reading is a highly specialised form of object recognition. Neuro-cognitive structures that evolved for visual information processing are re-purposed through education to perform the specialised task of word recognition (for streamlined discussion see Dehaene, 2009). However, this re-purposing of the neuro-cognitive functioning requires significant effort. Phonemic awareness and learning grapheme-phoneme correspondence are at the core of learning to read (Ziegler & Goswami, 2005). Debate on whether phonemic awareness precedes learning grapheme-phoneme correspondence is ongoing (Castles & Coltheart, 2004; Castles, Rastle, & Nation, 2018). However, being aware of separate speech sounds and connecting these sounds to letters enables an emergent reader to engage in self-teaching, as voicing out the letters of a word in a sequence establishes a feedback loop—if letters are converted into correct phonemes and uttered in a correct sequence, then they form a word and hearing that word acts as feedback (Share, 1995). Whether this feedback loop can be feasibly formed is influenced by many factors. One major obstacle to or facilitator of, depending on point of view, this process is the language in which an emergent reader is learning to read (Seymour, Aro, & Erskine, 2003). In languages with opaque orthographies like English, where the letter-phoneme correspondence is ambiguous and finding the correct phoneme depends on the surrounding letters, learning phonological decoding is cumbersome and essentially requires parallel processing of multiple letters to be successful (Ziegler & Goswami, 2005). Conversely, in transparent orthographies like Finnish, where the letter-phoneme correspondence is in practical terms unambiguous,[1] the learning of letter-phoneme correspondence is feasible and emergent readers can usually proceed to the self-teaching stage relatively fast. Through practice, an emergent reader manages to process a larger amount of letters parallel or even all letters of a word parallel.

---

[1] The sole exemption in Finnish is the phoneme /ng/.

However, for some individuals, in the absence of more severe neurological ailments or particularly low IQ, learning to read is an almost impossible achievement. This neurodevelopmental disorder is called dyslexia (Lyon, Shaywitz, & Shaywitz, 2003; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Impaired phonological processing is proposed to be present in dyslexia (Guttorm et al., 2005; Lyytinen et al., 2004), as speech sound discrimination in early development is shown to predict the incidence of reading difficulty in later development. Multiple findings suggest that genetic background is involved (Démonet, Taylor, & Chaix, 2004; Galaburda, 2005). Features of the writing system in a given language have an influence on the phenotype of the symptoms of reading difficulty. In languages with transparent orthography, reading difficulty manifests as an overall slowness when reading aloud, whereas, in languages with opaque orthography, reading difficulty manifests as numerous mistakes during reading aloud (Seymour et al., 2003). Poor discrimination of speech sounds hinders learning of the letter–phoneme correspondence and this, in turn, leads to a variety of behavioural symptoms (Ziegler & Goswami, 2005). An alternative explanation suggests that problematic cognition is the attentional capacity to process multiple letters simultaneously (Bosse, Tainturier, & Valdois, 2007; Valdois, Bosse, & Tainturier, 2004).

To summarise, reading can be considered to be a highly specialised form of object recognition, a skill that is attained through learning to connect phonemes to letters. One of the most efficient techniques to study reading is recording eye movements with an eye-tracker. Hence, the rationale behind tracking eye movements, eye movement patterns during reading and how symptoms of poor reading skill manifest through eye movements are explored next.

1.2 Eye movements during natural reading

At the very core of the rationale behind examining how eyes move during any task lies the concept of the eye–mind link. Simply put, the eye–mind link is a theoretical assumption that there is coupling occurring between eye movement behaviour and ongoing cognitive processes. To simplify this even further, the assumption is that what is looked at is what the mind is processing. Previous research has shown that neither absolute decoupling of the eye and mind nor one-to-one coupling between the eye and mind survive empirical testing. However, the ongoing debate in eye movement research of reading relates to the degree of decoupling between what word is looked and what word is processed. In computational eye movement models, the simplified general pattern is that most of the processing effort is directed towards the word that is fixated. In case of the EZ-reader model, which is a model of eye movement control during reading that assumes serial shifts of attention (Reichle, Rayner, & Pollatsek, 2018; Reichle, Tokowicz, Liu, & Perfetti, 2011), if the processing is successful enough, then the attention is allowed to move to the next word—which is not directly fixated. In another eye movement control during reading model—SWIFT (Engbert, Nuthmann, Richter, & Kliegl, 2005) — the processing is assumed to be distributed parallelly over several words simultaneously and yet it is thought to be most intense for the word that is directly fixated.

When adults without any developmental difficulties read normal text, a significant part of their eye movement variation can be attributed to three word properties: word frequency, word length and predictability (Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, 1997). Fixation durations and gaze durations are longer for low-frequency words than high-frequency words (Kliegl et al., 2004). Conversely, long words receive longer fixation durations and gaze durations than short words do (Hautala, Hyönä, & Aro, 2011; Kliegl et al., 2004). Long words are also skipped less often. Words that are predictable receive shorter fixations and gaze durations (Kliegl et al., 2004; Rayner, 1997). In addition, predictable words are skipped more often than unpredictable words (Kliegl et al., 2004).

Learning to read is clearly reflected in how eye movement patterns change during development. The eye movements of beginning readers are characterised by long fixations (Blythe, Häikiö, Bertam, Liversedge, & Hyönä, 2011; Blythe & Joseph, 2012; Tiffin-Richards & Schroeder, 2015) and short saccades (Blythe & Joseph, 2012). As saccades are shorter, more fixations are allocated to each word. Through practice, as reading becomes more fluent, fixations become shorter (Blythe et al., 2011; Blythe & Joseph, 2012; Tiffin-Richards & Schroeder, 2015) and saccades become longer (Blythe & Joseph, 2012), which reflects cognitive processing becoming more efficient. Low-level factors, such as word length, play a significant role earlier on in the development process but this influence diminishes as reading skill improves (Tiffin-Richards & Schroeder, 2015). Simply put, the eye movement pattern transitions from slow and piecewise to fast, accurate and largely automatic.

The eye movements of individuals with reading difficulty do not manifest this transfer. Overall, the eye movement pattern of dyslexics during reading is reminiscent of a pattern that is present earlier in an earlier developmental stage. Saccades are shorter and there is a higher amount of fixations per each word (De Luca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002; De Luca, Di Pace, Judica, Spinelli, & Zoccolotti, 1999; Dürrwächter, Sokolov, Reinhard, Klosinski, & Trauzettel-Klosinski, 2010; Hawelka, Gagl, & Wimmer, 2010). In addition to overall difference, word length effect is stronger for dyslexics than typical readers (De Luca et al., 1999, 2002; Dürrwächter et al., 2010; Hawelka et al., 2010).

In summary, the measurement of eye movements during reading is motivated by the eye–mind link assumption. This means that the modulations to eye movement patterns by text features are thought to be associated with the modulations in cognitive processes. During reading acquisition, the eye movement pattern progresses from a piecewise slow and cumbersome pattern to a fluid, swift and accurate pattern. Individuals with dyslexia do not exhibit this transfer to a similar degree but, instead, they manifest eye movement patterns that their typical reading peers exhibited earlier on in their development.

1.3 Miniature eye movements

In addition to fixation / saccade eye movement patterns exhibited during natural reading, the eye movements carry informative signals on a more miniature scale (Martinez-Conde, Macknik, & Hubel, 2004). These miniature eye movements include microsaccades, tremor and drift. Drift is a low frequency and low amplitude movement of the gaze position and tremor is a high frequency and low amplitude movement of the gaze position. Whether drift and tremor have an association with cognitive functions is unclear and they are mostly considered to serve the basic biological aspects of vision.

Microsaccades, on the other hand, are associated with spatial attention functions (Engbert & Kliegl, 2003; Laubrock, Engbert, & Kliegl, 2005; Meyberg, Werkle-Bergner, Sommer, & Dimigen, 2015; Rolfs, Kliegl, Laubrock, & Engbert, 2015) and sensitivity to auditory discrimination (Valsecchi & Turatto, 2009; Widmann, Engbert, & Schroger, 2014) and are reactive to letter string characteristics (Hautala & Parviainen, 2014; Yablonski, Polat, Bonneh, & Ben-Shachar, 2017). During visual tasks, microsaccades are suggested to optimise the foveal location in a way that supports the task at hand (Kagan & Hafed, 2013). However, there is an ongoing debate about whether microsaccades and voluntary small saccades should be distinguished from one another (Sinn & Engbert, 2016).

In summary, the eye exhibits informative behaviour during the period that, in the context of natural reading, is understood as the period in which the eye is still. A couple of studies have examined microsaccades during single word recognition (Hautala & Parviainen, 2014; Yablonski et al., 2017) and no studies have been conducted on the co-occurrence of microsaccades and brain activity during single word recognition.

1.4 Neuroanatomical and electrophysiological correlates of reading

It is established above that reading, reading skill and encountered text features have a profound impact on the pattern of eye movements. Unsurprisingly, a complex network of brain areas is involved in the generation of saccades (Martinez-Conde, Otero-Millan, & Macknik, 2013). Several cortical areas are implicated in the voluntary control of voluntary gaze control and these are the bilateral frontal eye fields, the supplementary eye fields, the lateral intraparietal area and the dorsolateral prefrontal cortex. Two components of basal ganglia are involved: the caudate nucleus and the substantia nigra pars reticulata and they are thought to perform inhibitory functions to avoid unwanted eye movements. Reticular formation and superior colliculi are the brainstem regions involved in producing the motor command needed to move the eyes. The fastigial oculomotor region and the oculomotor vermis are structures of the cerebellum that provide a feedback loop to control saccade accuracy (Martinez-Conde et al., 2013).

Reading, in the absence of eye movements, utilises large parts of visual processing hierarchy and areas related to semantic processes, which are essentially in the left hemisphere (Price, 2012). Visual information is processed in the occipital and ventral occipitotemporal cortex. Words, as a specific form of visual information, are found to specifically activate the left inferior temporal cortex, also known as the left fusiform gyrus or the visual word form area (McCandliss, Cohen, & Dehaene, 2003). Semantic properties of words are associated with activity in the middle regions of the left temporal lobe and the inferior frontal gyrus (Jobard, Crivello, & Tzourio-Mazoyer, 2003). In addition, activity in anterior temporal lobes are associated with semantic processing (Price, 2012). Propagation of brain activity during reading begins with initial visual processing in the occipital cortex, at around 100 ms, while at around 170 ms the activity is in the posterior part of the occipitotemporal cortex, spreading to the middle temporal regions and inferior frontal areas at around 250 ms (Marinkovic et al., 2003). The summary of the cortical areas that are plausibly connected to brain activity during natural reading is presented in Figure 1.

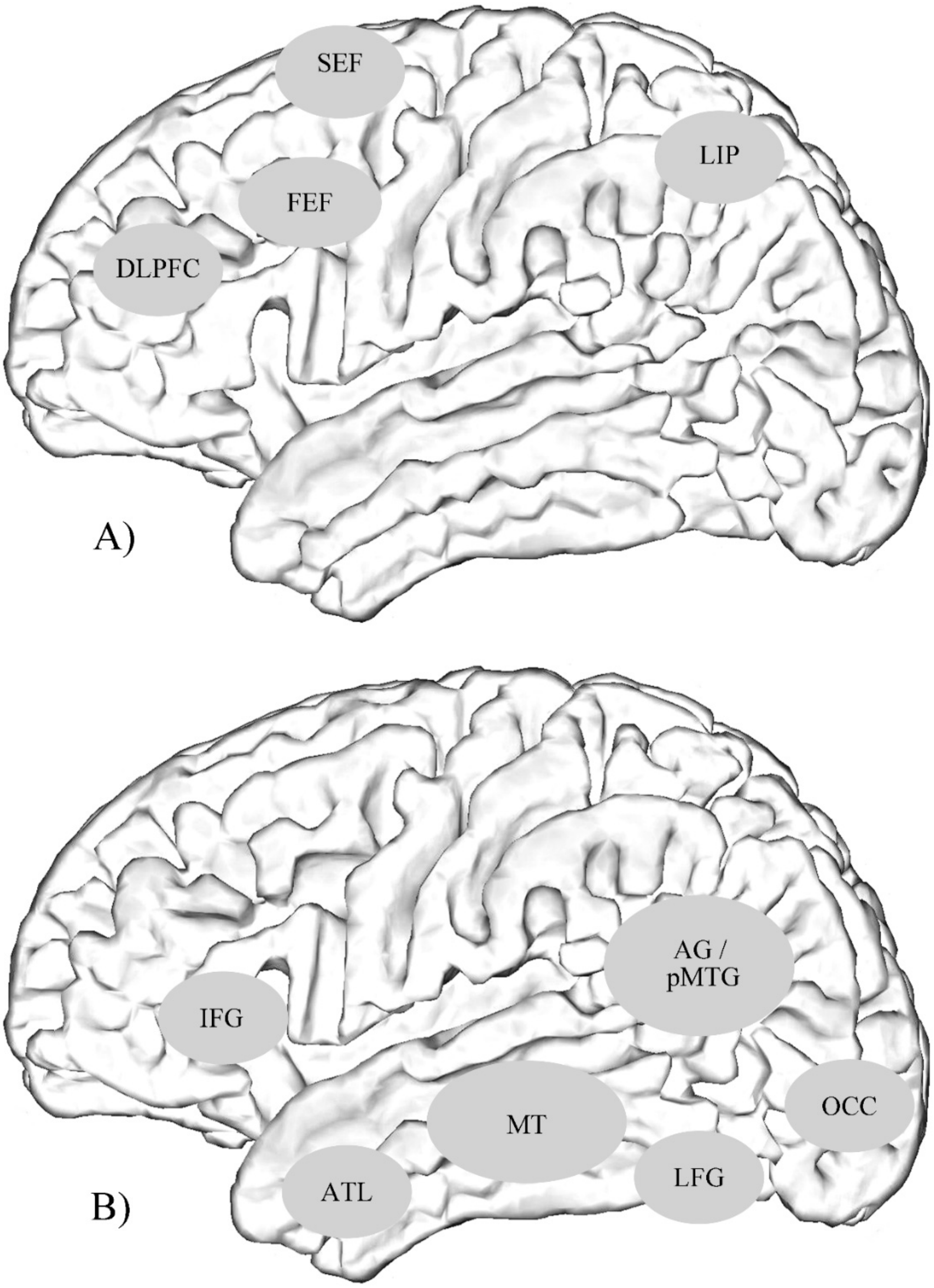


FIGURE 1 **Rough schematic** of the cortical areas involved in natural reading. A) Cortical areas involved in voluntary gaze shifts. LIP = Lateral intraparietal area. FEF = Frontal eye field. SEF = Supplementary eye field. DLPFC = Dorsolateral prefrontal cortex. B) Cortical areas involved in visual word recognition and processing of semantic content during reading. OCC = Occipital cortex. LFG = Left fusiform gyrus. MT= Middle temporal cortex. AG = Angular gyrus. pMTG = Posterior middle temporal gyrus. ATL= Anterior temporal lobe. IFG = Inferior frontal gyrus.

In the event-related potential (ERP) literature, the visually evoked occipital N1, especially the left-lateralised variant, has been associated with word recognition processes (Sereno & Rayner, 2003). Word length, the number of word neighbours and word frequency are found to influence the ERP during the early time window, preceding 150 ms, while words are shown to be differentiated from pseudowords by as early as 160 ms (Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006). Thus, word features that influence eye movements, word length and frequency are found to generate differences in ERP within the constraints set by the fixation duration. Semantic processing is a slightly different matter because related ERP signatures often span outside the fixation duration limits. The ERP responses P600 and N400 are associated with semantic processing (Friederici & Weissenborn, 2007). Even though P600 was originally thought to reflect syntactic processing (Osterhout & Holcomb, 1992), in the end, these two parietally maximal responses are shown to be evoked by fairly similar manipulations (DeLong, Quante, & Kutas, 2014; Friederici & Weissenborn, 2007). The critical distinction between the two is that N400 has stronger amplitude when attempts to fit an encountered word into the preceding context is difficult. P600, on the other hand, is found using manipulations that render the entire sentence incomprehensible, which, in part, explains why the P600 modulations are also reliably found with syntax violations. Also, the suggestion that P600 would reflect the conscious processing of semantic anomalies is raised (Sanford, Leuthold, Bohan, & Sanford, 2011).

Electrophysiological brain activity for letters and words changes as reading skill is acquired. N1 to visual words is larger for children in the second grade than children in kindergarten (Maurer, Brem, Bucher, & Brandeis, 2005; Maurer et al., 2006, 2007). In addition, when kindergarten children receive grapheme-phoneme correspondence training, their N1 response to visual words becomes stronger (Brem et al., 2010). Similar to acquiring reading skill, the difficulties in doing so are reflected in occipital responses. Magnetoencephalography (MEG) studies, comparing adult dyslexics to typical readers, have indicated that the left inferior occipitotemporal cortex shows different activity at around 150–200 ms from stimulus onset (Helenius, Tarkiainen, Cornelissen, Hansen, & Salmelin, 1999; Salmelin, Service, Kiesila, & Uutela, 1996). Furthermore, a clear association between word-related visual N1 and reading difficulty is established (Fraga González et al., 2014; Hasko, Groth, Bruder, Bartling, & Schulte-Körne, 2013; Maurer et al., 2007) but the pattern of effects is not uniform. Two studies found N1 to be weaker in dyslexics than in typical readers (Hasko et al., 2013; Maurer et al., 2007), while another study found N1 to be stronger in dyslexics than typical readers (Fraga González et al., 2014). However, while successful remedial intervention removed this N1 difference between dyslexics and typical readers, the unsuccessful one did not (Fraga González et al., 2016).

In summary, reading involves multiple cortical structures in the brain that are also related to oculomotor control, visual processing and semantic processing. From the point of view of the brain, learning to read is essentially re-purposing the existing structures to make connections between symbols and sounds, as well as symbol arrays and semantic meanings. Learning to read changes the functioning of the brain upon exposure to written words and people with reading difficulties do not exhibit these changes to the same extent. However, research on brain activity during reading that uses traditional methodological implementations is limited in its generalisability for the reading process in natural settings. Next, the core findings from limited but growing body of literature about co-registered EEG and eye-tracking are briefly summarised.

1.5 Brain activity during natural reading

Recently, studies examining brain activity during reading have implemented research techniques that allow for monitoring of electrical brain activity during reading as it occurs in natural circumstances. In co-registered eye-tracking electroencephalography, this is achieved by pairing an eye tracker with an EEG amplifier, which allows for monitoring eye movement patterns concurrently with brain activity patterns (Baccino & Manunta, 2005; Degno et al., 2019; Dimigen et al., 2011; Henderson, Luke, Schmidt, & Richards, 2013; Hutzler et al., 2007; Kornrumpf, Niefind, Sommer, & Dimigen, 2016; Kretzschmar, Bornkessel-Schlesewsky, & Schlesewsky, 2009; Kretzschmar, Schlesewsky, & Staub, 2015; López-Peréz, Dampuré, Hernández-Cabrera, & Barber, 2016; Metzner, von der Malsburg, Vasishth, & Rösler, 2016; Niefind & Dimigen, 2016; Simola, Holmqvist, & Lindgren, 2009).

Overall, electrophysiological signature time-locked to the fixation onset is highly similar to the one evoked at the stimulus onset (Dimigen et al., 2011; Kornrumpf et al., 2016; Metzner, von der Malsburg, Vasishth, & Rösler, 2015; Niefind & Dimigen, 2016). Studies utilising the fixation-related potential (FRP) method have managed to establish that those component modulations that are observed in traditional ERP studies can also be observed in FRP studies as well (Dimigen et al., 2011). The overall pattern shows that FRP components are similarly responsive to the same modulations as ERP components. For example, N400 is similarly responsive to semantic manipulations as in traditional ERP designs (Dimigen et al., 2011). New associations have also been uncovered. For example, P600 was recently found to co-occur with regressive saccades (Metzner et al., 2016). This is a reasonable pattern, as both have been associated with semantic processing. In addition, microsaccades generate FRP equally as longer saccades (Dimigen, Valsecchi, Sommer, & Kliegl, 2009; Meyberg et al., 2015) and parafoveal masking increases N1 amplitudes when a previously masked word is fixated (Degno et al., 2019; Niefind & Dimigen, 2016). Overall, these findings illustrate that many parts of the natural reading process are still unknown.

However, some discrepancies are also observed between eye movement effects and FRP modulation. For example, the influence of word frequency on the N1 time window in FRP is not observed during sentence reading, while eye movements corresponding to FRP show the effect of longer fixations for less frequent words (Degno et al., 2019; Kretzschmar et al., 2015). In ERP, the earliest that this effect is found is at 150 ms (Hauk & Pulvermüller, 2004). As a general note, these discrepancies emphasise the question: Which findings from traditional ERP research reliably generalise to a natural reading setting?

This thesis largely utilises co-registered EEG and eye-tracking methodology (**Study II, III & IV**) and, as such, a detailed summary of the intricacies of the method is warranted. The following summary illustrates that estimating unconfounded FRP's is not as simple as it would initially seem.

2 INTRODUCTION PART II: CO-REGISTERED EEG AND EYE-TRACKING

In order to study the impact of sublexical factors on brain activity during natural reading, co-registered ET and EEG methodology—henceforth ET-EEG—was chosen. In this part of the thesis, the intricacies of the ET-EEG methodology are explained. Particular attention is allocated to critical distinctions between fixation-related potentials (FRP) and event-related potentials (ERP), which are both electrophysiological brain measures. With ET-EEG, it is possible to observe the brain activity and gaze behaviour parallel but achieving this poses very specific requirements for recording, experiment setup and signal processing.

2.1 Measuring the ET-EEG signal

The first thing to consider with ET-EEG is the recording of the signals. Most EEG recording setups contain a rudimentary technique for registering eye movements—the electrooculogram (EOG). The EOG is a system of electrodes that are strategically placed around the eyes to capture the electrical voltage changes resulting from the rotation of the cornea-retinal dipole and from the muscle activity of the extraocular muscles. However, the EOG is quite ill-equipped for detecting relatively small differences in gaze position on a screen, as electrical perturbations caused by the smallest eye movements are indistinguishable from noise (Dimigen et al., 2009; Plöchl, Ossandón, & König, 2012). For purposes of successful ET-EEG measurement, modern video-based eye-tracking is a more suitable choice. Video-based eye-tracking exploits the reflection properties of the different parts of the eye by projecting infrared light into the eye. Then, the eye and resulting reflections are recorded using a high-speed video camera. The pupil appears as a dark circle/ellipse in the image and reflection from the cornea as a bright spot. The relationship between these two parts of the image changes based on the orientation of the gaze. The relationships between the pupil and the corneal reflection is measured at pre-determined points on the screen and, subsequent to this calibration procedure, it is possible to determine the gaze position on the screen using simple trigonometry—the requirement being a priori knowledge of the distance between the eye and the screen. From the gaze position, an angular velocity vector is calculated and the main movement states of the eye (fixation vs saccade) are typically defined either by a fixed velocity threshold or an adaptive velocity threshold (Engbert & Mergenthaler, 2006; Holmqvist et al., 2011). The gaze position measured with eye-tracker is an optical signal and the EEG is an electrophysiological measure of brain activity that is measured as a voltage difference between the reference electrode and the measurement electrode. Generally, only one reference electrode is used but the number of measurement electrodes can range from 1 to 256, depending on the capabilities of the amplifier and on the architecture of the measuring caps. The EEG signal is generated by the synchronous activity of thousands of cortical neurons; more specifically, by the synchronous post-synaptic potentials of these neurons (Nunez & Srinivasan, 2006).

Two things are critical when recording ET-EEG. 1) The two data modalities need to be perfectly aligned in time. This can be achieved either by converting the gaze position into voltage and recording it as an auxiliary data channel using the EEG amplifier or by inserting a sufficient number of time markers to both streams of data and aligning the data modalities offline (Dimigen et al., 2011). The more aligned markers there are in the data, the more resilient the dataset is to error resulting from clocks that perform differently, missing samples and other small-scale inaccuracies that sometimes occur. 2) The sampling rate of the two-time series should be close to each other. The sampling rates can be computationally matched offline but the temporal precision of the data is determined by the modality that has a lower sampling rate during recording. Synchrony should be checked after recording, either by examining the timings of the shared markers or by cross-correlating the optical and EEG signals from a channel that is likely to reflect the corneo-retinal dipole shift from eye movements (Dimigen et al., 2011). After recording (and synchronising, if done offline) the ET-EEG dataset is conceptually the same as the EEG dataset with an EOG or any other auxiliary channel attached to it. The true distinction comes with the paradigm used and the accompanying signal processing that utilises the optical signal.

2.2 Comparison of typical experimental designs in ERP and ET reading studies

The obvious strength of eye-tracking as a methodology is related to the capability to estimate the visual information that is passing through the neuro-cognitive system of an individual without restricting an individual's behaviour in any strict sense. The participant is free to view stimuli in any chosen manner, although the eye movement sequence for reading is largely set by the characteristics of the writing system that the text being read uses. In western alphabetic languages, the eye movement sequence when reading a sentence is typically left to right, with one to two fixations per word, depending on word characteristics and word predictability (Rayner, 1997). However, participants often skip words or return back within the text. Depending on the research question, different eye movement events are either considered meaningful or discarded as noise. The stimuli are segmented to areas of interest and to different eye movements in relation to them and their characteristics are quantified to make meaningful inferences about related cognition. In reading research, one of the more prominent stimulus types are single sentences and, quite often, analysis focuses on a single target word. This is because counterbalancing and other experimental control procedures are simpler to execute properly if only a single target word is manipulated. With high-end devices, it is possible to manipulate the stimulus based on the eye movements of participants. The basic limitation of eye movement research is the same as that of any behavioural research—the observed variables are end-state variables. In other words, eye movements are products of certain computational process in which researchers cannot determine what part of the computation in two fixations is different simply by comparing them even if they are of different duration.

With ERPs recorded using traditional designs, the differences in the process can be observed through direct comparison. If processing is differentiated immediately after stimulation, then it is observable at a corresponding part of the signal. Contrary to ET experiment designs, ERP experiment designs pose very strict limits for participants' behaviour. In reading experimentation, this is reflected in the instrumentation of rapid serial visual presentation (RSVP). In RSVP, participants are instructed to focus on some central point while words from larger text material—for example, a sentence—are presented in a successive manner to the foveal field (Dimigen et al., 2011; Hutzler et al., 2007; Kornrumpf et al., 2016). The duration of individual words is fixed to a certain value, which is typically longer than the average fixation duration, and there is typically a gap between words. The RSVP instrumentation thus radically differs in terms of how people behave when they are reading. RSVP does not allow participants to modulate the time used for each word, the rate of words is typically much slower than when reading freely, lookbacks and word skips are not possible and there are no parafoveal words to pre-process during exposure to a current word. To be thorough, it should be mentioned that there have been some attempts to introduce volition and parafoveal processing to RSVP—namely, the self-paced RSVP (Ditman, Holcomb, & Kuperberg, 2007), where proceeding from word to word is done by participants triggering stimulus change, and RSVP with flankers (Barber, van der Meij, & Kutas, 2013), where participants focus on words appearing in the centre while the centre word is flanked by previous and subsequent words. While these approaches partially introduce aspects central to natural reading to RSVP, the new unnatural aspects are also introduced. In self-paced RSVP, conscious monitoring of the stimulus rate is present, whereas eye movements are largely automatised in natural reading. Flanker RSVP, on the other hand, introduces the requirement of inhibiting saccades to the next word and, as eye movement and spatial attention have a close relationship, this inhibition of saccades might impact attentional functioning. Overall, there are good reasons for the methodological restriction of RSVP and these become apparent when the FRP methodology is discussed later on in this paper; however, as a general RSVP critique, the problem lies in its rather poor generalisation to the actual reading situation regarding visual presentation and participant behaviour. Figure 2 illustrates the differences between RSVP and unrestricted sentence reading. The differences between RSVP and natural reading are summarised in Table 1.

TABLE 1 Differences between natural reading and RSVP at a glance. RSVP = Rapid Serial Visual Presentation.

Feature	Natural reading	RSVP
Activity time lock	Fixation onset Saccade onset	Stimulus onset
Time allocated to each word	Determined by participant performance (~200–300 ms)	Determined by experiment program (>250 ms) Can be tied to participant response
Time between words	Determined by participant performance (~20–30 ms)	Determined by experiment program (>250 ms)
Returning in text allowed	Yes	No
Word skips	Yes	No
Parafoveal information available	Yes	Typically no Can be introduced with flankers

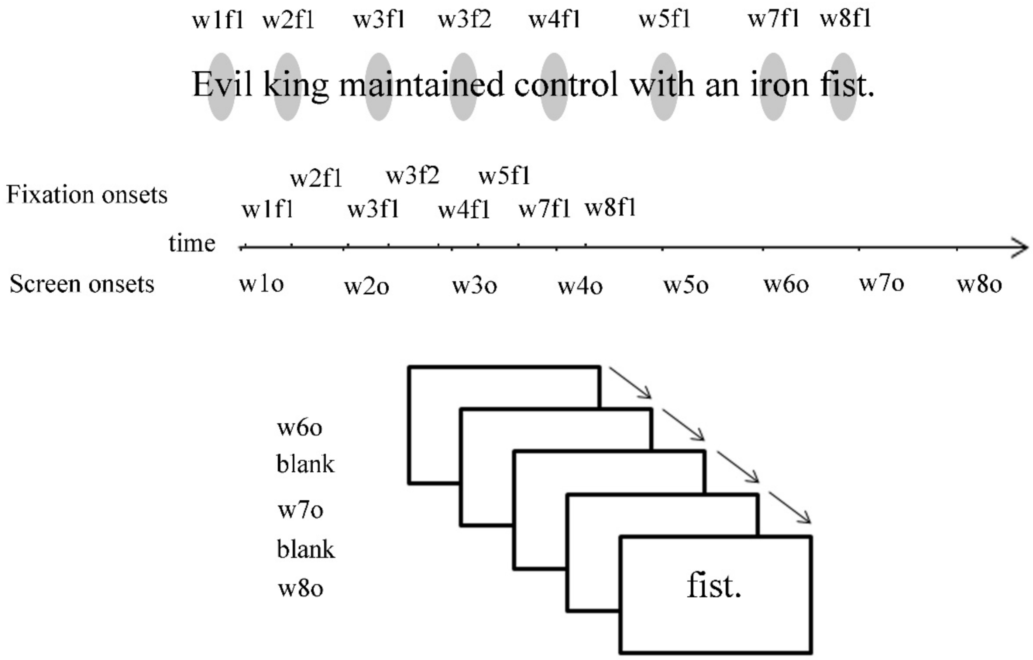


FIGURE 2 Comparison of free reading and RSVP time point selection principles. In free reading, the time points of interest are selected based on fixation onsets (tags w1f1–w8f1). The timing of the fixations and how many fixations are allocated for each word depends on participant behaviour. Grey ellipses show placement of the fixations on the words. The RSVP example is shown below, where the time points of interest (tags w1o–w8o) and direct exposure to words are determined by the program running the experiment. Note that, in RSVP, every second screen is often blank. Note: w = word, f = fixation, o = onset. Thus, w1f1 denotes word 1 fixation 1 and w2o denotes word 2 onset and so on.

Now, in ET-EEG, the main idea is to enable the use of experiment designs used in eye movement studies of reading and to capitalise on the window to processing stages present in ERP. This improves the ecological validity of the brain measures and the "black box" nature of eye movement variables is removed. However, the combination of free eye movements during EEG recording leads to some well-known and not so well-known problems at the response estimation stage.

2.3 Principles and issues of FRP estimation

For understanding the specific demands of FRP estimation, understanding the standard ERP principles is a good starting point.[2] In ERP estimation, general steps include filtering data to a certain frequency range, removing and interpolating channels that have excessively noisy signals, segmenting data based on events provided by the experimental program and removing bad segments that contain strong voltage fluctuations outside the reasonable brain activity range (Luck, 2005). In traditional ERP's, correcting artefacts that have systematic spatial patterns, such as eye movement artefacts, computationally is optional. EEG segments containing eye movements or blinks can also be rejected but computational correction is generally recommended if rejecting artefactual segments would lead to catastrophic data loss. The ERP estimation procedure culminates in a final step in which the segments are baselined and averaged (Luck, 2005).

Simply put, in FRP estimation the stimulus onset—as a defining event—is replaced by the fixation onset at a certain part of the stimulus—for example, a word in a sentence. The phenomenon that allows for equal treatment between stimulus onset and fixation onset is called saccadic suppression. Saccadic suppression refers to a pause in the neuronal input from the retina to the visual cortex during saccades (Bridgeman et al., 1975; Thiele et al., 2002). In fixation onset, this pause ends and a response that is very similar to visually evoked potential is generated. However, the eye movements and how they are modulated by the cognitive demands of each task cause issues that need to be solved before reliable inference from the FRP can be made. These issues and their known solutions are listed in Table 2.

---

[2] In this part, procedures are discussed on conceptual stage. For specific mathematical implementations, please refer to the publications cited.

TABLE 2 Known methodological issues in fixation-related potentials arranged in an ascending order of severity and effective solutions. ICA = Independent Component Analysis, MSEC = Multiple Source Eye Correction.

Issue	Impact	Influence	Solution(s)
Cornea-retinal dipole rotation.	Fast onset, slow dissipation strong artefacts.	Linear	ICA MSEC
Spike potential.	Short transient artefacts with relatively low power. Potential confound, especially in frequency analyses of the gamma-band.	Linear	ICA
Experimental manipulations and stimulus features modulate saccade amplitude, which modulates visual response.	Potential confound in early visual responses between conditions.	Non-linear	Saccade length matching Spline regression model
Experimental manipulations and stimulus features modulate fixation durations, which modulates how responses from fixations overlap.	Differently overlapping activity from adjacent fixations confounds the inference between conditions. Differently overlapping activity from adjacent fixations makes commonly used baseline sections non-uniform, which can lead to strong voltage shifts in overall FRP.	Linear	Fixation duration matching Linear Deconvolution

The first problem on the list—the artefacts from the cornea-retinal dipole rotation—is also present in any set of EEG data, as eye movements occur to some degree during any task. In FRPs, this problem is more severe because ocular artefacts are time-locked to the signal of interest. However, this problem can be solved with a commonly used blind source separation technique known as Independent Component Analysis (ICA)[3] (Lee, Girolami, & Sejnowski, 1999; Makeig, Bell, Jung, & Sejnowski, 1996; Onton, Westerfield, Townsend, & Makeig, 2006). The ICA algorithm produces a group of channel weights that, when multiplied with the data, generate a set of time courses that have as low zero lag dependencies to one another as possible. In order to work with EEG, ICA relies on several

---

[3] There is large amount of ICA algorithms, but for purposes of this thesis when ICA acronym is used, extended Information Maximisation algorithm is what is meant. Also even though ICA can be utilised to perform blind time deconvolution, focus here is on the application to source separation problem.

assumptions: first, the signals from different brain sources are assumed to be temporally independent; second, the spatial sources are stationary through time; third, the activities from brain sources combine linearly to form the recorded signal. Given that these assumptions are correct, the following procedure is available:

Decompose EEG → Identify artefact time courses → Rebuild EEG without artefact time courses.

The logic behind the procedure is simple but its implementation often proves difficult. For corneo-retinal dipole rotation, which causes the most prominent artefacts for ET-EEG recordings, the procedure works efficiently and is generally a basic requirement for observing FRP characteristics. However, when applying ICA for removing EEG perturbations from extraocular muscles, the spike potential, some adjustments to the procedure need to be made. The first issue related to spike potential (SP) is that the associated topography in channel weights is not as easily identified by visual inspection as the corneo-retinal dipole rotation is. This can be solved by comparing saccade timings to component time courses—if a component shows activity mainly during saccades, then it is likely to be generated by an eye-related source (Plöchl et al., 2012).[4] The temporal covariance principle also effectively identifies the components that correspond to the corneo-retinal dipole rotation artefacts. The second spike potential issue is related to the fact that the spike potential amplitudes can be weaker than the event-related activity that we are interested in. Thus, if there is a high number of temporally non-stationary artefacts, such as low-frequency fluctuation from sweating or impedance changes on individual electrodes, then ICA might be unable to reliably separate spike potentials from other components. However, it has recently been suggested that SP detection could be enhanced by creating a training dataset in which high pass filtering is much stricter (1–3 Hz) and data-segments that contain SP are copied to the end of the dataset (Dimigen, 2018). The training dataset is used to derive component weights with ICA and these weights are then used to decompose the original dataset, which has typical pre-processing parameters, where the time courses of SP and other eye-related artefacts are left out in the reconstruction phase. The optimisation of the dataset to improve ICA results can also be beneficial in traditional ERP designs.

It must be mentioned that ICA is not the only available method for correcting eye movement related artefacts. For example, early FRP research (Dimigen et al., 2011) utilised the Multiple Source Eye Correction (MSEC) (Berg & Scherg, 1994) approach, which relies on establishing a source model capable of separating eye movement related artefacts and genuine brain activity from one another.

---

[4] Comparing IC time course to optical signal has the benefit of neuronal activity not being directly reflected in the optical signal. For example, identifying artefactual IC by dependency to an EOG channel risks identifying actual brain sources as artefacts. This is because EOG electrodes also measure brain activity in addition to registering the corneo-retinal dipole rotations and extraocular muscle activity.

The first problem exclusive to FRPs is that eye movement behaviour is modulated by stimuli and task and that eye movement characteristics, in turn, directly modulate the FRP components. This introduces confounds for determining whether the observed FRP effect is directly due to interesting manipulation or indirectly due to change in EM behaviour resulting from manipulation. One known example of such influence is saccade amplitude, which has been reported to influence the P1 amplitudes of the FRP (Nikolaev, Meghanathan, & van Leeuwen, 2016; Ries, Slayback, & Touryan, 2018). An example of this modulation is presented in Figure 3. This relationship is problematic because a stimulus feature, or participant characteristic, can influence saccade amplitudes in some cases and, thus, a confound for interpretation is introduced. For example, it could be difficult to determine whether an observed P1 modulation is a reaction to the stimuli or is, in fact, a result of the mean saccade amplitude difference or whether a lack of effect in FRP is due to contrary stimuli effects or group characteristic and saccade amplitude. For example, impaired reading skill is associated with shorter saccades, on average, than typical reading skill (De Luca et al., 1999) and this confound needs to be handled if FRP P1 responses are to be reliably compared between slow and typical readers.

The first proposed solution to this issue was subsampling from available fixations so that there would be no statistical difference in previous saccade amplitude, on average, between conditions or groups, followed by testing FRP for differences between conditions. More sophisticated versions of this approach involved calculating and minimising the Mahalanobis distance, which is an effect size measure in multivariate statistics on the basis of matching eye movement features between conditions (Nikolaev et al., 2016). This matching procedure suffers from a few issues. First, it is contradictory to the aim of using more naturalistic situations to study brain activity because matching the eye movement correlates of cognition limits the manifestation of the observed pattern and, thus, diminishes the ecological validity. Second, even if we remain in the simple example of saccade amplitude, this matching procedure might lead to critical loss of signal for the estimates to be reliable. Third, not all influences of eye movements on FRP are known and, thus, the inclusion of additional factors would make the previous problem more severe. Fourth, it is unclear what eye movement features scale linearly with brain activity, thus removing statistical significance between conditions in eye movement factors might not remove their significant influence on FRP. Essentially, the flaws of the matching procedure stem from the requirement that categories of observations need to be established for traditional averaging-based estimation. Fortunately, category-based averaging can be switched to a more flexible framework.

Instead of trying to balance the confounding factors in condition averages, the FRP estimation can be approached through a General Linear Modelling framework (GLM) (Cornelissen, Sassenhagen, & Võ, 2019; Kristensen, Rivet, & Guérin-Dugué, 2017). Categorical waveforms can be estimated with procedures that are nearly identical to averaging, with the key difference being the treatment of the noise—which is expressed as a separate term in GLM but is incorporated in the average (Smith & Kutas, 2015a). This, in itself, does not help with the problem of the covariance of eye movements and FRPs causing confounds—but the capability of the GLM to handle continuous variables does. Now, saccade amplitude can be introduced as a continuous predictor to separate its influence on FRP in a separate parameter and the categorical effect in another parameter, given that there are enough observations. It must also be noted that stimulus features can be estimated as continuous factors when they are truly continuous in nature. In the domain of reading research, such features would be, for example, word frequency and word length.

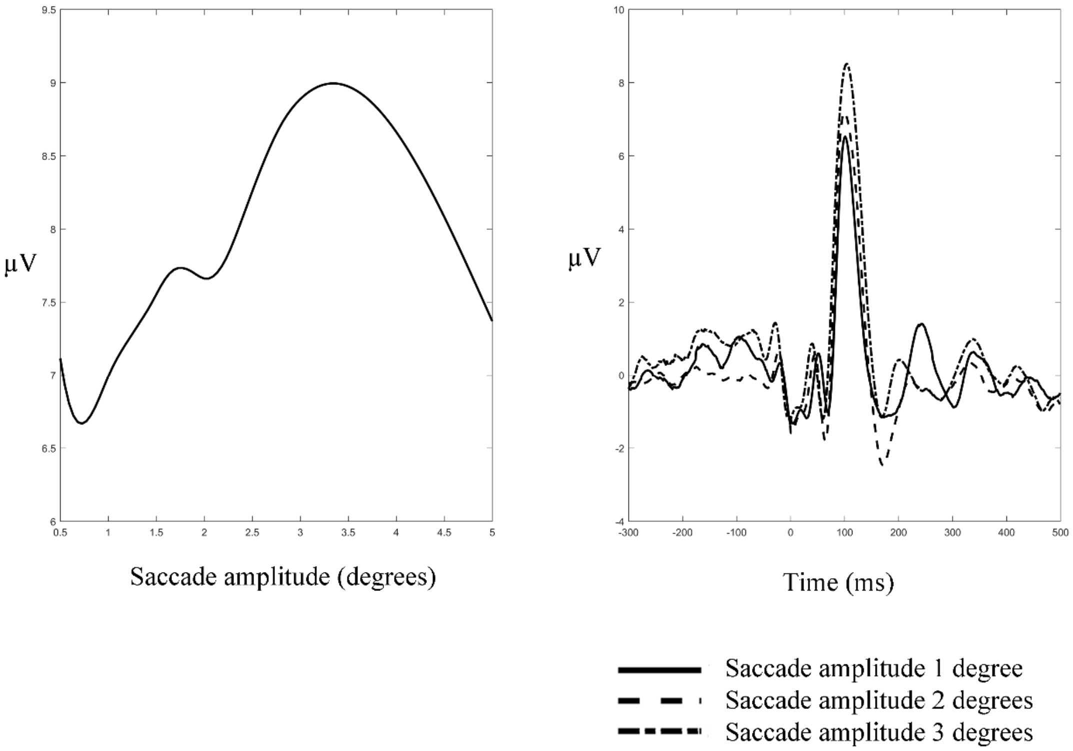


FIGURE 3 Example of the influence of saccade amplitude on FRP morphology. Deconvolved estimates were calculated from data of 10 participants and then averaged. **The left panel** shows the influence of saccade amplitude on channel E75 at 100 ms from fixation onset. This time point corresponds with the P1 peak. Note that saccade amplitude also has an impact on the P1 peak latency and, thus, a decrease in voltage after 3.5 degrees can also result from the P1 latency modulation. **The right panel** shows FRP on channel E75 with different preceding saccade lengths. Note that the impact on the P1 latency seems to be systematic, albeit non-linear.

Of the known issues with FRP, the temporal overlap is the most problematic. It occurs because fixations during reading have a very close temporal proximity and neural correlates of cognition, which are often examined span outside the fixation duration. Timing of fixations and saccades is affected by cognitive processing, which consequently modulates the overlap. Thus, different timings of the fixations might blur the components of subsequent and prior FRPs in FRP in a way that could mask itself as a differential activity between conditions. An example of overlap modulation by fixation duration is provided in Figure 4. Also, differential overlap in segments prior to the fixation onset is problematic because using such segments as the baseline causes distortions to the values of the entire waveform. This problem applies to both averaged FRPs and GLM estimated FRPs if the standard application of estimating from epoched data is used. The matching procedure that was suggested for saccade amplitude, was also suggested for fixation durations to minimise overlap confounds (Nikolaev et al., 2016); however, the same problems (as those covered for saccade amplitudes) render this solution unsuitable.

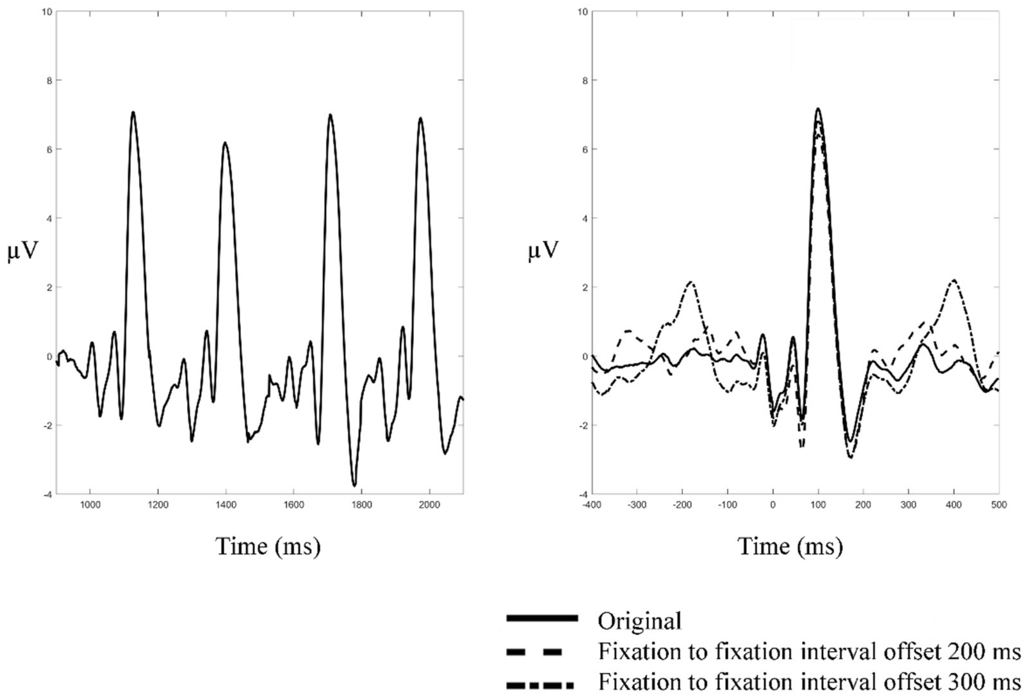


FIGURE 4 Example of confounds to averaged FRP estimates generated by a differential overlap between conditions on channel E75. **The left panel** shows a segment from a signal that was generated by summing grand average deconvolved FRP (-700 to 500 ms around fixation onset) with 300 ms fixation to the fixation offset and Gaussian random jitter (SD = 44.73 ms). Note that the same segment repeats in the sequence; thus, the difference in between peaks and in peaks is caused by overlap differences. Hence, the differences in P1 responses are due to the overlap. **The right panel** compares the original FRP to FRP segmented and averaged from the signal generated with either 200 ms or 300 ms fixation to the fixation offset. Note that after roughly 200 ms, the fixations with simulated overlap begin to clearly differ from the original, introducing potential confounds for the interpretation.

The GLM framework, however, can be adapted to handle the overlap directly (Cornelissen et al., 2019; Ehinger & Dimigen, 2018; Kristensen et al., 2017; Smith & Kutas, 2015b). Basic GLM approaches the FRP estimation by solving multiple instances of the same model for each time point in each channel across the trial dimension of the epoched segment; however, for overlap correction relative time, from activity generating events like fixations, is used. Figure 5 illustrates how including time points as predictors in the GLM framework can separate the influence of overlapping FRPs. For similar explanations, please refer to Figure 2 in Ehinger and Dimigen (2018) and Figure 1 in Cornelissen et al. (2019).

To explain in more detail, the linear deconvolution method sets up the entire dataset encompassing model in which each time point is expressed as a set of predictors and noise. Overlap correction is introduced in the form of relative time point predictors in samples prior to the fixation onset and after the fixation onset. Then, the model is solved using least squares approximation to obtain the estimate for contributions of each relative time point and predictor.

For example, in Figure 5, time point 26 is modelled as a linear combination of the relative time point 15 of the first fixation onset on word 1, the relative time point 4 of the re-fixation on word 1, the relative time point -2 of the first fixation on word 2 and noise. Each time point in the entire dataset model is derived based on the event-relative window selected—if a relative time point does not belong to a relative window, then that relative time point is not used as a predictor for brain activity during that time point. Here, it is important to make a decision about the length of the response one wishes to estimate, as this decision sets the degree to which the overlap is assumed to be present (Smith & Kutas, 2015b). It is better to have a too long window than a too short one. This is because if a brain response from any given fixation is particularly long lasting, as in case of P600, then it will overlap multiple fixations, and if the set window extent does not correspond to the actual overlap, then the overlap in that part will remain a confounding influence, unless it is not captured by the noise parameter. For example, in Figure 5, time point 13 is expressed with relative time point predictors from the first fixation onset on word 1 and the re-fixation on word 1 but the relative time point of the first fixation onset on word 2 is not. In this situation, if the first fixation onset on word 2 would have an impact on time point 13, then this influence would not be modelled.

Up to this point, the explanation has only focused on how relative time points are used for overlap correction as a simple intercept but, as explained before, the potentially confounding influences, such as saccade length, need to be considered as well. This is done by also adding these factors as relative predictors. For example, if we add the saccade amplitude as a linear continuous predictor, then time point 13 could be expressed as: relative time point 2 of the first fixation onset to word 1, continuous influence of saccade length preceding the first fixation onset to word 1 in relative time point 2, relative time point -9 to the additional fixation onset to word 1, continuous influence of saccade length preceding the additional fixation onset to word 1 in relative time point -9 and noise.

The linear deconvolution has recently been implemented into an easy to use toolbox, *Unfold* (Ehinger & Dimigen, 2018), where complex and entire dataset encompassing deconvolution models can be expressed using the Wilkinson notation. For example, the previous overlap model example with a model of saccade amplitude can simply be expressed as:

"y~1 + previous saccade amplitude"

and, with definition of the desired time window for the overlap, the entire dataset regression equation is established automatically. After solving the model, the individual contributions of each relative time point predictor, eye movement predictor and possible stimulus related predictor can be extracted and presented as meaningful waveforms that have a very close resemblance to averaged FRPs, the distinction being that the problematic overlap has been corrected. The linear deconvolution approach has been shown to be a reliable way for estimating overlap corrected FRPs (Cornelissen et al., 2019; Kristensen et al., 2017).

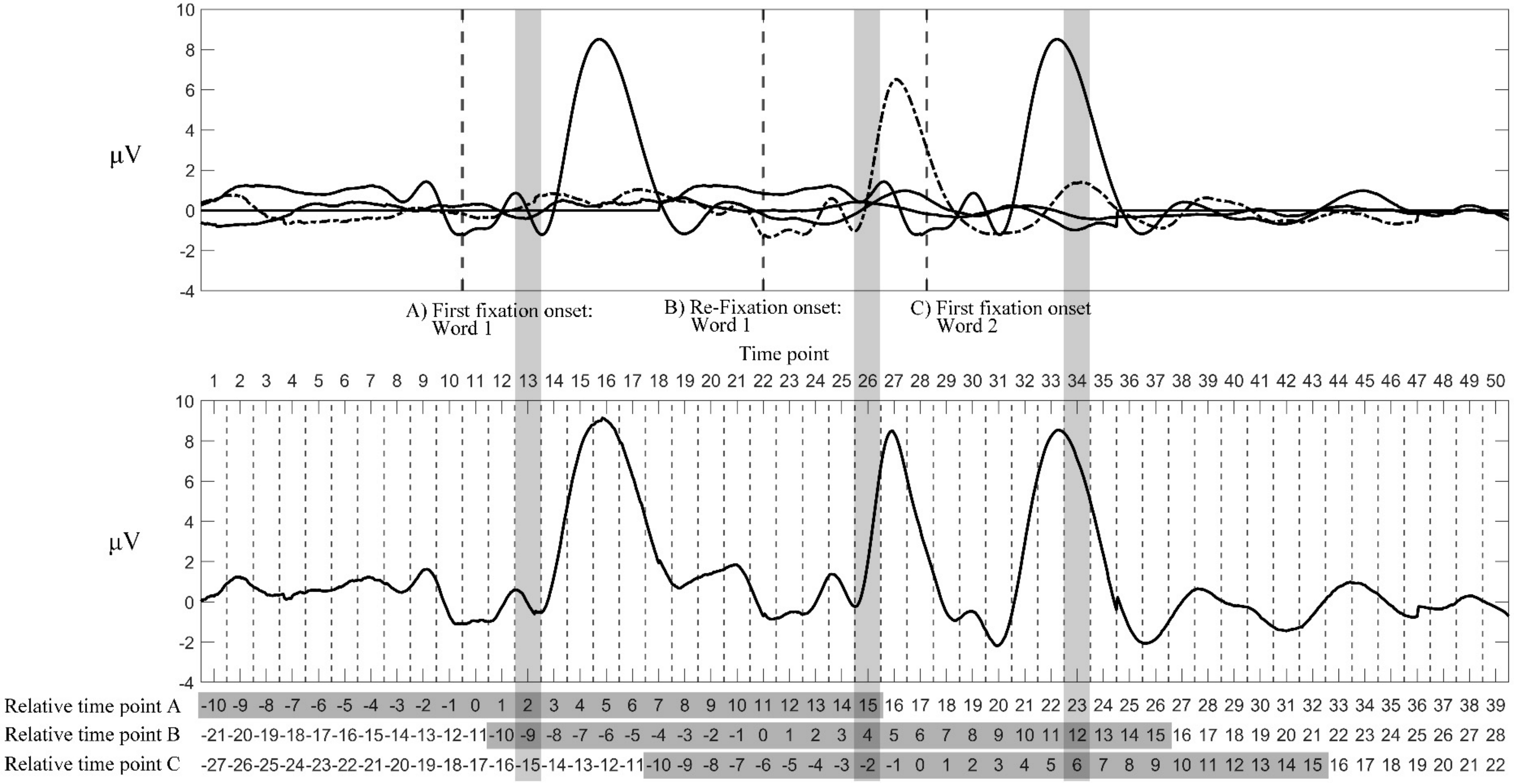


FIGURE 5 Underlying principle of the linear deconvolution method. Upper panel: Original responses from fixation onsets. Lower panel: Linearly combined signal.

In conclusion, FRPs are alternatives to ERPs, where the stimulus onset has been switched to a fixation onset when estimating the response. This switch increases the complexity of the estimation procedure. Eye movement-related artefacts need to be reliably removed. Contributions of eye movements to brain activity need to be modelled in order to separate factors from factors of interest. Most importantly, the overlap resulting from close proximity to previous and subsequent fixations needs to be corrected. Artefacts can be modelled using ICA and removed, which is a straightforward standard procedure in modern EEG processing. The latter two can be solved by switching averaging to a GLM-based regression framework, which is better suited for FRP estimation than averaging. Figure 6 shows a simplified overview of the alternative pipelines for FRP estimation.

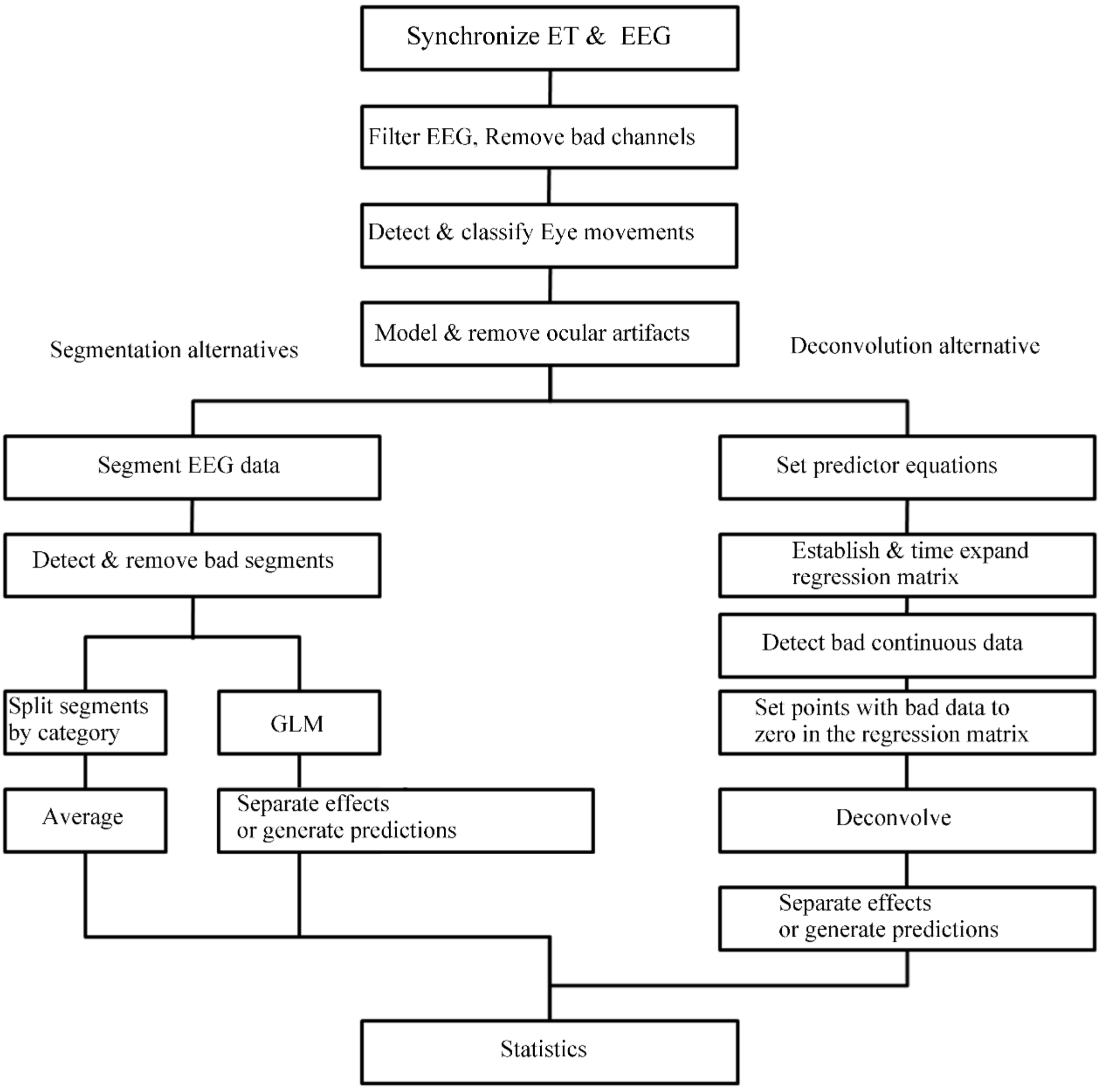


FIGURE 6 Comparative summary of the FRP pre-processing steps with and without overlap correction. The pipelines diverge after removing ocular artefacts to a segmentation-based averaging/GLM pipeline that does not implement overlap correction and to a GLM-based linear deconvolution pipeline that corrects for overlap.

2.4 Considerations in the statistical analysis of FRP

After estimating the FRP and quantifying the eye movements, there are no FRP specific steps that would need to be considered, as long as eye movements during discarded time points are not included in the statistics in order to have the same information in the statistical analysis of both modalities: eye movements and FRP. Even though the FRP estimation has been under significant methodological development during the last five years, the literature does not yet recognise any specialised statistical approach to FRP analysis. Thus, the differences in statistical approaches that are present in the eye movement and ERP literature need to be considered to make valid conclusions.

Statistical approaches for eye movements can be divided into two classes: approaches that require aggregating over single observations and those that allow for single observation structure to be preserved. When performing statistical analysis with Analysis Of Variance (ANOVA), categorical mean estimates are first calculated either across items (F1) or across subjects (F2). ANOVA, or its variant, is then performed on these estimates to determine whether there is statistical support for the difference between categories. However, this is often problematic because the amount of signal in the estimates might depend on the phenomenon studied. In addition, F1 and F2 tests often give results that have discrepancies.

The mainstream of the existing eye movement literature has largely replaced the traditional ANOVA approach with Linear Mixed Effects modelling (LME) (Baayen, Davidson, & Bates, 2008). In LME modelling, there is no aggregation of data over items or subjects. Instead, the LME model makes a distinction between fixed effects and random effects. Fixed effects are those effects that a researcher is interested in, such as, for example, experimental manipulation or group comparison, while random effects are the confounding influences present in data, for example, the variability between stimulus items and non-systematic differences in the subject performance. LME separates random effects from individual items or subjects into independent parameters so that they do not confound the fixed effects' results. The primary benefit of preferring LME to traditional ANOVA is that there is no longer a risk of interpretational ambiguity arising from a discrepancy between F1 and F2 analyses, as they are completed simultaneously. LME modelling also handles missing data flexibly (Baayen et al., 2008) without a need for data imputation. In addition, the statistical power for observing the same effect is considerably increased as each observation adds to the error degrees of freedom. The increase of statistical power needs to be treated carefully because it allows for the detection of very small effects.

Conversely, the ERP/FRP statistical analysis quite strictly follows the logic of the F1 analysis of repeated measures in ANOVA. For establishing statistical significance, there is no methodological difference between FRP that was estimated using averaging, basic GLM or linear deconvolution methods. The problem for ERP/FRP is that the amount of potential comparisons is very high because any of the time x channel pairs can be tested. One of the approaches to this problem has been to calculate grand averages, looking for a sweet spot of the effect and subjecting that spot to a statistical test (see a critique of localising the statistical test in this way in Luck & Gaspelin, 2017). This approach, however, elevates the probability of falsely rejecting the null hypothesis in favour of the alternative hypothesis (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009), similarly to the multiple comparison error. The more channels are estimated, the more precarious this approach becomes. The simplest and most effective way to mitigate the risks of elevating the probability for false positives is to derive strong a priori predictions on the location of the tested effect in spatial and temporal dimensions before seeing the data. This is feasible in well-established research traditions, such as ERP research on auditory processing using the oddball paradigm. However, the amount of existing FRP research is severely limited to support strong a priori hypotheses. An obvious argument, then, is to use results of previous visual ERPs to guide the formation of the hypotheses; however, because of all the criticisms established in the previous section on RSVP research, it is largely unknown how the RSVP ERP results generalise to naturalistic FRP research. From this, it essentially follows that statistical approaches, which consider all channels and time points of the data, are preferable to scalar value statistic methods like ANOVA.

Non-parametric cluster-based permutation tests are the more prominent approaches that allow for a comparison of the ERP/FRP responses without specifying tight temporal or spatial constraints (Maris & Oostenveld, 2007). In this approach, the desired statistic is first performed on all channels and time points and the results are combined into clusters by adjacency. Cluster formation relies on the fact that most ERP/FRP effects are present on multiple channels and time points simultaneously and these channels and time points are adjacent. In the temporal dimension, determining adjacency is trivial because time is a continuous factor but, for the spatial dimension, some thought is required because channels have gaps between them. Thus, adjacency needs to be defined by estimating spatial relations, often referred to as defining the neighbourhood, between channels. Two techniques for neighbourhood definition can be distinguished. A channel can be considered to be a neighbour to another channel if it is found within a defined radius of that other channel. Another approach relies on triangulation, which is independent of the distance between sensors. With triangulation, the algorithm neighbourhood is defined by building as many triangles as possible by connecting channels with lines without the lines crossing. Now, with temporal adjacency and channel neighbourhoods defined, similar test results are combined into a cluster and a test statistic for that cluster is defined. Again, two major alternatives are recognised: the extent of the cluster, which is simply the number of observations that constitutes the cluster, and the mass of the cluster, which is the sum of the base statistics, such as t-value, from the observations that constitute the cluster.

After cluster definition, a Monte Carlo simulation is run to establish a surrogate null distribution for determining the probability of observing clusters with such a test statistic in the given dataset. In practice, this means that the condition labels (or group labels if performing the between-subjects test) are randomly assigned, the base test is performed on this randomised data, the clusters are established and the largest test statistic from detected clusters is then deposited to the surrogate null distribution. These steps are run a sufficient number of times (>500) to establish a robust surrogate null distribution. The original clusters are then compared to the surrogate null to see whether the clusters observed in the main test are less probable than the ones observed from random permutations. If one or more of the clusters from the main test is less probable than the desired alpha level, then the effect is considered to be statistically significant.

Non-parametric cluster-based permutation tests have two general caveats. While they are highly effective in controlling the familywise error rate, they do so at a cost to sensitivity. In particular, in the presence of stronger clusters, the weaker effects might get corrected out. This can generally be countered by limiting the span of the test or splitting the test into separate time windows. For this to be a proper solution, a priori justifications are needed—otherwise, the non-parametric cluster-based permutation test is similarly subjective to double dipping as any other statistical solution. The second problem is that, when clusters are often presented with a high amount of detail to aid the interpretation of an underlying phenomena, it is tempting to make conclusions about specific time points or channels contained in the cluster that are too strong. The problem here is that only the test statistic—for example, mass of the cluster—is corrected for familywise error rate. Thus, drawing extensive and strong conclusions on any other feature of the cluster, which has not been tested, needs to be avoided (Sassenhagen & Draschkow, 2019). This means that, for example, the onset, offset or extent of the cluster cannot be determined with statistical criteria when using non-parametric cluster-based permutation tests even though such parameters can be extracted from the test.

However, using non-parametric cluster-based permutation tests or any F1 test that requires calculating participant-level estimates for establishing effects in FRP waveforms leads to a situation in which statistical tests for eye movement variables are much more sensitive to detecting effects. This is because the single-trial structure is preserved in LME statistics, which, in turn, translates to very high degrees of freedom. For example, in a hypothetical study with 20 subjects and 2 stimulus categories with 100 individual stimuli in each category, the dependent samples' FRP t-test would have 19 (20 – 1) degrees of freedom. On the other hand, the LME model with random factors for subject and stimulus items would have 3,779 (4000 – 20 – 200 – 1) error degrees of freedom. From this, it follows that the current state-of-the-art approaches are not equally sensitive for detecting effects in both eye movements and FRPs. Thus, one possible explanation for situations in which an effect in eye movements is found—but no corresponding effect is found in the FRP estimates—is that there are imbalanced degrees of freedom between the tests.

It must be noted that, recently, pipeline incorporating cluster-based permutation statistics were suggested as a step for targeting single-trial LME on ERP segments (Frömer, Maier, & Rahman, 2018). The general problem with this suggestion is that, if this approach is performed within a single dataset, the statistical analysis would be guided by the results of a statistical test—which can be considered as double-dipping (Kriegeskorte, 2009). If this approach is implemented so that the LME targeting is done based on a separate independent dataset, then there is no problem. For FRP, on the other hand, this procedure is not particularly beneficial because correcting for temporal overlap is a critical issue and linear deconvolution does not preserve the single observation structure. Thus, the benefit gained from analysing the deconvolved estimates using LME would come from handling random influences from subjects, provided that the subject populations are large enough for this to be efficiently capitalised on, as estimating subject random factor saps error degrees of freedom, but the potential for handling missing data by adjusting item and subject weights is not available.

In summary, both free reading and how tasks and stimulus properties modulate the probabilities on what data is obtained place specific demands on statistical approaches used. In ET literature, this has been solved using LME models that handle the imbalances arising from missing data by adjusting weight based on how much data each subject and item contribute; however, for FRP, there is no satisfactory approach for handling this issue yet, as single observation structure is not preserved in the overlap correction process. It can be argued that methodological development has not yet achieved a stage at which all problems could be simultaneously solved with reasonable computational costs. Thus, when performing ET-EEG research that employs FRP methodology, a critical choice has to be made: Which one of the following three methodological weaknesses is your research capable of tolerating:

A. A less naturalistic set of observations constituting the FRP estimates,
B. A temporal overlap in the FRP estimates or
C. An unbalanced signal in the FRP estimates and eye movement variables by variable of interest?

By combining current methodologies, two of the three problems can be solved simultaneously but one would remain. For example, if one chooses to study word skips and how predictability influences brain activity associated with word skips, one would need to consider how the number of word skip observations is correlated with the predictability of the words that are skipped (Kliegl et al., 2004). If we conceptualise this example in a simple statistical design, where one would wish to examine the 2*2 (high vs low predictability * skipped vs not skipped) factorial structure, we would likely observe a high incidence of word skips for words with high predictability and a high incidence of no word skips for words with low predictability—from these two combinations, we would then most likely observe a reasonable amount of signal for deriving reliable FRP estimates. However, in order to disentangle the brain activity related to predictability and the brain activity related to word skipping from one another, we would need to estimate FRP from words with high predictability, which are not skipped, and from words with low predictability that are skipped. Thus, the FRP estimates from the latter two would have a considerably smaller amount of signal than the estimates from the former two. Because the information about the amount of signal in the estimate is not considered in the statistical comparison, it may confound the results due to a different signal to noise ratio, leading to the third abovementioned methodological weakness (C). One could try to remedy this by balancing the FRP estimates through discarding some of the observations of the categories with a higher amount of signal. However, subsampling eye movement observations leads to problem A, where distributions of eye movements no longer reflect how they naturally occur. Similarly, one could choose not to perform an overlap correction but to analyse trial level EEG segments over subjects with LME, letting the signal discrepancies be handled with weighting procedures internal to the method. This, however, leaves the overlap—problem B above—as a potential confound. It must be noted that linear deconvolution and LME are mathematically compatible (Ehinger & Dimigen, 2018) but that the estimation of a massive multiple subjects model requires vast computational resources that are often not available. As such, A, B and C can, in theory, be simultaneously solved but, in practice, this capability is severely limited.

3 AIMS OF THE STUDY

Part I of the Introduction established that processes related to reading are reflected in both eye movements and brain activity. However, the two are rarely examined together. This is not surprising because measurement instrumentation, data pre-processing, response estimation and statistics are thoroughly affected by examining reading behaviour in natural circumstances instead of strict, specialised experiment designs. All these factors need to be approached by considering how to handle the challenges introduced by natural reading behaviour. Required considerations were laid out in part II of the Introduction.

This dissertation aimed to expand the existing knowledge of the influences that sublexical factors extend on eye movements and brain activity during natural reading. It also examined, to a lesser extent, the influences of reading skill on brain activity and eye movements during natural reading. Overall, one overarching aim of this dissertation was to examine eye movements simultaneously with brain activity in more naturalistic settings, which is something that has not been done in majority of existing reading research. This was achieved through utilising co-registered EEG and eye-tracking methodology.

Study I aimed to disentangle the contributions of spatial width and the number of letters on eye movements in dysfluent- and normal-reading adults. Previous studies have shown that that spatial width of the words has contributions on eye movements during natural reading that are independent from the number of letters (Hautala et al., 2011). In proportional fonts, the spatial extent of a word also is closely related to visual crowding, as words with multiple j or i letters have a smaller visual extent than words with multiple m or w letters. Crowding has recently been suggested to have an influence on reading speed for both normal and dysfluent readers (Perea & Gomez, 2012; Perea, Panadero, Moret-Tatay, & Gómez, 2012). A higher number of letters in a restricted spatial space, as it naturally occurs with proportional fonts, was therefore expected to impair word recognition and this effect was expected to be more severe for dysfluent readers.

Study II aimed to investigate the impact of word length and reading skill on brain activity during natural reading in slow- and typical-reading children.

The question was investigated using an exploratory approach on an unexamined part of data from Study III. Previous studies have shown that dyslexics had reflected a stronger word length effect in their eye movements than normal readers (Hawelka et al., 2010). In ERP studies, dyslexics were shown to have different N1 response modulations than normal readers (Fraga González et al., 2014; Maurer et al., 2007). However, the impact of word length on brain activity has not been investigated in children during natural reading. Thus, it was expected that the effect of word length on brain activity would differ between slow- and typical-reading children within the time constraints of fixation during natural reading.

Study III aimed to investigate the impact of semantic anomalies on eye movements and brain activity in typical-reading children during natural reading. A particular point of interest was the impact of orthographic similarity of anomalous words to plausible words and whether this similarity would influence brain activity in a distinct manner. In previous studies, semantic anomalies have been shown to impact brain activity (Kutas & Federmeier, 2011; Metzner et al., 2016) and eye movements (Metzner et al., 2016; Veldre & Andrews, 2016, 2018). Hence, it was expected that orthographic similarity to plausible words would alter the eye movement and brain activity dynamics for detecting semantic anomalies during natural reading.

Study IV aimed to investigate the impact of spatial location and types of sublexical deviation from actual words on brain activity, small saccades[5] and reaction times during single word recognition. In earlier studies, it has been established that sublexical deviations attract microsaccades (Hautala & Parviainen, 2014), suggesting that modulations by spatial attention influence the process. Therefore, it was expected that deviations inserted into words would modulate small saccade orientation and latency. Furthermore, it was expected that both co-occurring and prior brain activity effects would be found.

---

[5] The saccades in this study are considered to be under voluntary control and, thus, the term small saccades is used instead of microsaccades.

4 METHODS

Methodological details of the four studies are discussed in this chapter. A summary of the methods is presented in Table 2.

TABLE 3 Summary of methodology in the studies. TW = Target word.

Study	**Paradigm**	**Number of participants**	**Saccade detection**	**EEG preprocessing pipeline type**	**Statistical approach**	**Source analysis method**
I	Sentence (TW)	34	Fixed velocity threshold	N/A	Repeated measures ANOVA(F1 & F2)	N/A
II	Sentence	92	Median velocity threshold	Linear deconvolution	LME Non-parametric permutation test	CLARA
III	Sentence (TW)	66	Median velocity threshold	Standard averaging pipeline	LME Non-parametric permutation test (Scalp & Source)	Cortical LORETA
IV	Lexical decision	19	Median velocity threshold	Linear deconvolution	LME Non-parametric permutation test (Scalp)	N/A

4.1 Participants

In **Study I,** 34 students from the Jyväskylä area, Central Finland, participated. **Studies II** and **III** were part of the eSeek! Internet and Learning Difficulties – Multidisciplinary Approach for Understanding Information Seeking in New Media project and, thus, had access to a large population sample of 12 to 13.5-year-old school children. In **Study II,** 92 elementary school pupils from 12 to 13.5 years of age were drawn as participants from a larger sample. In **Study III**, 66 elementary school pupils from 12 to 13.5 years of age were drawn as participants from a larger sample. In **Study IV,** 19 university students participated. For **Studies I** and **IV**, the participants signed written informed consent forms prior to being included in the study. For **Studies II** and **III,** parents of the participants signed written informed consent forms prior to their children being included in the study. The studies were conducted in accordance with the Declaration of Helsinki and the Ethical Board of the University of Jyväskylä approved the study protocols.

4.2 Cognitive measures

In **Study I**, the participants reading skill was assessed using word lists and text reading subtests from an assessment test battery for reading abilities in young adults. The participants were considered to be dysfluent readers if they performance was weaker than that of the 11% of the normed population in either test. The rest were classified as normal readers. The participants IQ was measured as a control variable using Raven's standard progressive matrices.

In **Studies II** and **III**, the reading skill of the participants was assessed using three tests: the word identification test, which is a subtest of the Finnish ALLU reading test (Lindeman, 1998); the word chain test (Nevala & Lyytinen, 2000) and the oral pseudoword reading test (Eklund, Torppa, Aro, Leppanen, & Lyytinen, 2015). The participants' scores on these tests were reduced into a single Reading Fluency factor with principal axis factoring using PROMAX rotation in IBM SPSS 24 statistical software. For **Study II**, the participants belonging to the weakest 15% on the Reading Fluency factor were classified as slow readers (N = 27) and the rest were classified as typical readers (N = 65). For **Study III**, the participants performing below 13% were removed from the sample while the rest were classified as typical readers (N = 66)[6] and included in the study.

In **Study IV**, the reading skill was measured using a text reading subtest from a reading difficulty screening test for young adults (Nevala, Kairaluoma,

---

[6] These group definitions were set as in the larger eSeek project. Slight discrepancy in the definition of Typical and Slow readers between Study II & III is due small change in the group definitions in the larger eSeek-project.

Ahonen, Aro, & Holopainen, 2006). The participants performing above 11% in relation to the normed population were accepted in the study (N = 19).

4.3 Procedures and stimuli

Study I utilised a silent free reading paradigm. Participants were seated in front of a screen and leaned their heads on the available forehead- and chinrest. A fixation cross appeared on the left side of the screen at a level vertical to the participants' eyes. When the experimental program detected a fixation on the fixation cross, the cross disappeared and a sentence appeared at the level of the fixation cross instead. This sentence was followed by another fixation cross on which the participants fixated after reading the sentence. When fixation was detected on the cross at the end of the sentence, the experiment program removed the sentence and terminated the trial. Participants read the sentences at their own pace and answered comprehension questions after specific sentences (24 sentences).

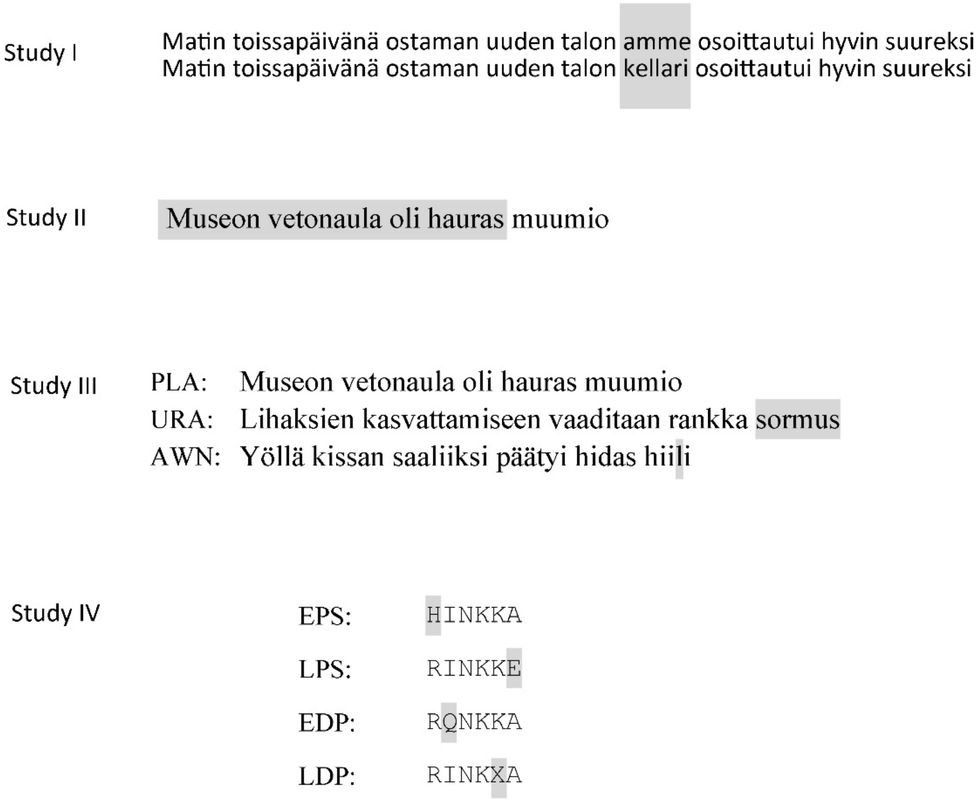
The stimuli were sentences with the target word located in the middle of it. There were 160 individual target words, which were embedded into 80 sentence frames that were presented with alternating target words, forming paired alternatives of the sentences (see Figure 7 for example sentence pairs). Sentences were presented in proportional font, Calibri, and thus the spatial width of the target words was orthogonally contrasted with the number of letters in some of the sentence pairs. Target words had eight categories that were paired in the following manner: narrow four-lettered words with wide seven-lettered words; wide four-lettered words with narrow seven lettered words; narrow and wide five-lettered words; narrow and wide six-lettered words.

Study II and **III** utilised a silent free reading paradigm. Participants were seated in front of a screen and rested their chin on a chinrest. First, a fixation point appeared on the left side of the screen and participants were instructed to fixate on it. When fixation was detected on the dot, the researcher approved the fixation and a sentence appeared at the screen at the level of the fixation dot and the trial commenced. Participants read the sentences as fast as they were able to and provided an answer after each sentence to the question "Is this sentence sensible or not?" by pressing a button. The sentence disappeared upon participant response.

The stimuli were single sentences with target words at the end of the sentence. There were three categories of stimuli: sentences with a plausible ending (100), sentences with an anomalous ending (50) and sentences ending with an anomalous target word, which was a word neighbour of a plausible word (50) (see examples in Figure 7).

In **Study IV**, a standard lexical decision task was used. Participants were seated in front of a screen and rested their chin on a chinrest. Participants fixated on a fixation dot at the centre of the screen. When fixation was detected on the fixation dot, the researcher monitoring the experiment accepted the fixation and the stimulus appeared at the centre of the screen instead of the dot. The stimulus remained on the screen for 1,500 ms. Participants were tasked to respond to the stimuli using a button box and to make a binary decision as fast as possible on whether the stimulus on the screen was a word or not.

There were 800 stimuli: 6-letter long words, non-words and pseudowords. There were 100 high-frequency words and 100 low-frequency words. Another 100 words had a uniqueness point early on in the word structure and 100 more words had a uniqueness point later on in the word structure. A uniqueness point is defined as a serial index, as letters from the beginning of the word that separate the word from all other words (Kwantes & Mewhort, 1999). Thus, a uniqueness point serves as natural control method for spatially distributed sublexical deviations. Early deviating non-words (100) were words that had their second letter replaced with a letter that is unusual in Finnish orthography (X, Q, W, etc.), while late deviating non-words had their fifth letter replaced with an unusual letter. Early deviating pseudowords (100) were generated by replacing the first letter with another letter that did not violate Finnish orthography and the same was done for the last letter of late deviating pseudowords (100). High- and low-frequency words acted as filler stimuli and were not analysed (see examples in Figure 7).



Study I

Matin toissapäivänä ostaman uuden talon amme osoittautui hyvin suureksi
Matin toissapäivänä ostaman uuden talon kellari osoittautui hyvin suureksi

Study II

Museon vetonaula oli hauras muumio

Study III

PLA: Museon vetonaula oli hauras muumio
URA: Lihaksien kasvattamiseen vaaditaan rankka sormus
AWN: Yöllä kissan saaliiksi päätyi hidas hiili

Study IV

EPS: HINKKA
LPS: RINKKE
EDP: RQNKKA
LDP: RINKXA

FIGURE 7 Examples of the stimuli in the studies. **Study I:** Example stimulus from Study I, target word highlighted. **Study II:** Example stimulus from Studies II and III; highlighted area illustrates the area from which eye movements and FRPs were extracted for Study II. **Study III:** Examples of the sentences used in Studies II and III; the last word of the sentence was the target word, which was analysed in Study III; the highlighted section shows the deviation from a plausible sentence. **Study IV:** Examples of non-lexical stimuli from Study IV; the highlighted letter is the inserted deviating letter. Acronyms: PLA = Plausible, URA = Unrelated anomalous word, AWN = Anomalous word neighbour, EPS = Early deviating pseudoword, LPS = Late deviating pseudoword, EDP = Early deviating non-word and LDP = Late deviating non-word.

4.4 Recordings and data pre-processing

In **Study I,** the eye movements were recorded from the participants' right eyes using the SMI HiSpeed eye-tracker at a 500 Hz sampling rate. Fixations and saccades were defined with a fixed threshold (40 degrees\second) algorithm. Calibration of the eye-tracker was completed with a 13 point sequence and repeated if the calibration–validation difference was higher than 0.2 degrees.

In **Studies II** and **III,** the eye movements were recorded from both eyes of the participants using the Eye-Link 1000 with a 1,000 Hz sampling rate for each eye. The EEG was recorded with a 128 channel net with AgCl electrodes (electrical geodesics) connected to the NeuroOne amplifier. The sampling rate for EEG recording was 1,000 Hz. The ET and EEG were synchronised offline, based on shared event markers. Calibration was completed with a 13 point sequence. Fixations and saccades were defined using a median velocity-based algorithm with 6 SD as the threshold (Engbert & Mergenthaler, 2006). Ocular artefacts were modelled with ICA (Makeig et al., 1996), selected with temporal covariance to saccades (Plöchl et al., 2012) and removed from the data.

In **Study II,** the FRP responses were estimated using a linear deconvolution pipeline. Fixations landing on words prior to the last word during the first pass were included in response estimation and eye movement analysis. Overlap was corrected and the influences of word length, fixation type (first fixation vs additional fixation) and saccade amplitude on brain activity were modelled from -700 ms to 500 ms in relation to the fixation onset. For comparison of intercept waveforms, the intercepts were set at a mean saccade amplitude. As word length and fixation type were found to interact in the statistical analysis of the deconvolved FRP parameters, the fixation type was separated into two deconvolution models during the post hoc analysis to interpret the interaction. A post hoc analysis of the source structure of the intercept waveform was run using the Classical Loreta Recursively Applied (CLARA) distributed source modelling algorithm to uncover the cortical generators of the response.

In **Study III,** the FRP responses were estimated using an averaging pipeline with Group-level ICA. FRPs were time-locked to the first fixation and segments -100 ms to 900 ms were extracted. The data were standardised within-subjects and then combined as a single dataset and subjected to ICA estimation. The resulting component weights were then applied to the original data segments, the components identified as artefacts or as noise components and removed, and then the responses at the IC and scalp level were averaged. Cortical sources were estimated using the cortically restricted LORETA source localisation algorithm for the remaining components that exhibited a different time course in statistical analysis.

In **Study IV,** the eye movements were recorded from both eyes of the participants using the Eye-Link 1000 with a 1,000 Hz sampling rate. The EEG was recorded with a 128 channel net with AgCl electrodes (electrical geodesics) connected to the Netstation 200 amplifier. The sampling rate for the EEG recording was 1, 000 Hz. Datasets were synchronised offline based on shared messages. Saccades were detected using a median velocity-based algorithm with 6SD as the threshold (Engbert & Mergenthaler, 2006). ERPs were estimated using a linear deconvolution, where brain activity related to saccades and behavioural responses were separated into individual parameters.

4.5 Statistical analysis

In **Study I,** eye movements (single fixation duration, gaze duration, total fixation duration, re-fixation probability, regression probability, relative landing position and skipping probability) were subjected to statistical analysis with repeated measures ANOVA. The first model had a 2 x 2 x 2 (4- vs 7-lettered words and spatial width controlled vs spatial width uncontrolled and dysfluent readers vs normal readers) factor design. The second model (for 5-lettered words) and third model (for 6-lettered words) had a 2 x 2 (narrow vs wide and normal readers vs dysfluent readers) factor structure. For all models, F1 and F2 analysis were performed.

In **Study II**, eye movements and deconvolved FRPs were subjected to statistical analysis. Two LME-models were estimated for eye movements. First models examined eye movement variables locked to the first fixation onset of a word (first fixation duration, gaze duration and re-fixation probability) with word length (5–13 letters) as a continuous factor and reading fluency group (slow readers vs typical readers) as a categorical factor. The second model analysed the fixation durations with word length as a continuous facto and the fixation type (first fixation vs additional fixation) and reading fluency as categorical factors. All models were defined with items and subjects as random factors. Deconvolved FRPs were analysed using non-parametric cluster-based permutation tests in a time window of 50–300 ms with all channels. For establishing effects within groups one-sample t-test was used as a base test, for comparing effects dependent samples t-test was used as a base test and for establishing between-group effects independent samples t-test as a base test. The sum of the base test parameters within clusters was used as a test statistic. The neighbourhood was defined using triangulation and minimum of two channels were required for establishing a cluster and 20,000 permutations were used to define the surrogate null distribution.

In **Study III**, eye movements, behavioural responses and FRPs were subjected to statistical analysis. Eye movement variables (previous fixation duration, first fixation duration, gaze duration and re-fixation probability) were analysed using LME models with three categorical predictors (plausible, unrelated anomalous, anomalous word neighbour). Items and subjects were included as random factors. Behavioural response accuracy was analysed using the Wilcoxon sign-rank test with pairwise contrasts. FRPs were analysed using non-parametric cluster-based permutation tests in a time window of 0–900 ms with pairwise contrasts. Dependent samples t-test was used as a base test and sum-t was the test statistic.

The neighbourhood was defined by a 3 cm radius and the minimum for cluster formation was 1 channel. The surrogate null distribution was defined with 10,000 permutations. Analysis of independent components was performed using the cluster-based non-parametric permutation test with the Wilcoxon sign-rank as the base test. Clusters were formed only across time and surrogate null distribution was defined with 10,000 permutations.

In **Study IV**, eye movements, reaction times and deconvolved ERP's were subjected to statistical analysis. For behavioural variables, (log-transformed small saccade onset, binary small saccade orientation to end of the stimuli and reaction time) were analysed using an LME model with a 2 x 3 (spatial location of the informative letter * stimulus type [word vs non-word vs pseudoword]) fixed factor structure. Items and subjects were introduced as random factors. Interactions were opened with pairwise comparisons. Deconvolved ERPs were analysed using non-parametric cluster-based permutation t-tests in a time window of 0–500 ms, including all channels. The neighbourhood was defined with triangulation and minimum of two channels were required for the cluster and 20,000 permutations were used to define the surrogate null distribution.

5 RESULTS

5.1 Study I: The number of letters and the spatial width of a word have separate contributions to eye movements during natural reading

Study I examined the separate influences of the number of letters and the spatial width of a word as well as whether crowding affects dysfluent readers differently from fluent readers. This was investigated by recording eye movements while participants read sentences that had target words embedded in the middle. In target words, the spatial width and the number of letters were orthogonally contrasted.

In analysis examining the number of letters and spatial width, a systematic effect of the number of letters was found in total fixation duration, gaze duration and in single fixation duration. Total fixation duration[7] is a sum of durations of the all fixations and reflects accumulated processing from first pass and regressions (Holmqvist et al., 2011). Gaze duration is the sum of all fixation durations during the first pass and reflects the processing allocated to the word during a first encounter (Holmqvist et al., 2011). Single fixation duration is the duration of fixations in which a word received only one fixation during the first pass—this variable is thought to reflect the difficulty of early word recognition processes (Kuperman & Van Dyke, 2011). All these variables had longer durations for the words with more letters. Spatial width affected the skipping probability, the refixation probability and the landing position. Skipping probability refers to the ratio of whether or not a word is fixated on during first pass reading and it is thought to be the earliest eye movement measure that reflects the parafoveal processing of a word (Inhoff & Radach, 1998). Re-fixation probability refers to the ratio of whether a word is fixated on only once or more during first pass reading and it is thought to reflect difficulties in early word recognition processes (Kuperman & Van Dyke, 2011). Landing position refers to the location within the

---

[7] This variable is listed as Total Dwell Time in Holmqvist et al. (2011).

word in which the incoming saccade lands and reflects attempts at optimising the fixation position for word recognition (Holmqvist et al., 2011). Spatially wide words were skipped less often, re-fixated more often and incoming saccades landed closer to the beginning of wide words than narrow words. However, for these three saccadic variables, the impact of the number of letters was on the interaction with spatial width. This was evident in how the number of letters had a weaker impact on skipping probability and landing position when the spatial width was controlled. Re-fixation effect was not observed when the spatial width was controlled. In summary, this illustrated that saccadic variables were more responsive to spatial width and that duration variables were exclusively responsive to the number of letters. Overall, dysfluent readers exhibited higher re-fixation probability and longer total fixation duration. There was no indication that the effect of the group would interact with the main effects.

Analyses controlling the number of letters and examining the spatial width as a factor found that, for narrow words, some of the duration measures were longer than for wide words. This was observed in single fixation duration and total fixation duration measures in five-lettered words. Total fixation duration showed this effect also in six-lettered words. Narrow words were also regressed back to more often and skipped more often than wide words. This was found for both five-lettered words and six-lettered words. Landing position also showed an effect in term of spatial width. For six-lettered words, a clear effect was observed—saccades landed more towards the word beginning in wide words than in narrow words. For five-lettered words, same was observed but the effect failed to reach a statistical significance. In summary, these analyses showed that while the spatial width of a word has a clear influence on saccadic targeting measures, some influence is exerted over the duration measures as well. Dysfluent readers exhibited overall longer gaze durations and total fixation durations for five-lettered words and there was no support for interaction between the group and spatial width.

In conclusion, the results of **Study I** demonstrated that measures that reflect saccadic targeting are mainly influenced by the spatial width of the words. The number of letters was clearly reflected in the duration measures, indicating that increasing the letter amount leads to increased processing costs. In addition, foveal crowding was found to influence the duration measures because more densely packed words increased processing costs in the form of longer duration measures. However, **Study I** found no support for the differential influence of foveal crowding in dysfluent readers, suggesting that foveal crowding is not a more severe problem during natural reading for dysfluent readers than for typical readers.

5.2 Study II: Brain activity and eye movements are different between slow readers and typical readers during natural reading

Study II investigated the word length effect on eye movements and fixation-related brain potentials in relation to reading fluency in school-aged children. This was done by recording eye movements and EEG while participants read sentences and made sensibility judgements about them. Eye movements and fixation-related brain potentials were extracted from all words prior to the last word. Eye movements and brain activity were analysed using word length as a continuous predictor and the fixation type (first fixation vs additional fixation) as a categorical predictor. Brain activity was examined from a window corresponding to the fixation duration of 50–300 ms.

In standard eye movement variables (first fixation duration, gaze duration and re-fixation probability), word length had systematic effects. These eye movement patterns are illustrated in Figure 2 of Study II. First fixations became shorter as a function of word length, gaze durations became longer as a function of word length and re-fixation probability rose as a function of word length. These effects were stronger for slow readers than for typical readers. In addition, slow readers had overall longer first fixation durations and gaze durations as well as overall higher re-fixation probability. Fixation durations were analysed using another LME model (word length x fixation type x group). The results of this model are illustrated in Figure 3 of Study II. This model revealed that word length primarily modulates the first fixation but not additional fixations. First fixations became shorter as a function of word length and this effect was stronger in slow readers than in typical readers. Slow readers had overall longer fixations.

In brain activity, word length was shown to be modulated during additional fixations instead of first fixations. These patterns are presented in Figure 7 of Study II. This effect was present as a negativity on the right occipital-parietal scalp from around 150 ms to around 250 ms. This effect was not statistically different between slow readers and typical readers, albeit the clusters did not have an identical structure or time course. Instead, the groups were differentiated by the intercept response that contains the activity common to all fixations, from around 150 ms to end of the epoch at 300 ms. This pattern is presented in Figure 4 of Study II. The difference was characterised by initial central positivity that moved to the occipital regions around 260 ms. The brain activity of typical readers exhibited the capability to differentiate first fixations from additional fixations and this was demonstrated with a positive voltage difference on the central scalp that persisted throughout the tested time window. There was no statistical support that brain activity of slow readers would differentiate between first and additional fixations.

In conclusion, the results of Study II demonstrate that slow and typical readers have different brain activity during natural reading that is early enough to affect eye movement behaviour. It is possible that the difference in brain activity reflects the foveal difficulty for slow readers in processing multiple visual units simultaneously.

5.3 Study III: Orthographic similarity of the anomalous word to plausible word delays the processing of the semantic anomaly during natural reading

Study III investigated how the orthographic similarity of an anomalous word to a plausible word influences the processing of the semantic anomaly during natural reading in school-aged children. This was done by recording eye movements and EEG while participants read sentences and made sensibility judgements about them. The last word of a sentence was a target word, which was either plausible in relation to sentence beginning, semantically anomalous to sentence beginning or semantically anomalous—but that was a word neighbour of a word that would have been plausible in relation to the sentence beginning. Eye movement variables (previous fixation duration, first fixation duration, gaze duration and re-fixation probability) were extracted from a target word and examined using LMEs. FRPs were time-locked to the first fixation on the target word and statistically examined on scalp and source level.

First fixation durations and gaze durations were longer for two categories of anomalous target words than for the plausible target words. However, there was no difference between the two anomalous target word categories. Similarly, re-fixation probability was higher for the two anomalous target word categories than for plausible target words and there was no difference between the two anomalous target word categories. Judgement accuracy for sentences ending with unrelated anomalous target words was higher than for sentences with either plausible or anomalous word neighbour target words. There was no difference in judgement accuracy between the latter two categories.

Analysis of the scalp FRP revealed that all three conditions were differentiated in brain activity. The results of the scalp statistics are illustrated in Figure 5 of Study III. The scalp difference between the anomalous condition and plausible condition was characterised by frontal negativity that persisted from around 100 ms to end of the epoch at 900 ms. There was also parietal positive difference from the plausible condition for both anomalous conditions but the time course of this positivity was not uniform for both anomalous conditions. A comparison of the unrelated anomalous and plausible conditions showed a positive difference in the parietal scalp from around 450 ms onwards, whereas in the comparison between the anomalous word neighbour and plausible conditions the parietal positive difference was observed from around 550 ms onwards. This difference in time course was complemented with the parietal positive difference between the unrelated anomalous and anomalous word neighbour conditions from around 460 ms to 700 ms.

Analysis of the independent component time courses revealed several differences between conditions. Component waveforms are presented in Figure 6 of Study III. Frontal negativity difference was attributed to a single independent component that produced more negative voltages to the frontal scalp for anomalous conditions than for plausible conditions, roughly from 200 ms to 700 ms. This component was localised to originate from the approximate area of the right frontal eye field.

Parietal positivity was attributed to three components that had parietal projections and differences between conditions. The first of these components had a right-lateralised projection to the parietal scalp and was localised at the proximity of the right angular gyrus. This component differentiated the unrelated anomalous condition from the plausible condition by projecting more positive voltage to the right parietal scalp from around 560 ms to around 810 ms.

The second of these components had a projection to the centro-parietal scalp and was localised to the bilateral superior parietal cortex and the right anterior temporal cortex. This component differentiated both anomalous conditions from the plausible one. In contrast between the unrelated anomalous and plausible conditions, more negative voltage was projected to the central scalp from around 260 ms to around 450 ms and positive voltage from around 790 ms to around 870 ms. In contrast between the anomalous word neighbour and plausible conditions more positive voltage was projected to the central scalp from around 650 ms to around 740 ms.

The third of these components had a left lateralised projection to the parietal scalp and was localised to a broad area at the left parietal and temporal cortex. The left angular gyrus resided within this broad area. This component had different activity in all three conditions. In contrast between the unrelated anomalous and plausible conditions, more negative voltage was projected to the left parietal scalp from around 250 ms to around 450 ms and more positive voltage from around 550 ms to end of the epoch. In contrast between the anomalous word neighbour and plausible conditions more negative voltage was projected to the left parietal scalp in from around 300 ms to around 540 ms and more positive voltage from around 710 ms to end of the epoch. In contrast between the unrelated anomalous and anomalous word neighbour conditions more positive voltage was projected to the left parietal scalp from around 480 ms to around 680 ms.

Also, two other components were detected to behave differently but they were not explicitly attributable to patterns observed in the scalp analysis. The first of these components had a centro-occipital projection and was localised to the occipital cortex. This component differentiated anomalous conditions from the plausible. In contrast between the unrelated anomalous and plausible conditions, more positive voltage was projected to the central occipital scalp from around 570 ms to around 770 ms. In contrast between the anomalous word neighbour and plausible conditions, more positive voltage was projected to central scalp from around 300 ms to around 160 ms and more positive voltage from around 270 ms to end of the epoch. The second of these components projected to the left temporal scalp and was localised to the middle of the left temporal cortex.

In contrast between the unrelated anomalous and plausible conditions, more negative voltage was projected to the left temporal scalp from around 340 ms to around 430 ms.

In conclusion, the results from **Study III** suggest that, when a semantically anomalous word is orthographically similar to a plausible word, the processing of the semantic anomaly becomes delayed. Yet, from the frontal pattern and eye movements, it is obvious that the semantic anomaly is detected early at some level of processing.

5.4 Study IV: The type and location of sublexical deviations from real words affect the processing of meaningless letter strings

Study IV investigated how the type and spatial location of a sublexical deviation inserted into an existing word influences brain activity, small saccade performance and reaction times. This was studied through a lexical decision task and concurrent ET-EEG recording. Six conditions were studied: early and late deviating non-words, where in an existing word a letter was replaced with a letter that is not common in Finnish orthography, thus generating a clear violation of Finnish orthography; early and late deviating pseudowords, where the replacement did not violate Finnish orthography but the pronounceable letter string was void of meaning. Words with early and late uniqueness points were included as natural control conditions.

The type of stimuli had multiple effects on brain activity and behaviour. In reaction times, non-words were responded to fastest and pseudowords slowest. Reaction time performance for the words was between the two not-lexicalised conditions. Behavioural effects are illustrated in Figure 3 of Study IV. Stimulus type had a systematic effect on brain activity in the later part of the tested window (from around 300 to 500 ms). Non-words in relation to words exhibited short frontal negativity and longer central positivity with slight right lateralisation. Pseudowords in relation to words exhibited frontal negativity and occipital positivity. Pseudowords in relation to non-words exhibited central negativity. These effects are illustrated in Figure 4 of Study IV.

There was no indication of the influence of the orthographical uniqueness point within words in any of the measured variables. In non-lexical conditions, the spatial location of the deviation had an influence on eye movement behaviour. When the deviation was late in the stimuli, small saccades were executed earlier than when the deviation was early. Location of the deviation also influenced probability to make the small saccade towards the end of the stimuli. If the deviation was at the end, the direction of the small saccade was much more probable towards the end of a word than when the deviation was at the beginning of the word. In brain activity, only the location of the deviation in non-words had support for different processing. Late deviating non-words in relation to early deviating non-words had a more positive response in the right occipital-parietal scalp from around 200 ms to around 300 ms. From around 400 ms to 500 ms, there was a central negativity and surrounding positivity at fringe channels. Pairwise, contrasting tests showed that the early part of the brain activity difference was generated by early deviating non-words having a more negative response in the right occipital-parietal scalp. These ERP effects are presented in Figure 5. Reaction times exhibited modulation by the spatial location of the deviation only for pseudowords. LPS received later responses than EPS.

In conclusion, **Study IV** found evidence in terms of reaction times and brain activity that a certain type of deviation from normal words alters the processing of letter strings. The spatial location of the deviation, on the other hand, had more mixed results, with eye movements being systematically responsive to manipulation and brain activity modulation being found only for non-words. Overall, this suggests that word recognition is particularly sensitive to abnormal letter combinations at the beginning of a word.

6 DISCUSSION

6.1 Study I

Study I aimed to separate the influences of the number of letters and spatial width of at word on eye movements and to see whether visual crowding, as it is present naturally in proportional fonts, affects dysfluent readers differently from fluent readers. In order to examine this, a sentence reading experiment in which the participants' eye movements were monitored with an eye-tracker was conducted. Eye movements in relation to target words were examined. The target words were embedded in the middle of the sentences and had spatial width and number of letters that were orthogonally contrasted.

The most substantial findings of **Study I** were that saccadic aspects of the eye movements to words are governed by the spatial width of the word and that duration measures were generally responsive to the number of letters. However, dense words showed an increase in duration measures in comparison to sparse words, which indicated that foveal crowding is a factor during natural reading. Crowding did not affect dysfluent readers and controls differently but dysfluent had overall longer fixation durations.

Crowding effects were clearly observable in five-lettered words and somewhat present in six-lettered words. Crowding effects being clearer in five-lettered words may be due to the fact that narrow five-lettered words had multiple instances of words that contained letters i and l which are highly similar and might therefore increase perceptual challenges. Nevertheless, these findings are in line with literature that recognises crowding as a considerable factor during reading (Martelli, Di Filippo, Spinelli, & Zoccolotti, 2009; Pelli et al., 2007; Perea & Gomez, 2012; Slattery & Rayner, 2013; Zorzi et al., 2012). A particularly noteworthy finding is that densely packed words are regressed to more often, which is indicative of crowding disrupting word recognition during the first pass in natural reading. Again, this is in line with previous research on the impact of decreased letter spacing (Rayner, Slattery, & Bélanger, 2010; Slattery & Rayner, 2013).

The number of letters had a consistent effect on the duration measures, as durations increased when the number of letters increased. This effect, however, was not due crowding because it manifested as the main effect and not as an interaction with controlling of spatial width. Similarly, in this model, no effects on regression probability were observed and, thus, there was no indication of disrupted word recognition due to the number of letters. Instead, the number of letters can be considered to increase the processing costs of a word, which is a view that is compatible with previous research (Hautala et al., 2011).

Spatial width of the words had a clear influence on the saccadic aspects of eye movements. Wide words were skipped less often and saccades were directed more towards the words beginning in wide words than in narrow words. Overall, these findings are in line with the view that saccadic targeting during reading is more dependent on spatial characteristics of words than linguistic aspects of the words (Inhoff, Radach, Eiter, & Juhasz, 2003).

Reading fluency manifested in overall longer duration measures. Previous studies have found that the word length effect is stronger for dyslexics (Hawelka et al., 2010). In addition, the stronger word length effect for slow readers was found in **Study II**. However, in this study, such influence of reading fluency was not observed. It may well be that the subjects in **Study I** do not have a severe enough deficit for the word length effect to be modulated here or they have managed to develop strategies to compensate for this deficit. The overall longer fixation durations for individuals with weaker reading skill has been found in previous studies (De Luca et al., 2002, 1999; Dürrwächter et al., 2010; Hawelka et al., 2010).

In conclusion, **Study I** provides support for the view that spatial width of a word and the number of letters in a word have separate influences on eye movements during reading. However, neither the crowding effect nor the word length effect was associated with reading fluency.

6.2 Study II

Study II examined how word length effect co-occurrs in eye movements and fixation-related potentials and whether reading fluency has relevant influence on word length effect in school-aged children. Participants read sentences and provided sensibility judgements while ET-EEG data was recorded. Fixation-related potentials and corresponding eye movements were extracted from the first pass reading, excluding the final word of a sentence. Influence of word length was analysed as a continuous factor, accompanied with categorical analysis of fixation type (first fixation vs additional fixation). Reading fluency was examined as a group effect (slow vs typical readers).

From the results of **Study II,** it is possible to establish that slow readers and typical readers have different brain activity within fixation duration. Slow readers also had overall longer fixation and faze durations than typical readers, as well as overall higher re-fixation probability. Another substantial finding was that word length influences brain activity during additional fixations and not first fixations. Word length effect on eye movements was different between the groups, yet there was no difference between the groups in how the word length effect manifested on brain activity.

Reading fluency had several clear influences on all eye movement measures. First fixation duration, fixations overall and gaze durations were longer for slow readers than for typical readers, which is in line with previous research (De Luca et al., 2002, 1999; Dürrwächter et al., 2010; Hawelka et al., 2010) Also, the between-group difference in intercept response was observed simultaneously with the overall difference in eye movements. In the FRP estimation with linear deconvolution, the intercept response reflected activity that is common to all the fixations, thus finding a difference in intercept response suggests that overall visual processing during natural reading is different between the groups. Even though statistical results point to central positivity, interpreting the waveform suggests that we actually observe stronger occipital N1 response in slow readers and later occipital positivity. Stronger N1 response is in line with previous RSVP ERP research with dyslexic children (Fraga González et al., 2016, 2014).

There was considerable support for the notion that cognitive processing is different between first and additional fixations. In eye movements, word length influenced the first fixations and not additional fixations. In the brain activity, this was the opposite. The brain activity of typical readers showed the main effect for fixation type, yet there was not sufficient support for a direct main effect of fixation type on brain activity for slow readers. The effect for typical readers persisted through the tested time window, characterised by central positivity, but, since the effect persists for the whole time window and is not clearly associated to known reading-related activity, **Study II** refrains from making conclusive claims on the nature of this effect. Rather, additional studies are advised to further examine this issue.

It should be remembered that word length is one of the main determinants of eye movements during reading (Kliegl et al., 2004; Rayner, 1997) and that it has been assumed to have a very early influence on the preparation of eye movements (Reichle et al., 2011; Sereno, Rayner, & Posner, 1998). In **Study II**, word length has a clear impact on eye movements and on brain activity. First fixations became shorter as a function of word length and gaze durations became longer and re-fixation probability became higher. First fixation duration and re-fixation probability thus indicate that word length influences cognition within the first fixation. The brain activity implicates otherwise. There, the influence of word length was implicated to occur during the additional fixations and this, in turn, is in direct confrontation with the assumptions of the early influence of word length. However, it is not entirely clear whether the brain activity effect from word length reflects the linguistic/number of letters aspect of the word length but could, in fact, be a reflection of the words spatial width increase. Spatial width and number of letters have been found to have separate influences on eye movements in previous research (Hautala et al., 2011) and in **Study I**.

In eye movements, a stronger word length effect for slow readers than typical readers was found. This is a conceptual replication of previous research (Hawelka et al., 2010). As the influence of word length does not differ between groups in brain activity, it is possible that significantly stronger word length effect exhibited by slow readers in gaze duration is due to the accumulation of multiple fixations from different overall processing.

Overall, **Study II** provides the first proof of electrophysiological brain activity during natural reading between school-aged slow readers and typical readers being different. In addition, the word length effect is found in the brain activity associated with additional fixations but not for first fixations during the time window comparable to fixation durations, which is in contradiction with strict interpretations of the eye-mind link that assume that gaze location and what is under cognitive processing are tightly aligned.

6.3 Study III

Study III examined the influence of orthographic similarity to processing of semantic anomalies during natural reading in school-aged children. The children read sentences while ET-EEG data was recorded and provided sensibility judgements. The last word of a sentence was the target word and it was either plausible or anomalous in relation to the preceding context. Semantically, anomalous words were either unrelated or orthographic word neighbours to a plausible word completion of the sentence. Eye movements, scalp level FRPs and source level FRPs related to the target word were analysed.

The main finding of **Study III** was that, when the semantically anomalous word is orthographically similar to the plausible word, the processing of the semantic anomaly is processed later than when there is no orthographic similarity to the plausible word. The semantic anomaly was detected early on. This was evident in the eye movement effects of first fixation duration and re-fixation probability. The timing of these effects points that the anomaly is detected prior to 400 ms as re-fixation probability and first fixation duration are modulated prior to that point. Also, gaze duration reflected the detection of the anomaly. These findings are well in line with previous research, where semantic features have been found to modulate gaze behaviour during first-pass reading (Abbott & Staub, 2015; Matsuki et al., 2011; Veldre & Andrews, 2016, 2017). This early detection of the anomaly was reflected also in the frontal negativity, which was present prior to the execution of the eye movements. Independent component corresponding to the frontal negativity was localised to the approximate location of right frontal eye field. This brain area is heavily involved in oculomotor control and spatial attention (Godoy-Fernandez, Lüders, Dinner, Morris, & Wyllie, 1990; Schall, 2004; Vernet, Quentin, Chanes, Mitsumasu, & Valero-Cabré, 2014), which suggests that frontal negativity difference is not a direct correlate of semantic processing but rather reflects the preparation and execution of oculomotor behaviour that is observed between conditions to be different.

The statistical differences in the scalp waveform between anomalous conditions and plausible ones were reminiscent of the P600 modulation. P600 has been associated with the integration of semantic meaning and responsive to clear semantic anomalies (DeLong et al., 2014; Friederici & Weissenborn, 2007) and reflect conscious processing of the semantic anomaly (Sanford et al., 2011). In contrast between the unrelated anomalous word and plausible word, the P600 can be considered to be present from around 450 ms onwards and in the contrast between the anomalous word neighbour and plausible conditions from 550 ms onwards. In addition, there was positive parietal difference between the unrelated anomalous and anomalous word neighbour conditions around 460 ms to around 700 ms. This pattern suggests that the semantic anomaly arrives at the integrative/conscious processing stage earlier for an unrelated anomalous word than it does for an anomalous word neighbour. Analysis of the source structure and time course of underlying components provided further support for this interpretation. Component projecting to the left parietal scalp was localised to cover an area of the left angular gyrus and the posterior temporal lobe. These structures have been implicated to heavily contribute to semantic processing in previous research. The activity in this component that differentiates the unrelated anomalous and plausible conditions occurs earlier than the activity that differentiates the word neighbour anomalous and plausible conditions. Furthermore, the unrelated anomalous and anomalous word neighbour conditions were differentiated in the activity of this component. Location of the source on the cortex and the time behaviour of the component support the interpretation that the P600 activity is delayed for the anomalous word neighbour condition.

Overall, **Study III** illustrates that semantic anomalies are detected fairly early. This is shown through a pattern of eye movements and brain activity. In addition, **Study III** shows that when there are shared visual features in the semantic anomaly and plausible content, the semantic anomaly is not as feasibly processed as it is when the semantic anomaly does not contain a word form similarity.

6.4 Study IV

Study IV examined how spatial location and type of sublexical deviation inserted in to a word influences the co-occurring small saccade behaviour, brain activity and reaction time performance. For this end, participants performed lexical decision task while ET-EEG data was recorded. The experiment had six conditions: early and late deviating non-words, where letter replacement created a violation of Finnish orthography; early and late pseudowords, where letter replacement preserved Finnish orthography. Words with early and late uniqueness points served as natural controls for the non-words and pseudowords. ERPs, small saccade characteristics and reaction times were analysed.

Main findings of **Study IV** were that the types of deviation from existing words had different impacts on brain activity and reaction time performance. Location of the deviation was found to have an influence on the non-words in brain activity and small saccade performance. For pseudowords, spatial location of the deviation influenced the reaction time performance and small saccade behaviour.

Based on the reaction time performance, detecting non-words as not words was the easiest task for this stimulus type, as the reaction times were the fastest. This finding is in line with previous research, where it was found that when non-words are dissimilar from real words, they are dispatched more easily (Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015) and inserting Xs, Qs, etc, into Finnish words makes them quite dissimilar from real words. It may well be that the dissimilarity acts as a salient cue for dismissing the letter string as not being a word. Brain activity pattern supports this, as there is central positivity that potentially implicates a lack of semantic processing. Conversely, pseudowords had the longest reaction times. Again this is in line with previous research in which non-lexical stimuli similar to lexical stimuli caused the reaction times to be longer than for dissimilar stimuli (Yap et al., 2015). In brain activity, fairly long-lasting frontal negativity is observed for pseudowords. It might be that this negativity is a reflection of the N400 (Kutas & Federmeier, 2011) family of components and, thus, because the task in Study IV lacks semantic context, the observed pattern would reflect attempts to access the semantics of the letter string. A prolonged search for semantic content would, in turn, lead to the longest reaction times. The contrast between pseudowords and non-words had a clear N400 difference, where pseudowords had stronger N400. Therefore, it can be argued that, for non-words, a very limited attempt or no attempt at all is made to access semantics.

The spatial location of the deviations had systematic effects in the non-lexical conditions. First of all, the probability of the first saccade towards the right was modulated in such a manner that, in practical terms, the saccade was most often launched towards the deviation. As such, the saccade orientation behaviour can be considered to be aiming to optimise foveal location in a manner that is most beneficial for the task at hand, as has been found in previous research (Kagan & Hafed, 2013). In addition, the spatial location of the deviation had an influence on the latency of the first saccade. The latency of the first saccade was shorter when the deviation was at the end. Now, this is opposite to what was found in previous research—but it must be noted that Study IV allowed for a much wider range of saccade latencies than previous research (Hautala & Parviainen, 2014). The latency effect in small saccades does not have the clearest of patterns; however, it must be remembered that the probability for making saccade towards the right was above the chance level in the word stimulus type. Thus, when the deviation is at the end, the prepared saccade can be executed as deviation is detected. However, when the deviation is at the beginning, the more probable orientation of the saccade towards the end needs to be reprogrammed towards the beginning and this inhibits execution.

Early deviating non-word has a distinct pattern of activity. It was characterised by the more negative response in relation to early deviating pseudowords,

words with early uniqueness point and late deviating non-words on the right occipital scalp. This activity was reminiscent of the N2pc response, a response that has been associated with the attentional shift to the contralateral hemifield in earlier attention studies (Hickey, Mcdonald, & Theeuwes, 2006). Recently, a view, according to which N2pc reflects the transient enhancement of processing at the locus of spatial attention, has been introduced (Zivony, Allon, Luria, & Lamy, 2018) and this enhancement has been suggested to support high-level processes, such as identification. The lack of similar modulation for late deviating non-words is noteworthy because it suggests that the mechanism involved is not symmetric, which, coupled with the saccade orientation effect, suggests that the early deviating non-words receive enhanced processing. It seems that the brain places particular emphasis on the processing of clear violations in the letter string beginning.

In addition, there was reaction time modulation by the spatial location of the deviation in pseudowords. Late deviating pseudowords took longer to classify as not being a word than early deviating pseudowords took. Similarly, the early deviating pseudowords and late deviating pseudowords were differentiated by small saccade performance in a way that would optimise the foveal position to the location of the deviation. Yet, there is no support for different processing in the brain activity found. It may well be that the process that leads to the reaction time difference is triggered after failing to access the semantics of the word. If this process is a serial letter-by-letter decoding process from left to right, then the reaction time results are plausible.

Overall, **Study IV** illustrates that the dynamics of behavioural responses, brain activity and oculomotor behaviour are altered during lexical decision by the type and location of deviations inserted into words. It would seem that if the violation is strong, then no attempt to access the semantic meaning of the letter string is made and the decision is made based on visual cues. When the deviation is not distinct but leaves the letter string very similar to real words, then the decision takes much longer, potentially due to attempts made to assign meaning to the letter string. Simultaneously, the oculomotor system optimises the foveal position on the deviation to optimise the processing.

6.5 General discussion on the common patterns in the studies

A common theme in the studies was that they all examined the sublexical features of the single words and their impact on eye movements (**Study I–IV**) and brain activity (**Study II–IV**). **Study I** manipulated two aspects of word length in a target word: spatial width and number of letters. **Study II** examined the impact of word length in a large number of words during the first reading pass. **Study III** examined the processing of semantic anomalies and impact of orthographic similarity. **Study IV** examined the impact of sublexical deviations from word form. **Studies I** and **II** also examined the impact of reading fluency on the effects.

If only behaviour is considered, then the studies provide a fairly simple pattern of results. **Studies I** and **II** replicated the typical finding that word length influences gaze duration. Words with more letters received longer gaze durations than words with fewer letters. This is thought to be indicative of larger processing demands. Similarly, both studies show that individuals with weaker reading ability exhibit longer duration measures. However, **Study I** does not exhibit stronger word length effects for the dysfluent readers, while **Study II** find that the influence of word length on eye movements is stronger for slow readers. This discrepancy might be related to the age difference between the participant groups in **Studies I** and **II**. This could imply that word length is a problematic factor for young readers with reading difficulties but not for adults with reading difficulties. This is, unfortunately, a tedious suggestion, as there are other noteworthy differences between the studies that could explain the discrepancy. In **Study I** there were fewer participants and eye movement variables were drawn from a smaller number of words than in **Study II**. In addition, **Study I** utilised repeated measures ANOVA and not LME models like **Study II**. All of these make **Study I** weaker in terms of statistical power and thus less probable to observe such effect.

Studies III and **IV** both included contrasts in which the differences between categories was established by one letter. In **Study III**, the anomalous word neighbour target word differed by one letter from its plausible alternative. In **Study IV,** non-words and pseudowords differed by only one letter from the real words. Overall, the results from **Studies III** and **IV** show that a single letter is sufficient to cause different eye movements and brain activity. In itself, this is an obvious statement because, in many cases, switching one letter changes a word to a different word. However, **Study III** shows that, in the presence of a plausible alternative, one letter difference influences the semantic processing of the word or even the whole sentence. Similarly, in **Study IV**, close proximity between the real word and the pseudoword or an absence of visual cues that is present in nonwords influence the semantic processing. Based on these two studies, it can be claimed that one letter is enough to cause increased processing costs in semantic processing if that is letter capable of confounding the identity of the word.

Overall, the manipulations were systematically reflected in the eye movement records—in many cases with substantial impact on behaviour. In some cases, eye movement effects were preceded by brain effects that could, by virtue of temporal order, be causally linked to the eye movement effects. For example, this was the case with the fixation duration, re-fixation probability modulations and frontal negativity in **Study III**, as the frontal negativity effect precedes the eye movement effect. Yet, often the eye movement effect was not accompanied by preceding brain activity effect. Most notable of these instances can be considered to be found in **Study II**, where the first fixations and the re-fixation probability were found to be modulated by word length but no accompanying word length effect in the brain activity was found corresponding to the first fixation. It, however, must be noted that the pattern of found eye movement effect and no effect on the fixation duration corresponding part of the FRP is not unique in FRP literature. For example, recent FRP studies on natural reading have contrasted high- and low-lexical frequency words (Degno et al., 2019; Kretzschmar et al., 2015) and found typical frequency effects in eye movements (longer duration measures) but no FRP effects. It should be remembered that lexical frequency is one of the main determinants for eye movements of reading and assumed to influence cognitive processes within fixation duration (Sereno & Rayner, 2003).

The FRP effects outside fixation confines have implications on the eye-mind link that either this thesis or the existing FRP literature is currently unable to specify with certainty because it is, in principle, reliant on the overall interpretation of FRP perturbations and what processes are reflected by the component. It could be that the FRP difference that we observed is the first time point at which the cognitive process differs between conditions. It may also be that the modulation of FRP does not reflect the first point of different processing—it has, for example, been suggested that late ERPs, like N400, reflect the feedback processes from earlier processing (Sereno & Rayner, 2003). The first sweep of the information propagation across neural networks is then proposed to occur in the early time window of C1 response, around 70 ms (Foxe & Simpson, 2002). Also, it is important to realise that FRPs are not the only possible EEG signals that could fit into the fixation time window. It may well be that the critical phenomena for modulation of eye movements are manifested, for example, as time-frequency perturbations or as connectivity patterns. Considering these alternative points of view would place commonly observed FRP modulations under the same interpretation as eye movements as a reaction to some previously occurred process. This mindset would explain some of the co-occurring FRP and eye movement modulations, where FRP modulations do not precede the eye movement modulations and do not contradict most of the interpretations of the eye–mind link.

In conclusion, **Studies II**, **III** and **IV** extend the knowledge about the co-occurring eye movement behaviour modulations and electrophysiological brain activity modulations. The findings of this thesis show that word length and one letter switches have concurrent influence on brain activity and eye movements. However, as complementing literature utilising FRP as a methodology is, at the moment, quite limited, it is premature to make far-reaching conclusions on the nature of this co-occurrence. In the future, along with substantial additional basic research, the results have a chance to contribute to a comprehensive understanding of the eye–mind link during natural reading.

6.6 Future directions

For the future of co-registered EEG and eye-tracking studies, two important lines of development can be distinguished: methodological development and large-scale studies.

The chapter on the methodology of co-registered ET-EEG closed on the notion that current implementation, while solving the most severe problems, is not yet optimal. This is due to the fact that current methodology does not allow for a simultaneous correction of the temporal overlap and the preservation of the single observation structure—but forces the analysis to be performed with participant-level estimates. However, there are no mathematical restrictions that would prevent LME models from being adapted to perform deconvolution (Ehinger & Dimigen, 2018). However, currently, there are practical restrictions to this approach, as such LME models are fitted simultaneously for all subjects and observations. In LME models for eye movements, this poses no issues but, for a combination of continuous datasets of EEG from 30 or so subjects, this would mean massive computational costs. Thus, the hypothetical deconvolution LME modelling would require access to sufficient computational hardware.

The introduction of GLM to the ET-EEG methodology allows multiple variables to act as predicting factors for FRP estimation. Thus, the field of ET-EEG reading research would benefit from large-scale studies in which the properties of stimulus materials are very well detailed instead of rigidly controlled. As a general example, this would comprise single sentences or multi-line texts in which, for example, the word-by-word predictability, the number of letters, syllables or morphemes, the spatial width of the words and the frequency are estimated prior to data collection. The resulting dataset could then be used for mapping the influence of eye movements, psycholinguistic factors and visual factors of words on brain activity. In addition, if published along with the guidelines of open science, such datasets could prove very valuable for the field of ET-EEG research. Eye movement research has several successful examples of this corpus-based approach (Cop, Dirix, Drieghe, & Duyck, 2017; Kliegl et al., 2004; Kliegl, Nuthmann, & Engbert, 2006; Luke & Christianson, 2018) and it is reasonable to assume that the ET-EEG corpus would be similarly beneficial.

In addition to these two critical lines of development, other interesting applications arise. Implementation of source analyses, in a larger extent to which it is present in this thesis, to ET-EEG data is a promising approach that would allow for mapping individual contributions of brain areas to natural reading. If source analysis proves successful, then this, in turn, opens the possibility of estimating the interactions between brain areas during natural reading with connectivity analyses.

6.7 Methodological limitations in the studies

Overall, the studies comprising this thesis had some methodological limitations. Some are specific to certain studies and others are general limitations arising from the chosen method or general limitations inherent to the field of research.

Studies had specific limitations. **Study I** used ANOVA instead of LME as its statistical solution—switching to LME models would have solved the F1 and F2 analysis discrepancy that is observed for some of the analysed variables. **Study II** was an exploratory study rather than a true experiment and, as such, there is no absolute certainty on whether there are confounding factors present.

Study III made explicit claims on the time course of the differences between conditions; thus, it must be noted that the statistical test used is not a generally suitable test for making such claims because non-parametric permutation tests do not correct individual observations of significant difference but control the familywise error rate at the cluster level (Sassenhagen & Draschkow, 2019). That being said, the timing differences that **Study III** argues for are on the scale of 50 or more milliseconds, which can be considered substantial. Thus, even though basing inference on specific observations within the cluster is not an optimal practise, due the scale of the difference in timing in the data, the observed pattern should be preserved if studied with statistics that are designed to estimate the onsets of differences. One potential approach for such analysis would be the so-called "jackknife" procedure (Kiesel, Miller, Jolicœur, & Brisson, 2008), which is a bootstrapping technique for establishing ERP latency distributions. Furthermore, **Study III** does not implement a direct overlap correction and yet, based on the component waveforms, it is fair to assume that the effect of the overlap is minor (see Discussion of limitations section in Study III).

A couple of general limitations can be identified. Participants were all native Finnish speakers and, thus, it is unclear how the findings would generalise to other languages. Participant populations were children in **Studies II** and **III**, therefore it is unclear to what extent the findings from these two studies generalise to adults and other groups.

Use of the Hydrocel caps might not be an optimal solution for ET-EEG research, as ocular artefacts need to be corrected using computational methods. Even though using saline-soaked sponges as a conduction method for the electrodes has obvious benefit when working with children, in terms of the set up speed, the relatively fast evaporation rate of the fluid from the sensors and resulting changes in impedance are problematic for accurate ICA estimation. This is because the ICA relies on spatial stationarity through time as a core assumption (Lee et al., 1999; Makeig et al., 1996) and, if conduction properties of an electrode change, this assumption is violated. In practice, this leads to split but highly similar ocular components, especially in situations in which contact is lost altogether. This, in turn, can affect the performance of the temporal covariance algorithm that is used in the classification of the components because only one of the split components is active during the saccade and, thus, dependency might not trigger the threshold.

Also, the limited amount of literature in the field of co-registered ET-EEG places a constraint on what can be generalised based on this thesis. Many of the implications of the findings presented are not clear because the field of co-registered EEG and eye-tracking is still in its infancy. To provide an example, this means that findings are difficult to fully relate to the existing body of work because the difference in the naturalistic free reading protocol vs RSVP is not fully mapped for many manipulations and there might be unknown interactions. Thus, the full impact of the pioneering work that constitutes this thesis will be realised when the field itself is more complete.

As with all scientific knowledge, the findings presented in this thesis would find support through replication, which can be considered to be the true test of the validity of any finding.

YHTEENVETO (SUMMARY)

Okulomotorisen käyttäytymisen ja elektrokortikaalisen aivoaktiivisuuden yhteisesiintyminen luonnollisen lukemisen ja sanantunnistamisen aikana

Luonnollisen lukemisen aikana katseen sijainti kohdistuu nopeiden nykäyksien ja pysähdyksien vuorottelevana sarjana luetun tekstimateriaalin ylitse. Näitä nykäyksiä kutsutaan sakkadeiksi ja pysähdyksiä fiksaatioiksi. Lukutaidon taso ja luetun tekstin ominaisuudet muokkaavat fiksaatioiden ja sakkadien ominaisuuksia. Esimerkiksi hitailla lukijoilla on taipumus tehdä pidempiä fiksaatioita kuin keskitasoiset lukijat tekevät. Luonnollisen lukemisen aikaista sähköistä aivotoimintaa ei kuitenkaan juurikaan tunneta. Tämä johtuu perinteisen herätevastetutkimuksen rajoitteista.

Herätevastetutkimuksen rajoitteet on kuitenkin mahdollista kiertää yksinkertaisella silmänliikekameran ja aivosähkökäyrän yhdistelmällä. Kun perinteisessä herätevastetutkimuksessa herätevaste aikalukitaan koeohjelman tapahtumaan, rinnakkaisrekisteröidyllä katseenseuranta-aivosähkökäyrällä (Co-registered Eye tracking Electroencephalography) on mahdollista käyttää valikoituja silmänliiketapahtumia, esimerkiksi fiksaatioita tiettyihin sanoihin, vasteen laskemisen kiintopisteenä. Fiksaatiovastemenetelmä tuo kuitenkin mukanaan analyyttisen tason haasteita. Silmänliikkeistä syntyvät sähköiset artefaktat ovat aikalukossa fiksaation alkuun ja silmänliikkeiden ominaisuudet, kuten esimerkiksi sakkadin pituus, ovat riippuvuussuhteessa aivoaktiivisuuteen. Tämän lisäksi fiksaatioiden keskinäisen ajallisen läheisyyden vuoksi kunkin fiksaation sähköinen aivoaktiivisuus päällekkäistyy rinnakkaisten fiksaatioiden aivoaktiivisuuden kanssa. Nämä haasteet on kuitenkin mahdollista ratkaista tarkoitukseen sopivilla menetelmillä, kuten riippumattomien komponenttien analyysillä ja yleistettyä lineaarista mallia soveltavalla lineaari-dekonvoluutiolla. Rinnakkaisrekisteröity katseenseuranta-aivosähkökäyrä ja erityisesti fiksaatiovastemenetelmä, jota käytin väitöskirjassani laajalti, mahdollistaa aivoaktiivisuuden tutkimisen luonnollisen lukemisen aikana.

Väitöskirjan ensimmäisessä osatutkimuksessa tarkasteltiin sanan kirjainmäärän ja sanan leveyden vaikutuksia silmänliikkeisiin. Myös lukutaidon yhteys näihin vaikutuksiin oli tutkimuksen kohteena. Tutkimuksenkoeasetelmassakoehenkilöt lukivat lauseita, joiden keskellä oli kohdesanoja, joiden leveys ja sisältämä kirjainmäärä oli ortogonaalisesti kontrastoitu. Aikaisemmissa tutkimuksissa on havaittu, että sanan kirjainmäärällä ja leveydellä on erilaiset vaikutukset, mutta näiden vaikutuksien yhteyksiä lukutaidon tasoon ei ole aiemmin tarkasteltu. Osatutkimuksessa toistettiin havainto, että silmänliikkeiden ajalliset aspektit (katselunkeston muuttujat) muokkaantuivat ensisijaisesti suhteessa sanan kirjainmäärään ja silmänliikkeiden spatiaaliset aspektit (sakkadien kohdentamiseen liittyvät muuttujat) suhteessa sanan leveyteen. Näillä vaikutuksilla ei kuitenkaan ollut yhdysvaikutusta lukutaidon tasoon. Sen sijaan hitailla lukijoilla oli keskimäärin suuremmat arvot katselunkeston muuttujissa. Ensimmäinen osatutkimus antaa tukea näkemykselle, että sanan kirjainmäärällä ja leveydellä on itsenäinen kontribuutio katselukäyttäytymiseen lukemisen aikana ja sitä kautta itsenäinen kontribuutio lukemisen kognitiivisiin prosesseihin. Tutkimuksen tulokset myös viittaavat siihen, että sanan kirjainmäärä tai sanan leveys eivät ole ongelmallisia piirteitä heikomman lukutaidon omaaville aikuisille lukijoille.

Väitöskirjan toisessa osatutkimuksessa tarkasteltiin sanan pituuden vaikutusta silmänliikkeisiin ja fiksaatiovasteisiin, sekä näiden vaikutuksien yhteisesiintymistä kouluikäisillä lapsilla. Lisäksi tarkasteltiin lukutaidon assosiaatiota näihin ilmiöihin. Tätä tutkittiin koeasetelmalla, jossa koehenkilöt lukivat lauseita ja antoivat arvioita lauseiden järkevyydestä sillä välin, kun heiltä mitattiin rinnakkaisrekisteröityä katseenseuranta-aivosähkökäyrää. Tutkimuksessa havaittiin, että sanan pituus vaikutti voimakkaammin hitaiden lukijoiden kuin tyypillisten lukijoiden silmänliikkeisiin. Sanan pituus vaikutti ensimmäisten fiksaatioiden kestoon mutta ei ensimmäistä fiksaatiota seuraavien fiksaatioiden kestoon. Fiksaatiovasteissa tämä toteutui päinvastoin, eli ensimmäiseen fiksaatioon liittyvä aivoaktiivisuus ei muokkaantunut suhteessa sanan pituuteen. Sen sijaan ensimmäistä seuraavien fiksaatioiden aivoaktiivisuus muokkaantui sanan pituuden funktiona. Siinä, miten sanan pituus vaikutti aivoaktiivisuuteen, ei ollut eroa hitaiden ja tyypillisten lukijoiden välillä. Sen sijaan kaikkien fiksaatioiden jakama aivoaktiivisuus oli erilaista ryhmien välillä. Löydöksien kokonaisuus on yllättävä, sillä aiemman tutkimuksen perusteella on oletettu, että sanan pituus vaikuttaisi voimakkaammin heikon lukutaidon omaaviin ja tarkasteltaessa pelkästään silmänliikemuuttujia tämä johtopäätös on oikeutettu. Fiksaatioihin lukittua aivoaktiivisuutta tarkasteltaessa näyttää kuitenkin siltä, että sanan pituus itsessään vaikuttaa samankaltaisesti molempiin ryhmiin, mutta aktiivisuuden perustaso on erilainen ryhmien välillä. Näyttää siis siltä, että heikkolukutaitoiset kykenevät prosessoimaan vähemmän kirjaininformaatiota kullakin fiksaatiolla ja tämä kenties kasaantuu suuremmaksi sanan pituusvaikutukseksi silmänliikkeissä.

Väitöskirjan kolmannessa osatutkimuksessa tarkasteltiin sitä, miten semanttisten anomalioiden havaitseminen heijastuu silmänliikkeisiin ja fiksaatiovasteisiin sekä näiden vaikutuksien yhteisensiintymistä kouluikäisillä lapsilla. Tätä tutkittiin koeasetelmalla, jossa koehenkilöt lukivat lauseita ja antoivat arvioita lauseiden järkevyydestä sillä välin, kun heiltä mitattiin rinnakkaisrekisteröityä katseenseuranta-aivosähkökäyrää. Lauseiden viimeistä sanaa manipuloitiin siten, että se oli joko uskottava päätös lauseelle, semanttinen anomalia tai uskottavan lauseen päätöksen sananaapuri ja samanaikaisesti semanttinen anomalia. Osatutkimuksessa havaittiin, että kirjaimen poikkeama uskottavasta lauseesta on riittävä aiheuttamaan semanttisen anomalian prosessointiin liittyviä yhteisesiintyviä silmänliiketapahtumia ja aivoaktiivisuutta. Tutkimuksessa havaittiin myös, että lauseen päättyessä semanttiseen anomaliaan, joka poikkeaa uskottavasta lauseen päätöksestä vain yhdellä kirjaimella, semanttisen anomalian prosessointia heijastava aivoaktiivisuus oli viivästynyttä suhteessa tilanteeseen, jossa koko sana poikkesi uskottavasta lauseesta. Tämä havaintokokonaisuus viittaa siihen, että mikäli semanttisesti poikkeava sana on visuaalisesti samankaltainen uskottavan sanan kanssa, semanttisen anomalian prosessointi on työläämpää, mikä mahdollisesti johtuu vaihtoehtoisten merkitysrepresentaatioiden kilpailusta.

Neljännessä osatutkimuksessa tarkasteltiin sitä, miten oikeista sanoista johdettujen epä- ja pseudosanojen prosessointi heijastuu mikrosilmänliikkeisiin ja aivoaktiivisuuteen. Tätä tutkittiin leksikaalisen päätöksenteon tehtävällä, jonka aikana koehenkilöiltä mitattiin rinnakkaisrekisteröityä katseenseuranta-aivosähkökäyrää. Epä- ja pseudosanat johdettiin vaihtamalla kirjain sanan alusta tai lopusta siten, että syntyvällä kirjainjonolla ei ollut semanttista merkitystä. Epäsanojen tapauksessa lopputulos oli suomen kielen ortografian vastainen, mutta pseudosanat olisivat rakenteeltaan voineet olla suomen kielen sanoja. Kokonaisuutena neljännessä osatutkimuksessa havaittiin, että poikkeaman tyyppi ja sijainti vaikuttivat yhteisesiintyvien aivovasteiden, pienten sakkadien ja reaktioaikojen dynamiikkaan. Näyttää siltä, että mikäli poikkeama rikkoo suomen kielen ortografisia sääntöjä, poikkeama havaitaan varhain ja kirjainjonon semanttista merkitystä ei edes yritetä löytää. Mutta jos poikkeaman seurauksena on suomen kielen sanaa muistuttava kirjainjono, aivot pyrkivät prosessoimaan kirjainjonon merkityksen.

Kokonaisuudessaan tämän väitöskirjan tulokset viittaavat siihen, että joissakin tapauksissa silmänliike-efektin yhteydessä voidaan havaita johdonmukaisesti yhteisesiintyvä aivotoiminnan modulaatio. Tämä ei kuitenkaan ole yksiselitteistä, sillä erityisesti toisessa ja neljännessä osatutkimuksessa esiintyi silmänliike-efektejä, joiden kanssa ei havaittu yhteisesiintyviä aivotoiminnan modulaatioita. Tämä ei ole ainutlaatuinen löydös, vaan muissa tutkimuksissa on löydetty samansuuntaisia tuloksia sanan yleisyyden silmänliike-efektien ja aivotoiminnan modulaation suhteen luonnollisen lukemisen aikana.

Tämän väitöskirjan tuloksia voidaan käyttää edistämään teoreettista ymmärrystä silmän ja mielen suhteesta luonnollisen lukemisen aikana. Tulokset viittaavat siihen, että tämänhetkistä teoreettista ymmärrystä on tulevaisuudessa laajalti uudelleenarvioitava.. On kuitenkin huomattava, että rinnakkaisrekisteröity katseenseuranta-aivosähkökäyrä on hyvin nuori tutkimusmenetelmä, joten tässä väitöskirjassa esitetyt tulokset tarvitsevat tuekseen lisää perustutkimusta, ennen kuin kauaskantoisia johtopäätöksiä voidaan tehdä.

REFERENCES

Abbott, M. J., & Staub, A. (2015). The effect of plausibility on eye movements in reading: Testing E–Z Reader's null predictions. *Journal of Memory and Language*, *85*, 76–87. https://doi.org/10.1016/j.jml.2015.07.002

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Memory and language mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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

Barber, H. A., van der Meij, M., & Kutas, M. (2013). An electrophysiological analysis of contextual and temporal constraints on parafoveal word processing. *Psychophysiology*, *50*(1), 48–59. https://doi.org/10.1111/j.1469-8986.2012.01489.x

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*(3), 229–241. https://doi.org/10.1016/0013-4694(94)90094-9

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 S. P. Liversedge, I. Gilchrist & S. Everling (Eds.), *The Oxford handbook of eye movements*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199539789.013.0036

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

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*(6), 719–722. https://doi.org/10.1016/0042-6989(75)90290-4

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. https://doi.org/10.1016/S0010-0277(03)00164-1

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eye-tracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*, *49*(2), 602–615. https://doi.org/10.3758/s13428-016-0734-0

Cornelissen, T., Sassenhagen, J., & Võ, M. L.-H. (2019). Improving free-viewing fixation-related EEG potentials with continuous-time regression. *Journal of Neuroscience Methods 313,* 77-94. https://doi.org/10.1016/J.JNEUMETH.2018.12.010

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

Degno, F., Loberg, O., Zang, C., Zhang, M., Donnelly, N., & Liversedge, S. P. (2019). Parafoveal previews and lexical frequency in natural reading: Evidence from eye movements and fixation-related potentials. *Journal of Experimental Psychology: General, 148(3), 453.* https://doi.org/10.1037/xge0000494

Dehaene, S. (2009). *Reading in the brain: The science and evolution of a human invention*. New York, USA:Viking.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

Démonet, J.-F., Taylor, M. J., & Chaix, Y. (2004). Developmental dyslexia. *The Lancet*, *363*(9419), 1451–1460. https://doi.org/10.1016/S0140-6736(04)16106-0

Dimigen, O. (2018). Optimized ICA-based removal of ocular EEG artifacts from free viewing experiments. *BioRxiv*, 446955. https://doi.org/10.1101/446955

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

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology*, *44*(6), 927–935. https://doi.org/10.1111/j.1469-8986.2007.00593.x

Dürrwächter, 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., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*(9), 1035–1045. https://doi.org/10.1016/S0042-6989(03)00084-1

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

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research*, *1146*, 50–58. https://doi.org/10.1016/j.brainres.2006.08.038

Frömer, R., Maier, M., & Rahman, R. A. (2018). Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. *Frontiers in Neuroscience*, *12*(FEB), 48. https://doi.org/10.3389/fnins.2018.00048

Galaburda, A. M. (2005). Dyslexia—A molecular disorder of neuronal migration. *Annals of Dyslexia*, *55*(2), 151–165. https://doi.org/10.1007/s11881-005-0009-4

Godoy-Fernandez, J., Lüders, H. O., Dinner, D. S., Morris, H. H., & Wyllie, E. (1990). Versive eye movements elicited by cortical stimulation of the human brain. *Neurology*, *40*(2), 296–299. https://doi.org/10.1212/WNL.40.2.296

Guttorm, T. K., Leppänen, P. H. T., Poikkeus, A.-M., Eklund, K. M., Lyytinen, P., & Lyytinen, H. (2005). Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. *Cortex*, *41*(3), 291–303. https://doi.org/10.1016/S0010-9452(08)70267-3

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

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., & Parviainen, T. (2014). Gaze position reveals impaired attentional shift during visual word recognition in dysfluent readers. *PLOS ONE*, *9*(9), e108937. https://doi.org/10.1371/journal.pone.0108937

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

Henderson, J. M., Luke, S. G., Schmidt, J., & Richards, J. E. (2013). Co-registration of eye movements and event-related potentials in connected-text paragraph reading. *Frontiers in Systems Neuroscience*, *7*(July), 28. https://doi.org/10.3389/fnsys.2013.00028

Hickey, C., Mcdonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, *18*, 604–613.

Holmqvist, K., Nyström, M., & Andersson, R. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford, UK: Oxford University Press.

Hutzler, F., Braun, M., Võ, M. L. H., Engl, V., Hofmann, M., Dambacher, M.,…Jacobs, A. M. (2007). Welcome to the real world: Validating fixation-related brain potentials for ecologically valid settings. *Brain Research*, *1172*(1), 124–129. https://doi.org/10.1016/j.brainres.2007.07.025

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Oxford, UK: Elsevier Science Ltd. https://doi.org/10.1016/B978-008043361-5/50003-1

Inhoff, A. W., Radach, R., Eiter, B. M., & Juhasz, B. (2003). Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56 A*(5), 803–827. https://doi.org/10.1080/02724980244000639

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *NeuroImage*, *20*(2), 693–712. https://doi.org/10.1016/S1053-8119(03)00343-4

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

Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*(2), 250–274. https://doi.org/10.1111/j.1469-8986.2007.00618.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., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12

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.

Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, 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., Schlesewsky, 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

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

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

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye movement patterns during sentence reading. *Journal of Memory and Language*, *65*(1), 42–73. https://doi.org/10.1016/J.JML.2011.03.002

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kwantes, P. J., & Mewhort, D. J. K. (1999). Evidence for sequential processing in visual word recognition. *Journal of Experimental Psychology*, *25*, 376–381.

Laubrock, J., Engbert, R., & Kliegl, R. (2005). Microsaccade dynamics during covert attention. *Vision Research*, *45*(6), 721–730.

https://doi.org/10.1016/j.visres.2004.09.029

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

Lindeman, J. (1998). *ALLU-Ala-asteen lukutesti* [ALLU-Reading test for primary school] (Vol. 90). https://doi.org/10.1016/S0022-3913(03)00368-8

López-Peréz, P. J., Dampuré, J., Hernández-Cabrera, J. A., & Barber, H. A. (2016). Semantic parafoveal-on-foveal effects and preview benefits in reading: Evidence from Fixation Related Potentials. *Brain and Language*, *162*, 29–34. https://doi.org/10.1016/j.bandl.2016.07.009

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639

Luke, S. G., & Christianson, K. (2018). The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior Research Methods*, *50*(2), 826–833. https://doi.org/10.3758/s13428-017-0908-4

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, *53*(1), 1–14. https://doi.org/10.1007/s11881-003-0001-9

Lyytinen, H., Ahonen, T., Eklund, K., Guttorm, T., Kulju, P., Laakso, M. L.,…Viholainen, H. (2004). Early development of children at familial risk for dyslexia—Follow-up from birth to school age. *Dyslexia*, *10*, 146–178. https://doi.org/10.1002/dys.274

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, *8*(3), 145–151. https://doi.org/10.1109/ICOSP.2002.1180091

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, *38*(3), 487–497. https://doi.org/10.1016/S0896-6273(03)00197-1

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

Martelli, M., Di Filippo, G., Spinelli, D., & Zoccolotti, P. (2009). Crowding, reading, and developmental dyslexia. *Journal of Vision*, *9*(4), 14–14. https://doi.org/10.1167/9.4.14

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*(3), 229–232. https://doi.org/10.1038/nrn1348

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

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 913–934. https://doi.org/10.1037/a0022964

Maurer, U., Brem, S., Bucher, K., & Brandeis, D. (2005). Emerging neurophysiological specialization for letter strings. *Journal of Cognitive Neuroscience, 17*(10), 1532–1552. https://doi.org/10.1162/089892905774597218

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

Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P.,…Brandeis, D. (2006). Coarse neural tuning for print peaks when children learn to read. *NeuroImage*, *33*(2), 749–758. https://doi.org/10.1016/J.NEUROIMAGE.2006.06.025

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7*(7), 293–299. https://doi.org/10.1016/S1364-6613(03)00134-7

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. *Cognitive Science, 41*(S6), 1232–1263. https://doi.org/10.1111/cogs.12384

Meyberg, S., Werkle-Bergner, M., Sommer, W., & Dimigen, O. (2015). Microsaccade-related brain potentials signal the focus of visuospatial attention. *NeuroImage*, *104*, 79–88. https://doi.org/10.1016/j.neuroimage.2014.09.065

Nevala, J., Kairaluoma, L., Ahonen, T., Aro, M., & Holopainen, L. (2006). Lukemis-ja kirjoittamistaitojen yksilötestistö nuorille ja aikuisille. *Jyväskylä: Niilo Mäki …*. Jyväskylä: Niilo Mäki Instituutti.

Nevala, J., & Lyytinen, H. (2000). *Sanaketjutesti* [Word chain test]. Jyväskylä: Niilo Mäki Institute.

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

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG*. Oxford, UK: Oxford University Press.

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience & Biobehavioral Reviews*, *30*(6), 808–822. https://doi.org/10.1016/j.neubiorev.2006.06.007

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*(2), 20. https://doi.org/10.1167/7.2.20

Perea, M., & Gomez, P. (2012). Subtle Increases in interletter spacing facilitate the encoding of words during normal reading. *PLOS ONE*, *7*(10), e47568. https://doi.org/10.1371/journal.pone.0047568

Perea, M., Panadero, V., Moret-Tatay, C., & Gómez, P. (2012). The effects of inter-letter spacing in visual-word recognition: Evidence with young normal readers and developmental dyslexics. *Learning and Instruction*, *22*(6), 420–430. https://doi.org/10.1016/J.LEARNINSTRUC.2012.04.001

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

Price, C. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

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., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, *17*(6), 834–839. https://doi.org/10.3758/PBR.17.6.834

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences, 26(4)*, 445-476.

Reichle, E. D., Tokowicz, N., Liu, Y., & Perfetti, C. A. (2011). Testing an assumption of the E-Z Reader model of eye movement control during reading: Using event-related potentials to examine the familiarity check. *Psychophysiology*, *48*(7), 993–1003. https://doi.org/10.1111/j.1469-8986.2011.01169.x

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

Rolfs, M., Kliegl, R., Laubrock, J., & Engbert, R. (2015). Microsaccades are an index of covert attention. *Psychological Science, 18*(4), 364–366.

Salmelin, R., Service, E., Kiesila, P., & Uutela, K. (1996). Impaired visual word processing in dyslexia revealed with magnetoencephalography. *Annals of*

Neurology, 40(2), 157–162.

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience, 23*(3), 514–523. https://doi.org/10.1162/jocn.2009.21370

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology, 56*(6), e13335. https://doi.org/10.1111/psyp.13335

Schall, J. D. (2004). On the role of frontal eye field in guiding attention and saccades. *Vision Research*, *44*(12), 1453–1467. https://doi.org/10.1016/j.visres.2003.10.025

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493. https://doi.org/10.1016/J.TICS.2003.09.010

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

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Simola, J., Holmqvist, K., & Lindgren, M. (2009). Right visual field advantage in parafoveal processing: Evidence from eye-fixation-related potentials. *Brain and Language*, *111*(2), 101–113. https://doi.org/10.1016/j.bandl.2009.08.004

Sinn, P., & Engbert, R. (2016). Small saccades versus microsaccades: Experimental distinction and model-based unification. *Vision Research*, *118*, 132–143. https://doi.org/10.1016/J.VISRES.2015.05.012

Slattery, T. J., & Rayner, K. (2013). Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text. *Attention, Perception, & Psychophysics*, *75*(6), 1275–1292. https://doi.org/10.3758/s13414-013-0463-8

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

Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K.-P. (2002). Neural mechanisms of saccadic suppression. *Science*, *295*(5564), 2460–2462. https://doi.org/10.1126/science.1068788

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

Valsecchi, M., & Turatto, M. (2009). Microsaccadic responses in a bimodal oddball task. *Psychological Research*, *73*(1), 23–33. https://doi.org/10.1007/s00426-008-0142-x

Veldre, A., & Andrews, S. (2016). Is semantic preview benefit due to relatedness or plausibility? *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 939–952. https://doi.org/10.1037/xhp0000200

Veldre, A., & Andrews, S. (2017). Parafoveal preview benefit in sentence reading: Independent effects of plausibility and orthographic relatedness. *Psychonomic Bulletin and Review*, *24*(2), 519–528. https://doi.org/10.3758/s13423-016-1120-8

Veldre, A., & Andrews, S. (2018). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during reading. *Journal of Memory and Language*, *100*, 1–17. https://doi.org/10.1016/J.JML.2017.12.002

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A., & Valero-Cabré, A. (2014). Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Frontiers in Integrative Neuroscience*, *8*, 66. https://doi.org/10.3389/fnint.2014.00066

Widmann, A., Engbert, R., & Schroger, E. (2014). Microsaccadic responses indicate fast categorization of sounds: A novel approach to study auditory cognition. *Journal of Neuroscience*, *34*(33), 11152–11158. https://doi.org/10.1523/JNEUROSCI.1568-14.2014

Yablonski, M., Polat, U., Bonneh, Y. S., & Ben-Shachar, M. (2017). Microsaccades are sensitive to word structure: A novel approach to study language processing. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-04391-4

Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 597–613. https://doi.org/10.1037/xlm0000064

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

Zivony, A., Allon, A. S., Luria, R., & Lamy, D. (2018). Dissociating between the N2pc and attentional shifting: An attentional blink study. *Neuropsychologia*, *121*, 153–163. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2018.11.003

Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M.,…Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences*, *109*(28), 11455–11459. https://doi.org/10.1073/PNAS.1205566109

ORIGINAL PAPERS

I

BREAKING DOWN THE WORD LENGTH EFFECT ON READERS' EYE MOVEMENTS

by

Jarkko Hautala & Otto Loberg, 2015

Language, Cognition and Neuroscience, *30* (8), 993–1007.

Reproduced with kind permission by Routledge.

Breaking down the word length effect on readers' eye movements

Jarkko Hautala[a, b,*] and Otto Loberg[b]

[a]Agora Center, University of Jyväskylä, Jyväskylä, Finland; [b]Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

(Received xx Month xxxx; accepted xx Month xxxx)

[*] Corresponding author. Email: jarkko.v.hautala@jyu.fi

Abstract

Previous research on the effect of word length on reading confounded the number of letters (NrL) in a word with its spatial width. Consequently, the extent to which visuospatial and attentional-linguistic processes contribute to the word length effect on parafoveal and foveal vision in reading and dyslexia is unknown. Scholars recently suggested that visual crowding is an important factor for determining an individual's reading speed in fluent and dyslexic reading. We studied whether the NrL or the spatial width of target words affects fixation duration and saccadic measures in natural reading in fluent and dysfluent readers of a transparent orthography. Participants read natural sentences presented in a proportional font that contained spatially narrow and wide four- to seven-letter target words. The participants looked at spatially narrow words overall for a longer duration partially due to more frequent regressions, which showed that crowding can disrupt word recognition during normal reading. In addition, reliable NrL effects on fixation duration suggest that letters are important attentional units during reading. Saccadic measures including relative landing position, refixation and skipping probability were strongly affected by spatial width and slightly affected by the NrL, which suggests that saccadic programming and parafoveal processing of upcoming words are limited by visual acuity more than by attentional factors. The dysfluent readers overall had longer fixation durations for words but did not show larger crowding or NrL effects.

Keywords: reading fluency; eye movements; word length, crowding; word skipping

Longer words require more time to be recognized (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Hautala, Hyönä, & Aro, 2011a; New, 2006; see Barton, Hashim, Eklinder, & Hills, 2014 for a recent review), which is pronounced in developmental dyslexia (e.g., De Luca, Di Pace, Judica, Spinelli, & Zoccolotti, 1999; Hautala, Aro, Eklund, Lerkkanen, & Lyytinen, 2013; Hautala, Hyönä, Aro & Lyytinen, 2011b; Ziegler, Perry, Ma-Wyatt, Ladner & Schulte-Körne, 2003). Generally, the temporal word length effect has been thought to stem from linguistic-attentional processing. However, there are several potent visuospatial explanations for the word length effect, and scholars have suggested that dyslexia may at least partially stem from a visuospatial processing deficit (Martelli, Filippo, Spinelli, & Zoccolotti, 2009; Zorzi, Barbiero, Facoetti, Lonciari, et al., 2012). Further, although during reading saccadic programming of the landing position is mainly based on spatial information about word spaces and fixation durations that reflect word recognition processes of textual information (Hautala et al., 2011a; Infoff, Eiter, Radach, & Juhasz, 2003), direct empirical evidence of how deep this functional dissociation is and whether it holds for all types of saccades, including refixations, regressions and word skips, is lacking. In the present study, for the first time, the spatial width of a word and the number of letters (NrL) were orthogonally manipulated, and their contribution to various eye movement measures of normal and dysfluent reading were examined.

Temporal word length effects are due to visual or perceptual limitations that force readers to make progressive refixations when they read long words (Vergilino-Perez, Collins, & Doré-Mazars, 2004). Human acuity foveal vision covers only two visual degrees, which typically equals six to eight letters (see Rayner, 1998), and acuity degrades rapidly in the parafovea. The size of the perceptual span in reading has been studied with a gaze-contingent display technique such as the moving window paradigm in which the number of upcoming letters visible to a reader is manipulated (see Rayner, 1998). These studies have shown that people learn to recognize letters that extend from foveal vision to the reading direction. The perceptual span increases during reading development (Häikiö, Hyönä, Bertram, & Niemi, 2009) so that fluent adult readers can identify seven to eight letters forward (McConkie & Rayner, 1975; see Rayner, 1998 for a review), while poor readers have a smaller span (Bosse, Tainturier, & Valdois, 2007; Hautala & Parviainen, 2014; Häikiö et al., 2009; Rayner, 1983). Reading is partly serial phonological decoding, a skill that becomes automatic during elementary school (Zoccolotti, De Luca, Di Filippo, Judica, & Martelli, 2009) but remains difficult for children with developmental dyslexia (Share, 1995; Wimmer, 1993, 1996a,b; Zoccolotti, De Luca, & Di Pace, 2005). When reading short words, the word length effects (New, 2006) of a single fixation (Hautala et al., 2011a,b) may also reflect attentional serial letter processing during word recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007; see especially Hawelka et al., 2010). In addition, problems in phonological decoding or whole-word recognition may explain the difficulty with reading that children with developmental dyslexia experience and their distinctive eye movement behavior during reading, including longer fixations, more frequent refixations and regressions and shorter saccades (see Bellochi, Muneaux, Bastien-Toniazzo, & Ducrot, 2013, for a recent review).

However, word length effects could reflect visuospatial processing. The leading models of eye movement control in reading assume that the speed of letter recognition increases by the distance of the letter from the fixation point (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Radach & Reilly, 2006; Reichle, Pollatsek, & Rayner, 2006), which is called here as the *visual acuity hypothesis for letter encoding*. In principle, this process may be responsible for the temporal word length effects within foveal vision, and deficits in this process may contribute to larger word length effects in readers with dyslexia. In contrast, *visual crowding* (Bouma, 1970, 1973), which is the degradation of the spatial frequency of vision as a function of eccentricity, impairs perception of nearby objects in peripheral and parafoveal vision. Pelli, Tillman, Freeman, Su, Berger, and Majaj (2007) provided extensive evidence that reading is essentially achieved within an uncrowded span that roughly equals the foveal vision and the perceptual span, beyond which the crowding makes identifying letters impossible. Visual interference of very near objects (< .1 visual degrees) within foveal vision is called foveal crowding (Levi, Klein, & Hariharan, 2002). Although this concept is controversial (Huurneman, Boonstra, Cox, Cillessen, & van Rens (2012), visually more densely packed words may be harder to read and lead to increased fixation times.

Regarding dyslexia, Martelli and colleagues (2009) found that children with dyslexia showed larger crowding effects in a peripheral letter identification task, and the magnitude of this crowding effect was associated with individual differences in reading speed. Bellochi et al. (2013) reported similar findings in a subgroup of individuals with dyslexia. Providing further support for the idea of a visual processing deficit in at least a subgroup of individuals with dyslexia, scholars recently reported that individuals with dyslexia showed abnormalities in several low-level oculomotor skills including binocular coordination (for a review, see Bellochi et al., 2013; Gori & Facoetti, 2015). Recent research reports that increasing letter and word spacing improve reading speed in children with dyslexia (Perea, Panadero, Moret-Tatay, & Gómez, 2012; Zorzi, Barbiero, Facoetti, Lonciari et al., 2012). Zorzi and colleagues (2012) suggested that wider letter spacing may help children with dyslexia focus on each letter during the phonological decoding process.

Word spaces and word length are important determinants of saccadic behaviors such as landing position (where to fixate on the word) and whether to skip or refixate a word (for a complete review of saccadic behavior during reading, see Rayner, 1998 and Schotter, Angele, & Rayner, 2012). Saccades are planned toward the preferred viewing location slightly left of the word center, which provides optimal visibility of the words. Corrective refixations toward the word center are made after mislocated fixations due to oculomotor error. If the word is long and extends over the foveal vision, the initial fixation location shifts to the word beginning, and the probability of making a progressive refixation increases, which suggests that refixations may even be preplanned according to word length information in the parafovea (Vergilino-Perez et al., 2004). Very short words are often skipped, especially when the preceding fixation is near, presumably because they fall within the acuity vision and can be parafoveally identified to a sufficient degree. Skipped words are then regressed more often because the parafoveal recognition may have been erroneous or words were skipped by accident due to oculomotor error in saccade targeting. Regressions are also made due to challenges in reading comprehension. Generally, this body of knowledge is almost exclusively based on studies that manipulated word spaces and word length in text presented in a monospaced font. Therefore, whether attentional or perceptual span (NrL), or visual (spatial width) factors are responsible for these saccadic effects, or both is not actually known. In addition, whether parafoveal preprocessing affects the landing position and the probability of skipping a word is governed by attentional or visual acuity limitations are not known (Schotter et al., 2012).

Some studies have been conducted on visuospatial influences on readers' eye movements. One method for studying these influences involves manipulating fonts and letter spacing. Rayner, Slattery, and Bélanger (2010) used the moving window technique to investigate whether letter spacing of proportional and monospaced fonts influences readers' eye movements. Although the researchers observed no effects on the perceptual span in the number of words, the decreased letter spacing led to a faster overall reading rate via the decreased number of fixations and their duration, but the number of regressions increased. In a follow-up study in which Slattery and Rayner (2013) manipulated letter spacing, text presented with standard letter spacing yielded the fastest reading times relative to the decreased or increased letter spacing condition. The authors also observed benefits in the average fixation duration for decreased letter spacing when the word spacing was increased, which suggested crowding effects were present at the letter and word levels during reading. Regarding font effects, Rayner et al. (2010) did not find an overall difference in reading rates; however, in Slattery and Rayner's study the proportional font was read faster than the monospaced font. A common finding for both studies and our previous study (Hautala et al., 2011a) is that proportional fonts tend to be read with fewer but longer fixations than monospaced fonts, which suggests two contrasting factors affect reading: While more letters in a proportional font fall in acuity vision, crowding is increased by the shorter distance between the letters.

However, in contrast, Perea and Gomez (2012) found that slightly expanded letter spacing provided weak benefits for reading speed and average fixation durations, but larger expanded spacing resulted an increased number of fixations per word. Perea and Gomez also found the initial fixation locations shifted to the word beginning when the letter spacing was expanded, but Slattery and Rayner (2013) did not confirm this finding. One possible reason for these conflicting findings across letter spacing and font studies is that manipulating letter spacing leads to unfamiliar typography for a reader, which may have consequences of its own for readers' eye movements.

In an early study, Morrison and Rayner (1981) studied eye movements when individuals read identical text at varying viewing distances. The researchers found viewing distance had no effect on saccade amplitude in letters, which led the researchers to conclude that attentional, not visual, factors determine how many letters are processed during a fixation. However, fixation durations were longer at larger viewing distances, suggesting that visual factors affect how efficiently letters are recognized.

Recently, Miellet, O'Donnell, and Sereno (2009) compensated for the degradation of visual acuity by magnifying parafoveal letters accordingly. The researchers assumed that if the perceptual span were limited by visual acuity, parafoveal magnification would improve parafoveal letter recognition and lead to increased perceptual span. This manipulation did not increase the perceptual span measured by varying the size of the moving window, which led the authors to suggest that the perceptual span is limited by attention, not visual acuity. We believe it is difficult to draw any firm conclusions about normal reading from such an unnatural reading condition; however, reading with and without parafoveal magnification seemed to be surprisingly similar. The authors agreed that providing their participants more training with such an unnatural reading condition may have made improvements in reading possible.

McDonald (2006) rendered all words in a text with an equal spatial width and compared eye movement measures with six- and eight-letter target words. He found that temporal fixation duration measures that reflected word recognition including first, single fixation and gaze durations and a saccadic measure of refixation probability were influenced by the NrL, whereas most of the saccadic measures, including skipping probability, landing position and launch distance, were not affected by the NrL in a word.

Hautala et al. (2011a) compared two NrL effects: the one controlled by spatial width by taking advantage of variability in letter widths (e.g., “m” and “i,” mama vs. flight) when presented in a proportional font and the NrL effect with spatial width confounded when presented in a monospaced font. Again, the NrL affected the fixation duration measures, and spatial width affected only the saccadic measures. Refixation probability was almost significantly affected by spatial width, not by the NrL as McDonald found. However, both studies had shortcomings. McDonald (2006) studied words rendered unnaturally with equal spatial width, which led to visually unnatural crowded words. Hautala et al. (2011a) confounded font type with spatial width manipulation.

In the present study, we optimized the experimental design by comparing reading of narrow and wide four- to seven-letter words, all presented in the same proportional font. By manipulating the orthography of the words’ spatial width and the number of letters, we aimed to resolve whether fixation duration and saccadic word length effects on readers' eye movements result from the visual (spatial width) or attentional (NrL) level of processing, and whether these effects are associated with reading fluency. According to the attentional view, the fixation duration and saccadic effects (except landing position) should be based on the NrL. According to the visual acuity hypothesis, the fixation durations should be longer for wider words, and the saccadic measures should reflect the increased processing demands of wide words (less skipping and more refixations and regressions). According to the visual crowding hypothesis, the fixation durations should be longer, and the saccadic measures should reflect the increased processing demands of narrow words that contain more objects in a given space. Dysfluent readers were expected to show generally increased fixation durations and more frequent refixations and regressions, and possibly a larger influence of NrL and crowding on these measures.

Materials and methods

Apparatus

An SMI Hispeed eye tracker with a 500 Hz sampling rate was used to record the eye movements of the participants' right eye. The computer screen (size 375 × 300 mm, resolution 1024 × 768 pixels) was located 670 mm from the participant's eye.

Participants

The participants were 37 native Finnish-speaking young adults (age *M* = 20, *SD* = 4.8 years). They were recruited by sending an email to the student mailing lists of local high schools, a vocational university and a university. In the letter, students with and without reading problems were invited to participate in the study. A written informed consent was obtained from the participants before their participation. The experiment was undertaken in accordance with the Declaration of Helsinki. The ethical committee of the University of Jyväskylä approved the research protocol. The participants' reading skills were assessed with text reading and word list reading subtests from an assessment battery for reading disabilities (Nevala, 2007). In the text reading subtest, the participant score was the number of words read aloud within a 3 min time limit. The word list reading subtest participant's score was the time taken to read aloud the 30-item word list. The participant was considered a dysfluent reader (DYS, $N = 11$) if he or she scored below the 11th percentile in population on either the text (controls 366 words, DYS 295 words, $t(32) = 5.3$, $p < .001$) or word reading subtask (controls 22.2 s, DYS 36.5 s, $t(32) = 7.3$, $p < .001$) while the other participants constituted the control group of fluent readers. The participants' IQ was assessed with the standard progressive matrices test (SPM test; Raven, Raven, & Court, 1998). One participant was excluded from the analysis due to a poor score on the SPM test, after the groups had equal IQs, $t < .62$, and another was dropped due to an incomplete measurement. Participants received movie tickets as compensation for their participation. Eye tracking of one participant was not possible due to her tendency to keep her eyes half closed. Thus, our total sample was 34.

Target words and sentences

To study the independent influences of the spatial width of a word, and the NrL in a word, narrow words that contain several narrow letters such as "l," and wide words that contain several wide letters such as "m" were selected as the target words. Each category of wide and narrow four- to seven-letter words contained 20 words, for a total of 160 target words. The visual and psycholinguistic descriptions of the stimuli are presented in Table 1, and the entire list of the stimulus sentences is in the Aappendix A. The categories were controlled for word and bigram frequency, based on a large newspaper text corpus of lemma frequencies (Language Bank of Finland, 2007). The number of orthographic neighbors was also controlled except between four- and seven-letter words since longer words have unavoidably fewer word neighbors in Finnish. The target words were embedded in 80 sentence beginnings, paired as follows: 1) narrow four-letter and wide seven-letter words, 2) wide four-letter and narrow seven-letter words, 3) narrow five-letter and wide five-letter words, and 4) narrow six-letter and wide six-letter words. A comparison of pairs 1 and 2 would reveal whether the NrL effect is a similar size when the spatial width of the words is the same (i.e., controlled) vs. very different (i.e., not controlled), whereas separate analyses of pairs 3 and 4 would reveal whether spatial width has an effect when the NrL is controlled. Pairs 3 and 4 could not be included in the same analysis since narrow five- and six-letter words and wide five- and six-letter words had different spatial widths. For the same reason and because of the differing sentence frames, we did not evaluate the main effect of spatial width in the analysis of pairs 1 and 2.

Word pairs were formed so that they shared the same part of speech (nouns with nouns, etc.) and that they had semantic likeness within a word pair. Examples of the stimuli embedded in the envelope sentences are shown in Figure 1. We constructed sentences with long words around the target word to minimize pre- and post-target word skipping. An online survey (Limesurvey; Schmitz, 2010) was used to evaluate the predictability, plausibility and emotional charge of the sentences, and there were no differences in these qualities within sentence pairs, *F*s < .1. Predictability was established with the standard cloze probability procedure. Plausibility evaluation was performed by presenting the sentence pair to the survey participants and asking whether sentence 1 was more plausible, or sentence 2 was more plausible, or the sentences were equally plausible. Emotional charge was evaluated by asking the survey participants whether sentence 1 evoked emotional reaction, sentence 2 evoked emotional reaction, both sentences evoked emotional reactions, or neither sentence evoked emotional reaction.

Figure 1. Sample sentence pair. Sentence translates literally "To my greatest misfortune binoculars/egg broke after falling from the table," with "binoculars" and "egg" being the target words.

Mitä suurimmaksi epäonnekseni kiikari hajosi pudottuaan pöydältä.
Mitä suurimmaksi epäonnekseni muna hajosi pudottuaan pöydältä.

Table 1. Mean spatial width in degrees, word frequency in a million words, number of word neighbors and bigram frequency in a thousand words of target word categories with standard deviations in parentheses.

Comparison	Spatial width controlled		Not controlled		Five-letter words		Six-letter words	
Letters	Four	Seven	Four	Seven	Narrow	Wide	Narrow	Wide

Spatial width	1.47	1.49	.93	2.18	.96	1.44	1.45	2.06
	(.08)	(.10)	(.06)	(.22)	(.09)	(.08)	(.07)	(.17)
Word	34	98	144	105	112	73	192	155
frequency	(69)	(210)	(154)	(95)	(117)	(96)	(578)	(286)
Word	2.90	.55	3.50	.50	2.15	3.00	1.15	1.40
neighbors	(1.94)	(.944)	(2.33)	(.89)	(1.93)	(1.97)	(1.22)	(1.09)
Bigram	1.7	3.0	5.9	4.1	5.5	5.3	5.4(4.8)	6.2
Frequency	(2.2)	(2.6)	(5.0)	(2.9)	(3.0)	(3.7)		(3.4)

Procedure

Participants leaned their head against a forehead and chin rest. A 13-point calibration procedure was repeated at the beginning of the experiment and after every 40 sentences, and was repeated if the deviation between the calibration and the validation was more than 0.2 degrees. Each trial started with the requirement to look a fixation cross at the left edge of the upper half of the screen for 500 ms, which triggered the stimulus sentence to appear. After the participant read the sentence, he or she fixated for 500 ms on a fainter fixation cross at the right edge of the screen, which triggered the sentence to disappear. Sentences were presented in the proportional font Calibri at 16 points (see Table 1 for the visual degrees of the stimuli words). After 24 specific sentences, a yes or no question about the previous sentence appeared, and the participants were instructed to answer by choosing the correct alternative with the mouse. All 160 sentences were randomized for every participant.

Data processing

Fixations were detected with the saccade–velocity-based algorithm developed by the eye tracker manufacturer (SMI). The area of interest was analyzed and the dependent measure was calculated with the manufacturer's analysis package for reading studies. The parameters for detecting a saccade were a saccade velocity threshold of 40 °/s and a minimum saccade duration of 22 ms; however, these parameters left some anomalies to saccadic amplitude data so saccades that exceeded 150 pixels in length were excluded from further analyses (56 cases). Fixations shorter than 50 ms and longer than 1000 ms were excluded from further analysis.

Data analysis

The following target word–specific measures that reflect word recognition processes were selected: Total fixation duration and first-pass gaze duration reflected the overall and first-pass word recognition processes, respectively. Single-fixation duration included occurrences when a word was recognized by one fixation. To study the influence of manipulated variables on eye movement control, saccading measures including relative landing position (percentages of a word's horizontal width) and refixation, skipping and regression probabilities were analyzed. In the within-subject F1 analyses, these measures were the subject of repeated measures analysis of variance with two-level within-subject factors of the NrL (four, seven) and spatial width (controlled, not controlled) in the analysis of four- and seven-letter words. In this analysis, we were not interested in the main effect of spatial width (as it was presented in different sentence pairs) but in the possible interaction of spatial width and the NrL. In the analyses of five- and six-letter words, there was a two-level spatial width factor (narrow, wide). A two-level between-subject factor of reading fluency (CONTROLS, DYS) was used in each analysis. Significant interactions were inspected with paired *t*-tests. In the F2 item analysis of four- and seven-letter words, there were two-level within-subject factors of NrL (four, seven) and Group (CONTROLS, DYS) and a two-level between-subject factor of spatial width (controlled, not-controlled). In the F2 analyses of five- and six-letter words, there was a two-level within-subject factor of spatial width (narrow, wide) and Group (CONTROLS, DYS).

The three analysis blocks were used to test the following hypotheses: If narrow words induced values that reflected increased processing demands, the crowding hypothesis was supported. If wider words induced values that reflected increased processing demands, the visual acuity hypothesis was supported. If spatial width had no influence but the NrL did, the letters as cognitive processing units hypothesis was supported.

Results

The means for all measures are given in Tables 2 and 3 separately for fluent and dysfluent readers. Only significant F1 and F2 effects or significant F1 effects with non-significant F2 result of particular interest are reported. The complete ANOVA tables are provided in the Appendix B. The groups did not differ in answering the comprehension control questions, $F < 1$; the DYS group answered the questions with accuracy 94.3% accuracy, and the controls with 93.1% accuracy.

Table 2. Means and standard deviations for eye movement measures for fluent readers.

Comparison	Spatial width controlled				Spatial width not controlled				Five-letter words				Six-letter words			
	Four-letter		Seven-letter		Four-letter		Seven-letter		Narrow		Wide		Narrow		Wide	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Single fixation duration	253	49	266	54	240	47	263	63	250	64	231	44	266	53	256	48
Gaze duration	259	46	277	52	242	45	280	64	251	55	236	45	266	51	272	50
Total fixation duration	309	74	337	75	288	66	333	93	326	88	270	60	342	102	323	77
Refixation probability	25	16	31	16	21	13	31	18	25	16	18	15	32	19	30	15
Regression probability	14	12	14	8	15	10	10	6	16	10	7	5	17	13	10	8
Relative landing position	58	7	52	9	58	8	45	7	58	9	57	9	55	8	48	6
Skipping probability	14	15	7	6	28	23	2	5	29	21	9	11	9	13	3	6

Table 3. Means and standard deviations for eye movement measures for dysfluent readers.

Comparison	Spatial width controlled				Spatial width not controlled				Five-letter words				Six-letter words			
	Four-letter		Seven-letter		Four-letter		Seven-letter		Narrow		Wide		Narrow		Wide	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Single fixation duration	288	79	323	95	280	74	291	92	294	68	252	66	287	90	292	85
Gaze duration	295	79	334	96	286	71	315	93	293	64	288	78	302	90	316	85
Total fixation duration	383	95	451	162	375	130	429	127	406	72	353	104	425	135	391	132
Refixation probability	39	15	40	19	30	15	46	17	31	17	30	15	39	16	35	15
Regression probability	15	8	14	9	18	11	14	9	18	6	10	6	18	9	9	8
Relative landing position	55	10	52	12	59	10	46	11	60	7	55	11	55	11	52	11
Skipping probability	9	13	4	5	26	15	3	6	24	17	9	10	9	9	5	13

Influence of orthogonal manipulation of spatial width and NrL

Single fixation duration

The NrL had a main effect, *F1*(1, 32) = 8.264, $p = .007$, $\eta_p^2 = .205$, *F2*(1, 38) = 3.196, $p = .082$, $\eta_p^2 = .078$; seven-letter words received longer fixation durations than four-letter words (286 vs. 265 ms).

Gaze duration

The NrL had a main effect, *F1*(1, 32) = 24.112, $p < .001$, $\eta_p^2 = .430$, *F2*(1, 38) = 17.984, $p < .001$, $\eta_p^2 = .321$; seven-letter words received longer fixation durations than four-letter words (302 vs. 270 ms).

Total fixation duration

The NrL had a main effect, *F1*(1, 32) = 20.515, $p < .001$, $\eta_p^2 = .391$, *F2*(1, 38) = 9.730, $p = .003$, $\eta_p^2 = .204$, seven-letter words received longer fixation durations than four-letter words (339 vs. 387 ms). The main effect of Group, *F1*(1, 32) = 8.189, $p = .007$, $\eta_p^2 = .204$, *F2*(1, 38) = 96.930, $p < .001$, $\eta_p^2 = .718$, indicated that overall the DYS group looked at the target words longer than the controls (409 vs. 317 ms).

Refixation probability

The main effect of NrL, *F1*(1, 32) = 14.998, $p = .001$, $\eta_p^2 = .319$, *F2*(1, 38) = 10.327, $p = .003$, $\eta_p^2 = .214$, was accompanied by the two-level interaction of Width x NrL, *F1*(1, 32) = 6.226, $p = .018$, $\eta_p^2 = .163$, *F2*(1, 38) = 2.733, $p = .107$, $\eta_p^2 = .067$. This interaction resulted from the increase in refixation probability, when spatial width was not controlled for (from .25 to .38; $t(33) = -4.40$, $p < .001$) but not when it was controlled (from .32 to .36; $t(33) = -1.67$, $p = .104$). The main effect of Group, $F1(1, 32) = 5.851$, $p = .021$, $\eta_p^2 = .155$, $F2(1, 38) = 44.129$, $p < .001$, $\eta_p^2 = .537$, indicated that overall the DYS group made more refixations than the controls (.39 vs .27, respectively).

Relative landing position

The main effect of the NrL, $F1(1, 32) = 60.548$, $p < .001$, $\eta_p^2 = .654$, $F2(1, 38) = 51.577$, $p < .001$, $\eta_p^2 = .576$, was accompanied by the two-level interaction of Width x NrL, $F1(1, 32) = 23.872$, $p < .001$, $\eta_p^2 = .427$, $F2(1, 38) = 10.964$, $p = .002$, $\eta_p^2 = .224$. This interaction resulted from a shift in the relative landing position to the word beginning in seven- vs. four-letter words when spatial width was not controlled for (from 58% to 46%; 9.08, $p < .001$), relative to when it was controlled (from 56% to 52%; 4.09, $p < .001$).

First-pass skipping probability

The main effect of the NrL, $F1(1, 32) = 38.991$, $p < .001$, $\eta_p^2 = .549$, $F2(1, 38) = 220.850$, $p < .001$, $\eta_p^2 = .853$, was accompanied by two-level interaction of Width x NrL, $F1(1, 32) = 29.165$, $p < .001$, $\eta_p^2 = .477$, $F2(1, 38) = 76.681$, $p < .001$, $\eta_p^2 = .669$. This interaction resulted from the larger decrease in the probability of skipping as a function of the NrL when spatial width was not controlled for (.27 to .02; $t(33) = 7.42$, $p < .001$) relative to when it was controlled (.12 vs .05; $t(33) = 3.44$, $p = .002$).

Ten point two percent of the word skips to narrow words were likely overshoots shown as missing tails in landing position distribution. Although this value was subtracted from the skipping probability in the narrowest word category (narrow four- to five-letter words), the effect of spatial width on word skipping remained highly significant, $F(1,34) = 14.6$, $p = .001$, $\eta_p^2 = .307$, when compared to wide four- and five-letter words.

Summary

There was a reliable NrL effect of 18 ms per letter on the total fixation duration measure, and a 10 ms per letter -effect in single fixation duration -measure, yet this latter effect was only a trend-like in item-analysis. Skipping and refixation probability and landing position were clearly affected by spatial width, with the exception that the effect in refixation probability was not significant in item analysis. These variables were also slightly affected by the NrL, since words with a higher NrL and wider words were landed more toward the word beginning, refixated more often and skipped less often than narrower or shorter words. The DYS group was associated with longer total fixation duration and more frequent refixations.

The effect of spatial width

Single fixation duration

In five-letter words, the main effect of Width, $F1(1, 32) = 14.837$, $p = .001$, $\eta_p^2 = .317$, $F2(1, 19) = 15.744$, $p = .001$, $\eta_p^2 = .453$, resulted from the longer fixation duration on narrow (272 ms) vs. wide words (241 ms).

Gaze duration

In five-letter words, the main effect of Group, *F1*(1, 32) = 5.621, *p* = .024, η_p^2 = .149, *F2*(1, 19) = 33.649, *p* < .001, η_p^2 = .639, resulted from longer durations for the DYS group (291 ms) relative to the controls (243 ms).

Total fixation duration

In five-letter words, the main effect of Width, *F1*(1, 32) = 15.561, *p* < .001, η_p^2 = .327, *F2*(1, 19) = 7.540, *p* = .013, η_p^2 = .284, resulted from longer durations for narrow (366 ms) vs. wide (312 ms) words. The main effect of Group, *F1*(1, 32) = 9.755, *p* = .004, η_p^2 = .234, *F2*(1, 19) = 24.572, *p* < .001, η_p^2 = .564, resulted from longer durations for the DYS group relative to the controls (379 vs. 298 ms). In six-letter words, the main effect of Width, *F1*(1, 32) = 7.059, *p* = .012, η_p^2 = .181, *F2*(1, 19) = 2.022, *p* = .171, η_p^2= .096, resulted from longer durations for narrow (384 ms) vs. wide words (357 ms).

Regression probability

In five-letter words, the main effect of Width, *F1*(1, 32) = 23.635, *p* < .001, η_p^2 = .425, *F2*(1, 19) = 8.385, *p* = .009, η_p^2 = .306, resulted from the higher regression probability for narrow (.17) vs. wide (.08) words. In six-letter words, the main effect of Width, *F1*(1, 32) = 12.165, *p* = .001, η_p^2 = .275, *F2*(1, 19) = 7.283, *p* = .014, η_p^2 = .277, resulted from the higher regression probability for narrow (.17) vs. wide (.09) words.

The narrow words might have been regressed because they were skipped more often (20.5% vs. 10.6% of trials). However, the same pattern of results was present when the analysis of the regression probability was restricted to non-skipping trials, which suggests that the regression probability findings were not only consequences of skipping for five-letter words, *F1*(1, 32) = 17.1, $p < .001$, η_p^2 = .341, and six-letter words, *F1*(1, 32) = 13.7, $p = .001$, η_p^2 = .294.

Relative landing position.

In five letter words, the effect of width only approached significance, *F1*(1, 32) = 3.940, $p = .056$, η_p^2 = .110, *F2*(1, 38) = 4.228, $p = .054$, η_p^2 = .182. Wider words were landed more towards word beginning (55 %) than narrow words (59 %).

In six letter words, the effect of width was highly significant, *F1*(1, 32) = 40.612, p < .001, η_p^2 = .559, *F2*(1, 38) = 15.9, $p = .001$, η_p^2 = .456. Wider words were landed more towards word beginning (50 %) than narrow words (55 %).

First-pass skipping probability

In five letter words, the effect of width was highly significant, *F1*(1, 32) = 29.915, $p < .001$, η_p^2 = .483, *F2*(1, 38) = 68.413, $p < .001$, η_p^2 = .783. Narrow words were skipped more often than wide words (0.26 vs. 0.09, respectively)

In six letter words, the effect of width was significant, *F1*(1, 32) = 5.390, $p = .027$, η_p^2 = .144, *F2*(1, 38) = 7.255, $p = .014$, η_p^2 = .276. Narrow words were skipped more often than wide words (0.8 vs. 0.04, respectively)

Summary

Narrow vs. wide five-letter words were fixated on longer in the single fixation duration and total fixation duration measures and regressed more often, and narrow vs. wide six-letter words were regressed more often and had longer total fixation durations in subject but not in item-analysis. However, these effects were substantially weaker than in five-letter words. The DYS group was associated with overall longer fixation durations in five-letter words in the gaze duration and total fixation duration measures. Similarly what was found in orthogonal analysis of spatial width and NrL, skipping probability and landing position were strongly governed by spatial extent of words.

Discussion

We studied how spatial width and the NrL in the target words affect eye movement measures during typical and dysfluent reading. We found that the total fixation durations (the sum of the durations of all fixations on a word) on narrow words were somewhat higher than those for wide words, which suggests that visual crowding in foveal vision may increase fixation durations, not visual acuity limitations, when individuals read relatively short words fit mostly on foveal vision (only the wide six- and seven-letter words were wider spatially by more than two degrees). This finding is in line with recent findings that visual crowding is an important factor in reading (Martelli et al., 2009; Pelli et al., 2007; Perea & Gomez, 2012; Slattery & Rayner, 2013; Zorzi et al., 2012). The crowding in our data partially resulted from more frequent regressions to narrow words, which suggests that crowding disrupts word recognition during first-pass reading, and thus, the readers had to return to check the word. This finding agrees with two recent reports on increased regression rates for text presented in decreased letter spacing (Rayner et al., 2010; Slattery & Rayner, 2013).

The crowding effect on fixation durations was especially strong in five-letter words but much weaker in six-letter words. Narrow five-letter words may have been perceptually the most challenging because they consisted of similar visual letters with high spatial frequency, for example*, tilli* (dill), whereas the most narrow six-letter words were somewhat less crowded because they consisted of letters of more variable width, for example*, piikki* (spike). This suggests that not only the letter spacing but possibly also the distance between letter center points or, in general, the overall discriminability of letters within words counts when it comes to visual processing of words. This view is in accordance with the recent understanding of the multilevel nature of crowding (Whitney & Levi, 2011). Overall, we conclude that the effect of visual crowding can be substantial, but only among very densely packed words. Further, visual and lexical processes may also interact if the visually challenging word has visually similar word neighbors, such as *tilli/tiili* (dill/brick). Although the number of orthographic neighbors between narrow and wide words was controlled in this study, future studies should investigate additional specific interplays between visual crowding and lexical processes.

In line with our previous study (Hautala et al., 2011a), there was a consistent NrL effect in summative fixation duration measures. In gaze duration, there was an NrL effect but no crowding effect, indicating that these effects can occur independently of each other. However, the NrL effect on words with equal spatial width can result from crowding since there are more objects in the given space, while the NrL effect on words with various spatial widths may result from refixations. Therefore, the single fixation duration measure may be the purest measure for comparing crowding and NrL effects. This is justified also from the viewpoint that visual effects should appear early during the time course of word processing, and therefore be present already in single-fixation duration. The results in this variable indicated the NrL effects were similar irrespective of whether the spatial width was controlled (a crowded condition) or not, suggesting that this effect was due to the NrL, not crowding. The temporal word length effect resulted mainly from the genuine NrL effect, whereas particularly crowded words seemed to provide an extra visual challenge for a reader. Thus, we believe letters are important attentional units in reading. However, letters of very familiar words may be processed in parallel (Coltheart et al., 2001; Perry et al., 2007); thus, the NrL effect may be even absent (see Hawelka et al., 2010 for a recent eye movement study).

The saccadic measures were heavily influenced by the spatial width of the word but also slightly by the NrL. The landing position shifted to the word beginning, refixations were more frequent and skipping was less frequent for wider and longer (NrL) words. Generally, this pattern of findings is in line with the view that spatial instead of linguistic information is predominantly used for saccade targeting while linguistic information is mainly used for word recognition processes (Inhoff et al., 2003). However, this finding suggests that this functional dissociation is not all or nothing, but linguistic-attentional demands of the upcoming word are used to fine-tune the saccade targeting. Since the number of letters is a factor in word recognition that consistently affects fixation times, the saccadic system is also affected by this processing demand to some extent: Words with fewer NrL could be more easily recognized parafoveally and thus skipped, and landing more toward the word beginning of words with more letters prioritizes processing of the word beginning (Hautala & Parviainen, 2014) and leaves more space for progressive refixation saccades (Vergilino-Perez et al., 2004).

The strong influence of spatial width on saccadic measures is most likely caused by visual acuity limitations. If a word is spatially wide, the initial saccade lands more toward the word beginning to provide a high-quality visual sample of the word beginning, while refixations may be done to provide a higher-quality visual sample of the word ending. Similarly, very narrow words may be skipped because they are within reach of the foveal vision from the previous fixation location and could therefore be parafoveally identified. Such identification, however, is futile as shown by generally more frequent regressions to skipped words. Further, the effect of spatial width on the skipping rate depends on some currently unspecified typographical factor, since Slattery and Rayner (2013) did not find clear effects of increased or decreased letter spacing (which affects words' spatial width) on the skipping rate, yet skipping occurred more frequently on a spatially more condensed proportional vs. spatially wider monospaced font in their study. The finding that spatial width has an important role in parafoveal processing in landing position and skipping (see also Hautala et al., 2011; McDonald, 2006) contradicts the view that the parafoveal preprocessing of words is mainly limited by attention, not visual acuity (Miellet et al., 2009; Schotter et al., 2012). Our results do not favor any models of eye movement control in reading (Engbert et al., 2005; Reilly & Radach, 2006; Reichle et al., 2006) but suggest that visual acuity limitations should be stressed over attentional limitations in processing of upcoming words.

In regards to reading fluency, dysfluent readers had overall longer fixation durations, and made more refixations. Since reading speed was not consistently associated with the effect of spatial width on temporal eye movement measures, we conclude that crowding was not related to reading fluency in our sample of adult readers. In contrast to several developmental studies (De Luca et al., 1999; Hautala et al., 2011b; Hutzler & Wimmer, 2004; Hyönä & Olson, 1995), but in line with a study with adult readers with dyslexia (Hawelka et al., 2010), dysfluent readers showed only insignificant trends toward a larger NrL effect. However, this is not to say that specific visual or letter-processing deficits could not be found in a subgroup of individuals with dyslexia suffering from problems in visual processing (Bellochi et al., 2013) or generally in readers who are more affected than the dysfluent readers studied here. In more severely affected readers, disturbances in letter processing are more likely to be detected (Moll, Hutzler, & Wimmer, 2005).

In conclusion, the present results support the view that letters are important attentional units in processing of foveally fixated words while visually very crowded words require longer viewing time to be correctly identified and still must be regressed in some cases. However, neither the NrL nor the crowding effect was associated with reading fluency in our data. Our results also provide strong new evidence that all saccadic behavior is more strongly governed by a visual (spatial width) rather than attentional (NrL) factor. Very narrow words within the reach of acuity vision while fixating on the preceding word can be parafoveally identified and skipped, while spatially wide words are landed on more toward the word beginning and refixated more often, presumably to attain a high-quality visual sample of the word beginning and end.

Acknowledgements

The authors would like to thank Jukka Hyönä for his comments and Asko Tolvanen for statistical consulting concerning this work. The working of JH was supported by the Academy of Finland grants nr. 269102 for years 2013-2014, and grants nr. 274022 and 274050 for year 2015. The working of OL was supported by the grant from Eino

Jutikkala foundation governed by Finnish Academy of Science and Letters, and Grant from Department of Psychology, University of Jyväskylä.

References

Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D., & Yap, M. (2004). Visual word recognition for single-syllable words. *Journal of Experimental Psychology: General, 133,* 283-316. doi:10.1037/0096-3445.133.2.283

Barton, J. J., Hanif, H. M., Eklinder, Björnström, L., & Hills, C. (2014). The word-length effect in reading: A review. *Cognitive neuropsychology*, *31*(5-6)*,* 378-412. doi: 10.1080/02643294.2014.895314

Bellocchi, S., Muneaux, M., Bastien-Toniazzo, M., & Ducrot, S. (2013). I can read it in your eyes: What eye movements tell us about visuo-attentional processes in developmental dyslexia. *Research in developmental disabilities*, *34*(1), 452-460. doi: doi:10.1016/j.ridd.2012.09.002

Bosse, M., Tainturier, M. & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104,* 198-230. doi: doi:10.1016/j.cognition.2006.05.009

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177-178. 10.1038/226177a0

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research, 13,* 767-782. doi:10.1016/0042-6989(73)90041-2

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204-256. doi:10.1037/0033-295X.108.1.204

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. doi:10.1016/S0028-3932(99)00038-X

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813. doi: 10.1037/0033-295X.112.4.777

Gori, S., & Facoetti, A. (2015). How the visual aspects can be crucial in reading acquisition? The intriguing case of crowding and developmental dyslexia. *Journal of vision*, *15*(1), 8. doi: 10.1167/15.1.8

Hautala, J., & Parviainen, T. (2014). Gaze Position Reveals Impaired Attentional Shift during Visual Word Recognition in Dysfluent Readers. *PloS one*, *9*(9). doi:10.1371/journal.pone.0108937

Hautala, J., Aro, M., Eklund, K., Lerkkanen, M. K., & Lyytinen, H. (2013). The role of letters and syllables in typical and dysfluent reading in a transparent orthography. *Reading and Writing*, *26*(6), 845-864. doi:10.1007/s11145-012-9394-3

Hautala, J., Hyönä, J. & Aro, M. (2011a). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Research*, *51*, 1719–1727. doi: 10.1016/j.visres.2011.05.015

Hautala, J., Hyönä, J., Aro, M., & Lyytinen, H. (2011b). Sublexical effects on eye movements during repeated reading of words and pseudowords in Finnish. *Psychology of Language and Communication*, *15*(2), 129-149. doi: 10.2478/v10057-011-0009-x

Hawelka, S., Gagl, B., & Wimmer, H. (2010). A dual-route perspective on eye movements of dyslexic readers. *Cognition, 115,* 367-379. doi:10.1016/j.cognition.2009.11.004

Hutzler, F. &Wimmer, H. (2004) Eye movements of dyslexic children when reading in a regular orthography. *Brain and Language*, *89*, 235–242. doi:10.1016/S0093-934X(03)00401-2

Huurneman, B., Boonstra, F. N., Cox, R. F., Cillessen, A. H., & van Rens, G. (2012). A systematic review on 'Foveal Crowding'in visually impaired children and perceptual learning as a method to reduce Crowding. *BMC ophthalmology*,*12*(1), 27. doi:10.1186/1471-2415-12-27

Hyönä, 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, and Cognition, 21,* 1430-1440. doi:10.1037/0278-7393.21.6.1430

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. doi:10.1016/j.jecp.2008.04.002

Inhoff, A., Eiter, B., Radach, R., & Juhasz, B. (2003). Distinct subsystems for the parafoveal processing of spatial and linguistic information during eye fixations in reading. *The Quarterly Journal of Experimental Psychology Section A*, *56*(5), 803-827. doi: 10.1080/02724980244000639

Language Bank of Finland. Frequency Lexicon of the Finnish Newspaper Language. Retrieved from: http://urn.fi/urn:nbn:fi:lb-201405271

Levi, D. M., Klein, S. A., & Hariharan, S. (2002). Suppressive and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking. *Journal of Vision, 2,* 140-166. doi:10.1167/2.2.3

Martelli. M, Di Filippo. G, Spinelli, D & Zoccolotti. P. (2009). Crowding, reading, and developmental dyslexia. *Journal of Vision*, *9*, 14. doi:10.1167/9.4.14

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578-586. doi:10.3758/BF03203972

McDonald, S. A. (2006). Effects of number-of-letters on eye movements during reading are independent from effects of spatial word length. *Visual Cognition, 13,* 89-98. doi: 10.1080/13506280500143367

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification visual acuity does not modulate the perceptual span in reading. *Psychological Science*, *20*(6), 721-728. doi: 10.1111/j.1467-9280.2009.02364.x

Moll, K., Hutzler, F., & Wimmer, H. (2005). Developmental dyslexia in a regular orthography: A single case study. *Neurocase*, *11*(6), 433-440. doi: 10.1080/13554790500263537

Morrison, R. E., & Rayner, K. (1981). Saccade size in reading depends upon character spaces not visual angle. *Perception and Psychophysics, 30*, 395-396. doi: 10.3758/BF03206156

New, B. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*(1), 45-52. doi: 10.3758/BF03193811

Nevala, J., Kairaluoma, L., Ahonen, T., Aro, M., & Holopainen, L. (2007). Lukemis- ja kirjoittamistaitojen yksilötestistö nuorille ja aikuisille. [Assessment Battery for Reading Disabilities in Young and Adults] Jyväskylä: Niilo Mäki Institute.

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of vision*, *7*(2), 20. doi:10.1167/7.2.20

Perea, M., & Gomez, P. (2012). Subtle increases in interletter spacing facilitate the encoding of words during normal reading. *PloS one*, *7*(10), doi:10.1371/journal.pone.0047568

Perea, M., Panadero, V., Moret-Tatay, C., & Gómez, P. (2012). The effects of inter-letter spacing in visual-word recognition: Evidence with young normal readers and developmental dyslexics. *Learning and Instruction*, *22*(6), 420-430. doi:10.1016/j.learninstruc.2012.04.001

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud.*Psychological review*, *114*(2), 273. doi: 10.1037/0033-295X.114.2.273

Raven, J., Raven, J. C., & Court, J. H. (1998). Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3, The Standard Progressive Matrices. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Rayner, K. (1983). Eye movements, perceptual span, and reading disability. *Annals of Dyslexia, 33,* 163-173. doi: 10.1007/BF02648003

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

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review, 17,* 834-839. doi: 10.3758/PBR.17.6.834

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7,* 4-22. doi:10.1016/j.cogsys.2005.07.002

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*(1), 34-55. doi:10.1016/j.cogsys.2005.07.006

Schmitz, C. (2010). Limesurvey [computer software]. Retrieved from: http://www.limesurvey.org

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, *74*(1), 5-35. doi: 10.3758/s13414-011-0219-2

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151-218. doi:10.1016/0010-0277(94)00645-2

Slattery, T. J., & Rayner, K. (2013). Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text. *Attention, Perception, & Psychophysics*, *75*(6), 1275-1292. doi: 10.3758/s13414-013-0463-8

Vergilino-Perez, D., Collins, T., & Doré-Mazars, K. (2004). Decision and metrics of refixations in reading isolated words. *Vision research*, *44*(17), 2009-2017. doi:10.1016/j.visres.2004.03.012

Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, *15*(4), 160-168. doi:10.1016/j.tics.2011.02.005

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14,* 1-33. doi:10.1017/S0142716400010122

Wimmer, H. (1996a). The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. *Journal of Experimental Child Psychology, 61,* 80-90. doi:10.1006/jecp.1996.0004

Wimmer, H. (1996b). The early manifestation of developmental dyslexia: Evidence from German children. *Reading and Writing, 8*, 171-188. doi:10.1007/BF00555368

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology, 86,* 169-193. doi:10.1016/S0022-0965(03)00139-5

Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., & Martelli, M. (2009). Reading development in an orthographically regular language: Effects of length, frequency, lexicality and global processing ability. *Reading and Writing*, *22*(9), 1053-1079. doi:10.1007/s11145-008-9144-8

Zoccolotti. P., De Luca. M., Di Pace. E., Gasperini, F. & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and language*, *93*, 369-373. doi:10.1016/j.bandl.2004.10.010

Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., ... & Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences*, *109*(28), 11455-11459. doi:10.1073/pnas.1205566109

Appendix. List of stimuli sentences with their alternative target words and sentence endings, separated by bolding and backlash.

Pair 1 - wide four letter vs. narrow seven letter words.
Tummuneiden ikkunoiden vuoksi **aamu näytti synkemmältä kuin se oli** / keittiö näytti synkemmältä kuin se oli
Matin toissapäivänä ostaman uuden talon **amme osoittautui hyvin suureksi /** kellari osoittautui hyvin suureksi
Vaarini kertoo, että vanhoina hyvinä aikoina **hame valmistettiin pehmeästä silkistä** / laituri valmistettiin paksusta männystä
Toistuvista poistoyrityksistä huolimatta **home olikin tarttunut seinään pysyvästi** / juliste olikin tarttunut seinään pysyvästi
Syksyn kylmyyden yllättäessä **hymy hyytyi Pekan huulilta /** viisari hyytyi Pekan mittarista
Pirteän kummipoikani piirtämä **maha sattui olemaan varsin suuri /** kirjain sattui olemaan varsin koukeroinen
Johtokunnan suosituksesta epäonnistunut **maku jätettiin kehitysasteelle** / poliisi jätettiin eläkkeelle tehtävistään
Luonnontieteellisessä museossa **mato herätti eniten kiinnostusta /** tiikeri herätti eniten kiinnostusta
Onnistuneen markkinoinnin vuoksi **mehu myytiin nopeasti** /liiteri myytiin nopeasti
Illasta erittäin railakkaaksi yltynyt **meno loppui poliisien saapuessa/** tiistai loppui poliisien saapuessa
Huonon kuntonsa ja parittomuutensa vuoksi **mono heitettiin roskalavalle/** luistin heitettiin roskalavalle
Poliisin huomion kohteeksi joutunut **mopo päätettiin lopulta pysäyttää/** turisti päätettiin lopulta pidättää
Mitä suurimmaksi epäonnekseni **muna rikkoutui tiputtuaan pöydältä**/ kiikari rikkoutui tiputtuaan pöydältä
Epähuomiossa lattialle jäänyt **muru päätyi hiiren vatsaan/** siivilä päätyi kirpputorille
Huomasin, että aamiaisena pelkkä **namu jättäisi ankean nälän tunteen**/ pilleri jättäisi ankean nälän tunteen
Tilanteen ollessa erittäin kriittinen **oksa varoitti räsähdyksellä katkeamisestaan /** juristi varoitti asiakastaan puhumasta
Tarkastuksessa omituisen näköinen **papu herätti erityistä huomiota/** siviili herätti erityistä huomiota
Tehokkaan suoriutumisensa vuoksi **pomo palkittiin bonuksilla/** tutkija palkittiin bonuksilla
Lukuisien vuosien saatossa **sumu alkoi vähentyä kylän mailta/** viljely alkoi vähentyä kylän mailta
Tyhmän ja ajattelemattoman käskyn aiheuttama **uhma yllätti johtajiston täysin**/ riitely yllätti johtajiston täysin

Pair2 - narrow four letter vs. wide seven letter words.
Kovassa kuumuudessa työskennellessä **hiki pääsi yllättämään**/ vahinko pääsi yllättämään
Pienen tovin kuluttua olikin **ilta koittanut juhlissamme/** kuolema koittanut potilaalle
Vastoin aiempaa käsitystäni **joki kuuluu kansalliseen suojeluohjelmaan**/ maisema kuuluu kansalliseen suojeluohjelmaan
Biologien tutkimusten mukaan **keli vaikuttaa nopeasti kasvien kasvuun**/ kosteus vaikuttaa nopeasti kasvien kasvuun
Jännittävässä kilpailussa viimeinen **kilo osoittautui raskaimmaksi /** laukaus osoittautui raskaimmaksi
Tiedemiesten laskelmien mukaisesti **kivi osoittautui varsin painavaksi /** aurinko osoittautui varsin vanhaksi
Yleensä vilkkailla markkinoilla jokainen **kori saadaan myytyä /** makkara saadaan myytyä
Ammattilaisten mukaan tasapainoinen **koti vaikuttaa lapseen positiivisesti** / rakkaus vaikuttaa lapseen positiivisesti

Taiteilijan kuuluisimmassa maalauksessa **kylä lepää tunturin juurella /** hevonen lepää tunturin juurella
Liikuntakeskuksen runsaasti käytetty **latu vaatii paljon huoltoa** / katsomo vaatii säännöllistä huoltoa
Nuorien vieraiden kovassa käytössä **lelu yllättäen hajosi** / summeri yllättäen hajosi
Viime perjantaina kovassa kiireessä **liha unohtui viikonlopuksi pöydälle** / hakemus unohtui viikonlopuksi pöydälle
Taitavan ruuanlaittajan käsissä **lohi muuttuu herkulliseksi ruuaksi** / hedelmä muuttuu herkulliseksi jälkiruuaksi
Ollin mielestä vaalean keltainen **olut sattuu olemaan maukkain /** kuutamo sattuu olemaan kaunein
Perinteinen Afrikan tähti olikin **peli parhaimmasta päästä!** / kokemus parhaimmasta päästä!
Oppikirjojen mukaan muinaisina aikoina **piru aiheutti runsaasti pelkoa** / ukkonen aiheutti runsaasti pelkoa
Useiden tuntien etsinnän jälkeen **raja löytyi metsän keskeltä /** vastaus löytyi tietosanakirjasta
Markkinoilla palkintoja voittanut **sika myytiin ennätyshintaan** /maalaus myytiin ennätyshintaan
Päätöksen mukaisesti puolueen **väri vaihdettiin toiseen** / ehdokas vaihdettiin toiseen
Heti alusta asti tiesin ettei yleisin **öljy sopisi autoosi!** /näkemys sopisi yleisöllesi!

Pair3 - wide vs. narrow five letter words.
Olohuoneen seinässä näytti olevan **aukko johtuen rikkinäisestä digikamerasta** /pilvi johtuen roskasta dikikameran linssissä
Matti tiesi, että valtion lakiin säädetty **eläke tuli maksaa aina kuun alussa** / tulli tuli maksaa aina yli viiden kilon kuormasta
Juhlissa tarjottavaksi varattu **kahvi loppui nopeasti kuokkavieraiden saavuttua** /viini loppui nopeasti kuokkavieraiden saavuttua
Karjalle tuhoisin eläin näyttäisi olevan **karhu aiheuttamiensa vammojen vuoksi** / hiiri aiheuttamiensa tautien vuoksi
Rakennusmestarin mukaan myrskyn rikkoma **katto täytyisi kunnostaa mahdollisimman pian** / tiili täytyisi vaihtaa mahdollisimman pian
Perinteeksi muotoutunut jokavuotinen **kerho tuottaa merkittävästi varoja kyläyhteisölle** / ralli tuottaa merkittävästi varoja kyläyhteisölle
Menneisyydessä kotien lämmönlähteenä käytetty **koivu tuotti paljon energiaa** / hiili tuotti paljon hiukkaspäästöjä
Lopulta osallistuneiden kyllästyttyä **lakko purkautui kaikessa hiljaisuudessa** /piiri purkautui eri puolille pihaa
Päinvastaisista odotuksista huolimatta **lento olikin nopeasti ohi** /leiri olikin nopeasti ohi
Ostoslistan hukkumisen vuoksi **maito unohtui kauppaan, mikä myöhemmin harmitti** /riisi unohtui kauppaan, mikä myöhemmin harmitti
Suuressa punaisesta laatikosta löytyi **matto vaikka Matti muuta väitti** /siili vaikka Matti muuta väitti
Myrskyn aikana seurakunnan varjeltu **metsä kärsi merkittäviä vaurioita /** risti kärsi merkittäviä vaurioita
Aamulla löysin pienen linnun **niska vaurioituneena ikkunan alta** / siipi vaurioituneena ikkunan alta
Tuomariston mukaan paras lastenkirja olikin **paksu satukirja hattupäisen kissan seikkailuista** /villi satukirja hattupäisen kissan seikkailuista
Maantieteellisen läheisyyden vuoksi **perhe asettui länsirannikolle** / kieli asettui länsirannikolle
Pojallesi eilen ostama punainen **pyörä pitää kammottavaa meteliä /** pilli pitää kammottavaa meteliä
Aikakausilehden tilaajalahjana tullut **reppu tuotti karvaan pettymyksen** / liite tuotti karvaan pettymyksen
Huomasin kesäkuun loppuessa, että **ruoho olikin jo polvenkorkuista** / riihi olikin yllättäen luhistunut
Tarkoista varotoimista huolimatta **ruoka meni pilalle loppuviikosta** / liike meni konkurssiin loppuviikosta

Pair 4 - wide vs. narrow six letter words.
Heti seuraavalla hetkellä **silmä tarkentui merellä kulkevaan laivaan** / litra marjoja oli hävinnyt parempiin suihin
Eilen illalla Maunon kipeytynyt **hammas alkoikin särkemään entistä enemmän** / lantio alkoikin särkemään entistä enemmän
Ruokaa tuottavan yrityksen kallis **kamera rikottiin yön aikana** / portti rikottiin yön aikana
Maija yllättyi huomatessaan, että **kangas täytyisi viedä pesulaan kuivapesuun /** turkki täytyisi viedä pesulaan puhdistettavaksi
Matin vanhassa kotitalossa odotti **karmea kellari hämähäkkeineen** / tilava kellari hämähäkkeineen
Kuulin, että Helsingin pörssissä **kauppa loppui jännittävässä vaiheessa** /viikko loppui jännittävässä vaiheessa
Pitkän ja kuuman työpäivän jälkeen **kuorma osoittautui liian raskaaksi** / leikki osoittautui liian raskaaksi
Toisin kuin huhu kertoi, maatalosta ostettu **lammas olikin kovin kallista** / sokeri olikin kovin kallista
Kylän rautakaupasta löytyi halpa **lamppu keittiöstä rikkoutuneen tilalle** / veitsi keittiöstä hukkuneen tilalle
Läheisen baarin asiakaskunnan mukaan **markka menetettiin poliittisen huijauksen takia** / voitto menetettiin ensimmäisessä erässä
Kokin viime tingassa lisäämä **mauste valitettavasti pilasi ruokalajin** / sipuli valitettavasti pilasi ruokalajin
Odotuksien mukaisesti herkullinen **munkki nousi kahvipisteen hittituotteeksi** / keitto nousi ruokapisteen hittituotteeksi
Maaliskuussa aavikolta esiin kaivettu **muumio tuotti museolle hyvät kävijämäärät** / piikki tuotti nuorelle arkeologille pettymyksen
Kuvaamataidon tunnilla askarreltiin **naamio Kaapon suureksi harmiksi** / lautta Kaapon suureksi harmiksi
Yöllisen tietokonevian vuoksi Leenan **numero katosi väliaikaisesti luettelosta** / osoite katosi väliaikaisesti luettelosta
Elinan isoisän mielipiteen mukaan **näkymä olisi kerrassaan upea** / kielto olisi kerrassaan raivostuttava
Muistatko, kun viime kesänä **pumppu kuulosti kuin se olisi ollut hajoamassa** / soitto kuulosti kuin se olisi tullut nauhalta
Yllättäen kovan äänen kuuluessa **salama räjähti takapihalla rikkoen keinun** / kallio räjähti takapihalla rikkoen keinun
Lehden mukaan näytelmän laimea **sanoma tuotti yleisölle suuren pettymyksen** / tanssi tuotti yleisölle suuren pettymyksen
Kuulin. että maailmanmestarin eilinen **temppu yllätti jokaisen katsojan** / ottelu yllätti jokaisen katsojan
Valheellisissa lupauksissa tarjottu **vapaus sattuu olemaan yleinen palkinto höynäytetylle** / taivas sattuu olemaan yleinen palkinto höynäytetylle

Appendix B.

Complete ANOVA tables for each of the dependent variables. Abbreviations: W = Width, NrL = Number of Letters, G = Group. Degrees of freedom are 1, 32 for all subject analyses (F1), and 1, 38 for item (F2) analysis of NrL x Width, and 1, 19 for item analysis of spatial width in five and six letter words.

Single Fixation Duration

Analysis	Factor	*F1*	*p*	η_p^2	*F2*	*p*	η_p^2
NrL x W	NrL	8.264	.007	.205	3.196	.082	.078
	G	3.387	.075	.096	65.810	.000	.634
	WxNrL	0.498			1.395	.245	.035
	WxG	0.871			1.915	0.174	0.048
	NrLxG	0.114			1.007	0.322	0.026
	WxNrLxG	2.890	.099	.083	2.596	.115	.064
Five-letter	W	14.837	.001	.317	15.744	0.001	0.453
	G	2.585	.118	.075	17.090	.001	.474
	WxG	2.009	.166	.059	3.353	0.083	0.15
Six-letter	W	0.137			0.048		
	G	1.624	0.212	0.048	18.274	0	0.49
	WxG	0.927			1.257	0.276	0.062

Gaze duration

Analysis	Factor	*F1*	*p*	η_p^2	*F2*	*p*	η_p^2
NrL x W	NrL	24.112	.000	.430	17.984	.000	.321
	G	3.923	.056	.109	89.387	.000	.702

	WxNrL	0.218			0.057		
	WxG	0.650			0.007		
	NrLxG	0.253			0.106		
	WxNrLxG	2.544	.121	.074	4.178	.048	.099
Five-letter	W	1.856	.183	.055	1.763	.200	.085
	G	5.621	.024	.149	33.649	.000	.639
	WxG	0.438			0.452		
Six-letter	W	1.880	.180	.055	0.727		
	G	3.198	.083	.091	26.556	0	0.583
	WxG	0.313			0.418		

Total fixation duration

Analysis	Factor	*F1*	*p*	η_p^2	*F2*	*p*	η_p^2
NrL x W	NrL	20.515	.000	.391	9.730	.003	.204
	G	8.189	.007	.204	96.930	.000	.718
	WxNrL	0.009	.924	.000	0.049		
	WxG	0.250	.876	.001	0.088		
	NrLxG	1.223	.277	.037	0.455		
	WxNrLxG	1.140	.294	.034	1.384	.247	.035
Five-letter	W	15.561	.000	.327	7.540	.013	.284
	G	9.755	.004	.234	24.572	.000	.564
	WxG	0.020			0.113		
Six-letter	W	7.059	.012	.181	2.022	.171	.096
	G	4.110	.051	.114	40.618	.000	.681

	WxG	0.564			0.243		

First-pass skipping probability

Analysis	Factor	*F1*	*p*	η_p^2	*F2*	*p*	η_p^2
NrL x W	NrL	38.991	.000	.549	220.850	.000	.853
	G	0.312			2.579	.117	.064
	WxNrL	29.165	.000	.477	76.681	.000	.669
	WxG	1.702	.201	.050	1.585	.216	.040
	NrLxG	0.377			2.463	.125	.061
	WxNrLxG	0.044			0.139		
Five-letter	W	29.915	.000	.483	68.413	.000	.783
	G	0.331			1.833	.192	.088
	WxG	0.514			1.138	.300	.056
Six-letter	W	5.390	.027	.144	7.255	.014	.276
	G	0.142			1.058	.317	.053
	WxG	0.258			0.359		

Refixation probability

Analysis	Factor	*F1*	*p*	η_p^2	*F2*	*p*	η_p^2
NrL x W	NrL	14.998	.001	.319	10.327	.003	.214
	G	5.851	.021	.155	44.129	.000	.537
	WxNrL	6.226	.018	.163	2.733	.107	.067
	WxG	0.002			0.022		
	NrLxG	0.054			0.027		
	WxNrLxG	1.583	.217	.047	1.415	.242	.036

Five-letter	W	2.218	.146	.065	1.262	.275	.062
	G	3.326	.078	.094	5.877	.025	.236
	WxNrL	0.757			.944		
Six-letter	W	1.438	.239	.043	0.799		
	G	1.214	.279	.037	5.659	.028	.229
	WxNrL	0.251			0.198		

Regression probability

Analysis	Factor	*F1*	*p*	η_p^2	*F2*	*p*	η_p^2
NrL x W	G	0.476			1.522	0.255	0.039
	NrL	2.169	0.151	0.063	1.55	0.221	0.039
	WxNrL	2.504	0.123	0.073	1.466	0.233	0.037
	WxG	1.898	0.178	0.056	1.698	0.2	0.43
	NrLxG	0.011			0.004		
	WxNrLxG	0.102			0.221		
Five-letter	W	23.635	.000	.425	8.385	.009	.306
	G	1.359	.252	.041	2.014	.172	.096
	WxG	0.089			0.134		
Six-letter	W	12.165	.001	.275	7.283	.014	.277
	G	0.004			0.002		
	WxG	0.294			0.454		

Relative landing position

Analysis	Factor	*F1*	*p*	η_p^2	*F2*	*p*	η_p^2
NrL x W	NrL	60.548	.000	.654	51.577	.000	.576

	G	0.007			0.316		
	WxNrL	23.872	.000	.427	10.964	.002	.224
	WxG	1.973	.170	.058	1.058	.310	.027
	NrLxG	0.556			1.149	.291	.029
	WxGxNrL	0.451			0.056		
Five-letter	W	3.940	.056	.110	4.228	.054	.182
	G	0.019			0.299		
	WxG	1.071	.309	.032	2.400	.138	.112
Six-letter	W	40.612	.000	.559	15.900	.001	.456
	G	0.693			2.382	.139	.111
	WxG	3.682	.064	.103	1.980	.175	.094

II

INFLUENCE OF READING SKILL AND WORD LENGTH ON FIXATION RELATED BRAIN ACTIVITY IN SCHOOL-AGED CHILDREN DURING NATURAL READING

by

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

Vision Research.

Reproduced with kind permission by Elsevier.

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 (Hyöna & 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 (Hyöna & 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 interest[1]. 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-back[3] 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, & Võ, 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–project[4]. 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 sentence[5]. 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 pyykkitelineellä 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 ~ 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 ~ 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 ~ 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.

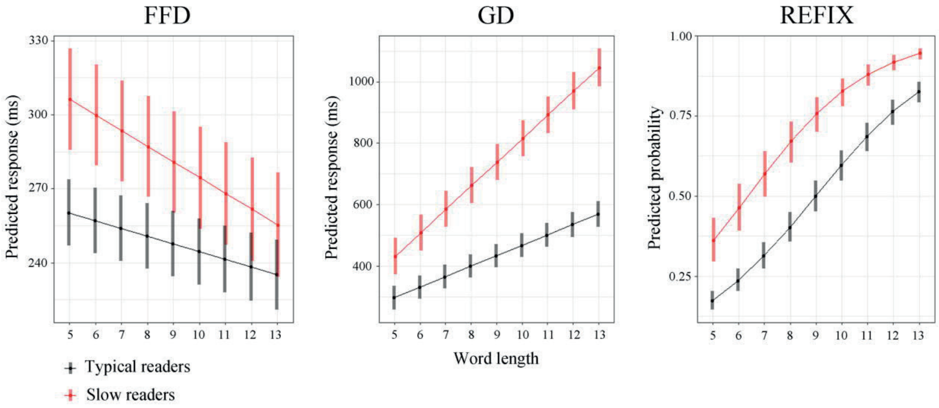


Figure 2. Predicted responses from linear mixed effect models exploring first fixation duration (FFD), gaze duration (GD) and generalized linear mixed effect model exploring Re-fixation probability (REFIX). Red lines represent slow readers (N = 27) and black lines represent typical readers (N = 65).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

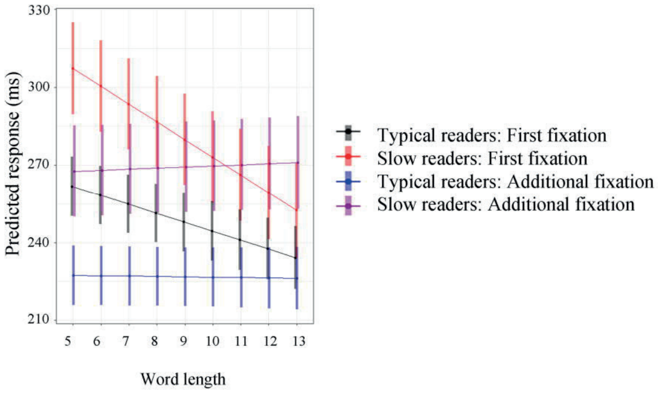


Figure 3. Predicted responses from the linear mixed effects model examining the effect of word length, group and fixation type on fixation duration. Black lines indicate predicted durations of first fixations for typical readers (N = 65). Blue lines indicate predicted durations of additional fixations. 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 $p = 0.70097$).

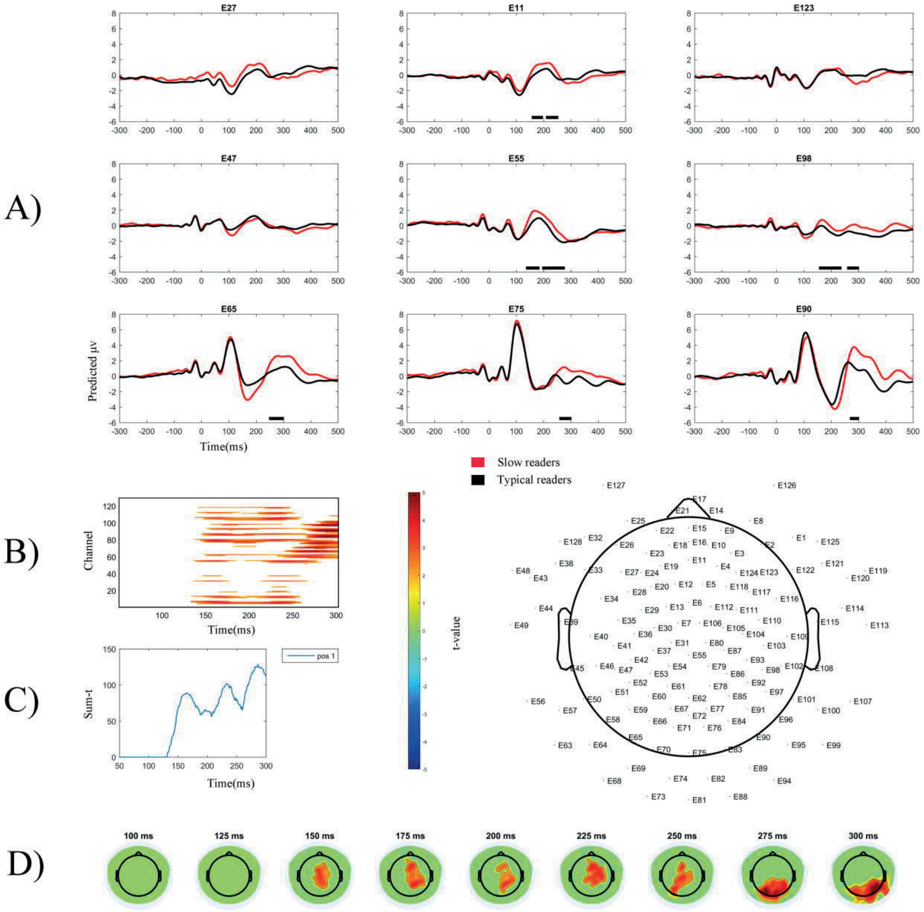


Figure 4: Between-group nonparametric cluster-based permutation test results for the intercept waveform that contains the shared activity of all responses. A) Grand average waveforms of the intercept at select channels. The red line represents the intercept response for slow readers (N = 27). The black 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 ($p = 0.04710$).

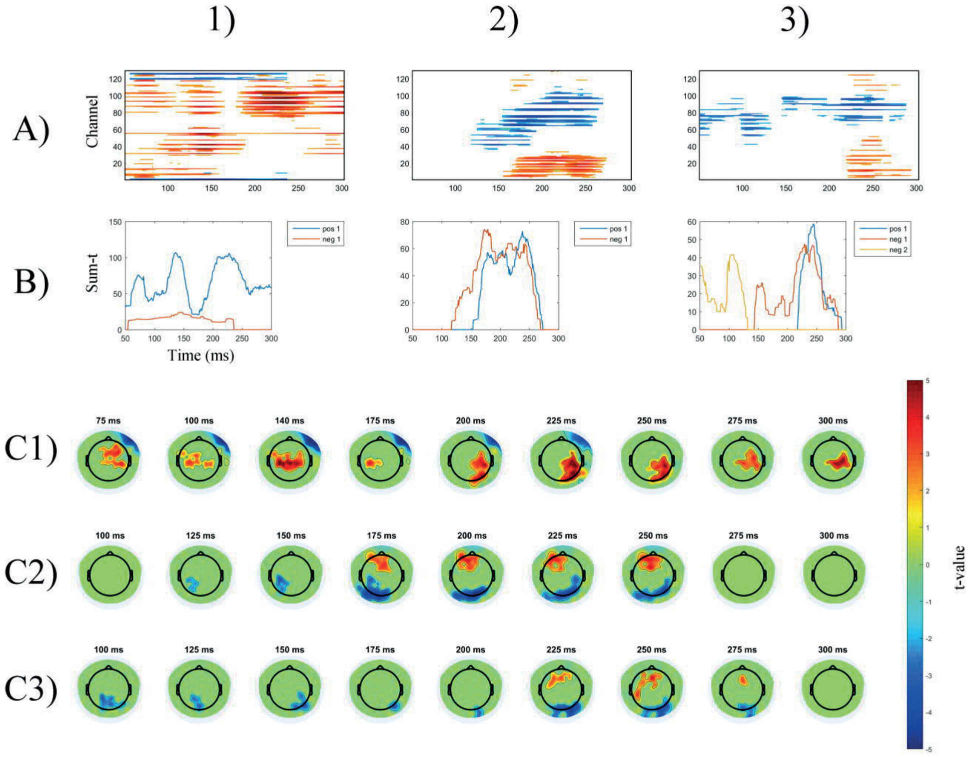


Figure 5. Within-group one-sample non-parametric cluster-based permutation test results for the factor beta waveforms, which reflect the modulation of the FRP waveform as a function of the modelled factors. A1) is the raster plot of additional fixation effect among typical readers (N = 65). A2) is the 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.

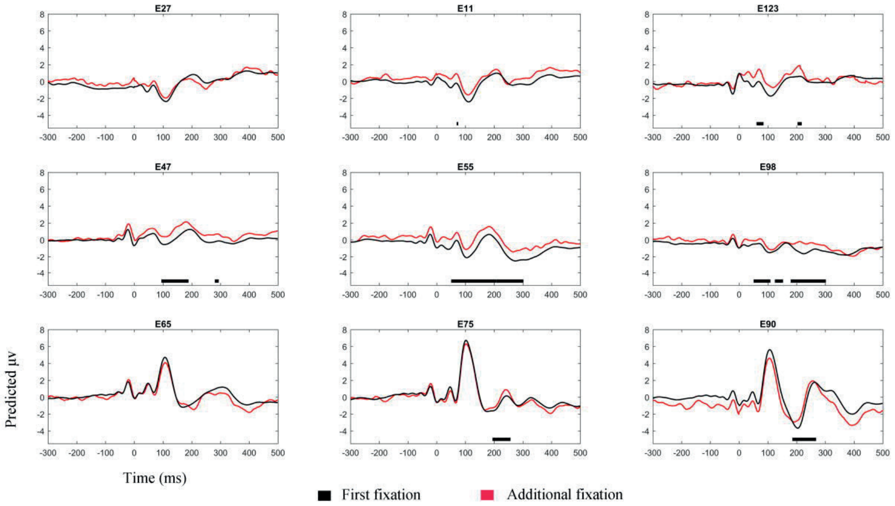


Figure 6: The predicted effect of additional fixation among typical readers (N = 65). The black line is the intercept which represents the first fixation in the full deconvolution model. The red line is the intercept with the effect of the additional fixation added to it. The black rectangles under the 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 ~ 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.

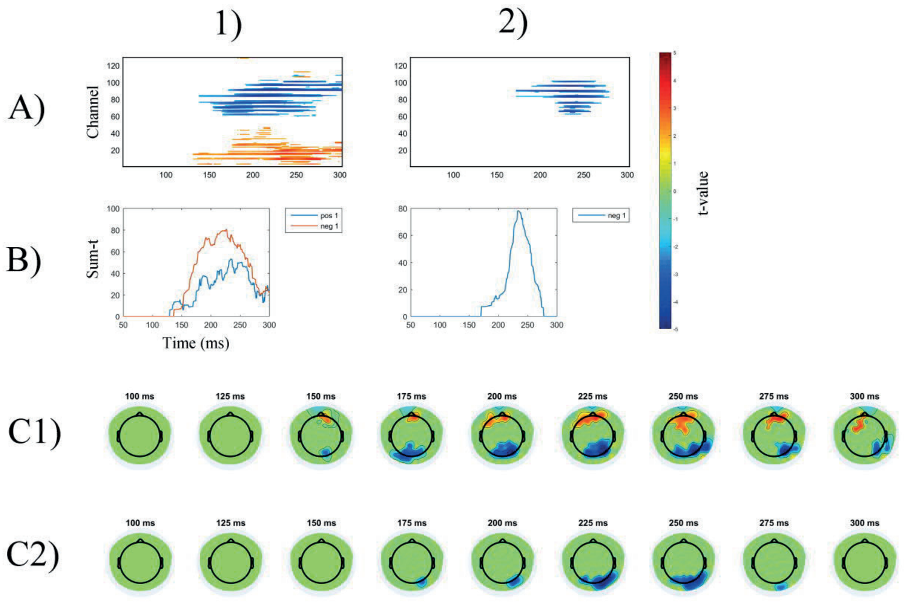


Figure 7: Post hoc analysis results of one-sample permutation t-test on word length effect in additional fixation. A1) Raster plot of word length effect in additional fixation for typical readers. A2) Raster plot of word length effect in additional fixation for slow readers. B1) Time behaviour of the 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.

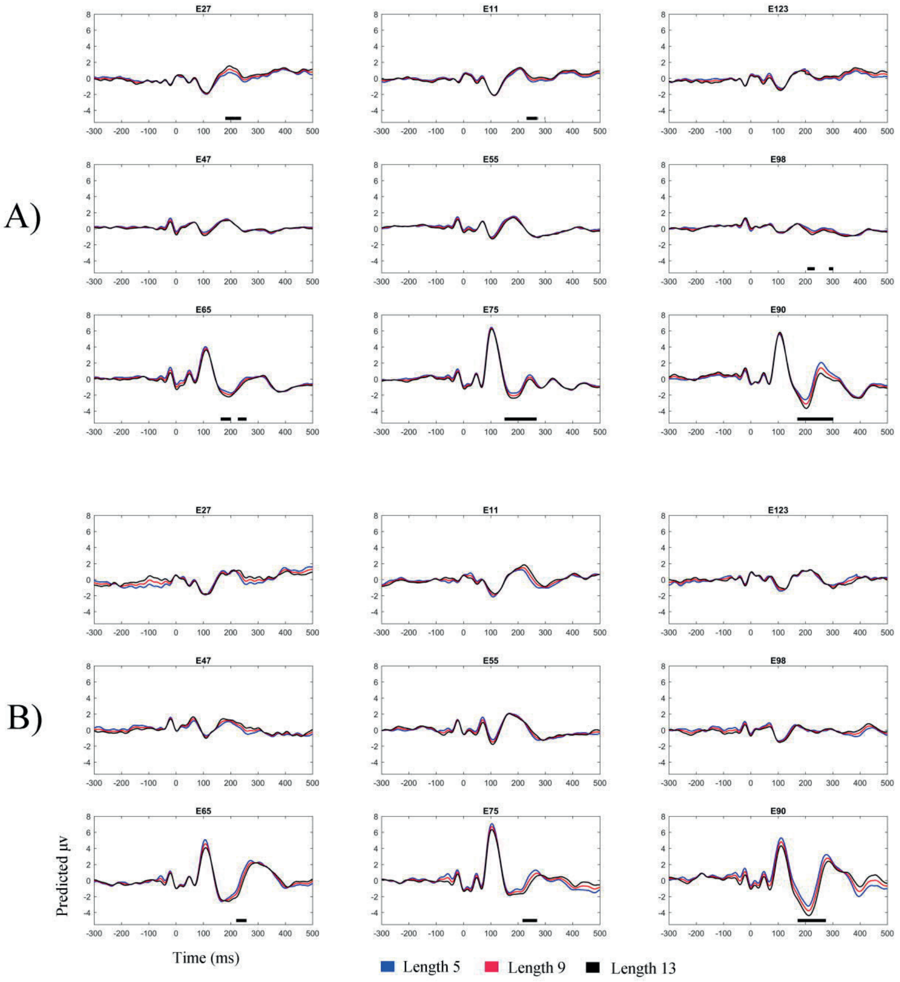


Figure 8: Predicted word length effects from additional fixations on specific word lengths added to the additional fixation intercept. A) Typical readers (N = 65). B) Slow readers (N = 27). The blue line indicates the word length effect prediction on five-lettered words added to the intercept. The red 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 assume that the sources at the occipital cortex or bilateral posterior temporal cortices are the main contributors for this difference[6]. 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

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., & 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

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

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

Dürrwächter, 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., Utriainen, J., Coiro, J., Kanniainen, 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

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

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, & Schlesewsky, 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, Schlesewsky, & 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, & Võ, 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., & Schlesewsky, 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., Schlesewsky, 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

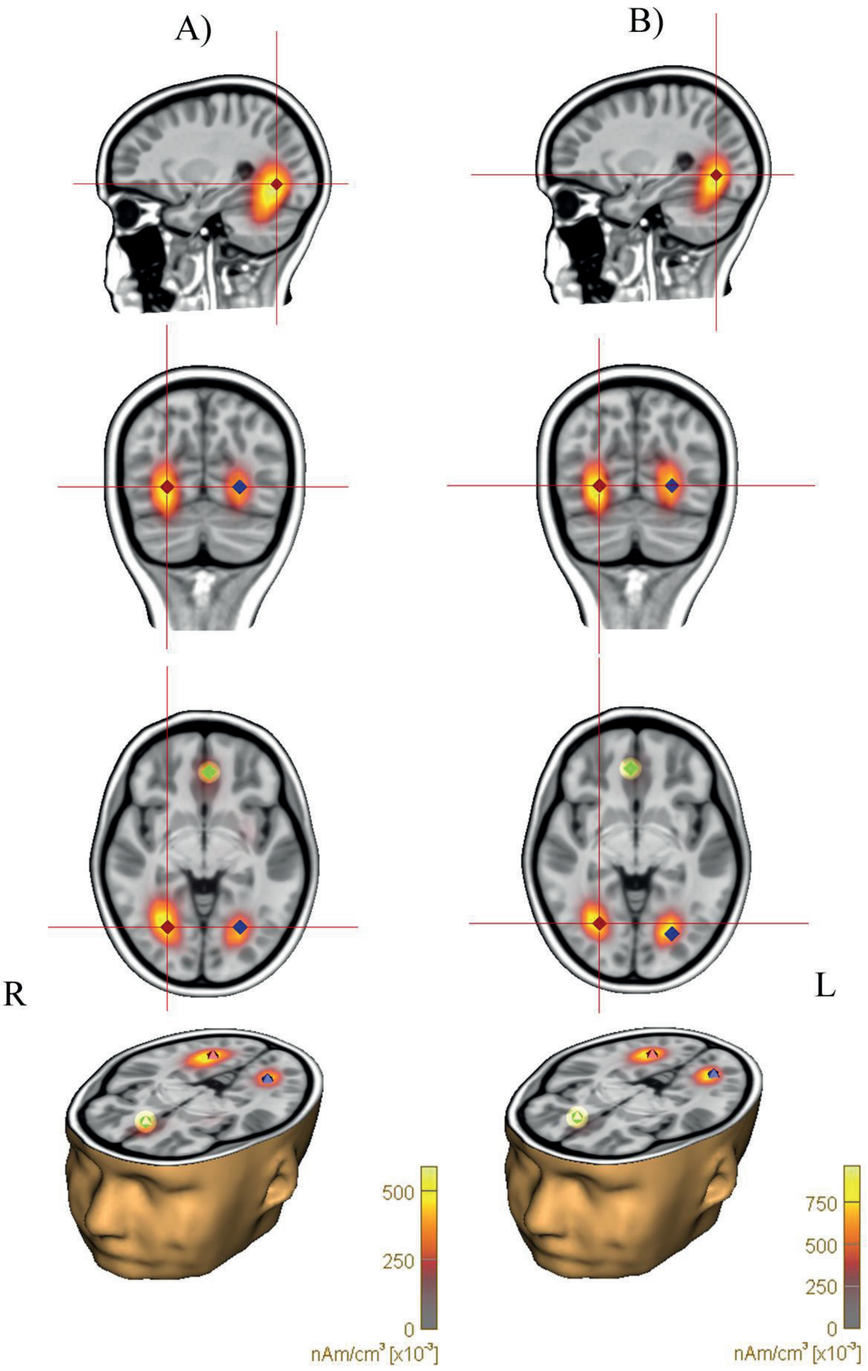
Woldorff, M. G. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. *Psychophysiology*, *30*, 98–119. https://doi.org/10.1111/j.1469-8986.1993.tb03209.x

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.


A)
B)
R
L
500
250
0
nAm/cm³ [x10⁻³]
750
500
250
0
nAm/cm³ [x10⁻³]

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				Talairach			
	x	y	z	C-RV%	x	y	z	C-RV%
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

III

SEMANTIC ANOMALY DETECTION IN SCHOOL-AGED CHILDREN DURING NATURAL SENTENCE READING —A STUDY OF FIXATION-RELATED BRAIN POTENTIALS

by

Otto Loberg, Jarkko Hautala, Jarmo A. Hämäläinen & Paavo H.T. Leppänen,
2018

PLOS One, *13*(12), e0209741.

Reproduced with kind permission by Public Library of Science.

PLOS | ONE

RESEARCH ARTICLE

Semantic anomaly detection in school-aged children during natural sentence reading – A study of fixation-related brain potentials

Otto Loberg[1]*, **Jarkko Hautala**[2], **Jarmo A. Hämäläinen**[1], **Paavo H. T. Leppänen**[1]

1 Department of Psychology, University of Jyväskylä, Jyväskylä, Finland, **2** Niilo Mäki Institute, Jyväskylä, Finland

* Loberg.o.h@gmail.com

Citation: Loberg O, Hautala J, Hämäläinen JA, Leppänen PHT (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

Editor: Jed Meltzer, Baycrest, CANADA

Received: February 5, 2018

Accepted: December 11, 2018

Published: December 27, 2018

Data Availability Statement: Data is available from principal investigator (Paavo H.T Leppänen, paavo.ht.leppanen@jyu.fi) to known collaborators upon request. Access to data can also be inquired from Ethical Board of University of Jyväskylä at secretary-ethicomm@jyu.fi. This data cannot be submitted to public/open repository as such use has not been included in the written informed consent that the legal guardians of the participants (which were children) have signed.

Funding: This work was supported by the Academy of Finland: Grant no. 274022 "Internet

Abstract

In this study, we investigated the effects of context-related semantic anomalies on the fixation-related brain potentials of 12–13-year-old Finnish children in grade 6 during sentence reading. The detection of such anomalies is typically reflected in the N400 event-related potential. We also examined whether the representation invoked by the sentence context extends to the orthographic representation level by replacing the final words of the sentence with an anomalous word neighbour of a plausible word. The eye-movement results show that the anomalous word neighbours of plausible words cause similar first-fixation and gaze duration reactions, as do other anomalous words. Similarly, we observed frontal negativity in the fixation-related potential of the unrelated anomalous words and in the anomalous word neighbours. This frontal negativity was larger in both anomalous conditions than in the response elicited by the plausible condition. We thus show that the brain successfully uses context to separate anomalous words from plausible words on a single letter level during free reading. From the P600 response of the scalp waveform, we observed that the P600 was delayed in the anomalous word neighbour condition. We performed group-level decomposition on the data with ICA (independent component analysis) and analysed the time course and source structure of the decomposed data. This analysis of decomposed brain signals not only confirmed the delay of the P600 response but also revealed that the frontal negativity concealed s more typical and separate N400 response, which was similarly delayed in the anomalous word neighbour condition, as was the P600 response. Source analysis of these independent components implicated the right frontal eye field as the cortical source for the frontal negativity and the middle temporal and parietal regions as cortical sources for the components resembling the N400 and P600 responses. We interpret the delays present in N400 and P600 responses to anomalous word neighbours to reflect competition with the representation of the plausible word just one letter different.

Introduction

Forming a mental representation of the semantic content of text is the end goal of the reading process. This is achieved through a sequence of visual inputs and the analysis of these inputs.

and learning difficulties: multidisciplinary approach for understanding information seeking in new media (eSeek)" to Phd Paavo H. T. Leppänen. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Exact features of this visual input sequence are modulated by the reading process itself. Indeed, previous research has shown that gaze behaviour, brain activation and the properties of processed text interact in a complicated manner [1,2]. Some attempts have been made to synthesise the knowledge obtained from separate eye-tracking and brain ERP (event-related potential) experiments, but fundamental differences in how these experiments are run impede the joint interpretation of the results. Studying eye-movement-locked brain activity (i.e. fixation-related brain responses) related to semantic processing during natural reading provides advances for the interpretation of both eye movements and brain responses. Specific methodological challenges arise from the dynamic nature of how eye movements occur during reading as well as the artefactual contamination of EEG from eye movements. However, the benefits of co-registering EEG and optical eye tracking in terms of increased ecological validity far outweigh the challenges, both of which are discussed below.

Semantic processing of sentences in gaze behaviour and in brain activity

The semantic processing of sentences involves not only the processing of the individual words; the relationships between words also need to be considered to grasp the meaning contained in the sentences. It has been proposed that forming a representation of a sentence proceeds through three stages: 1) building a local phrase structure, 2) assigning syntactic and thematic and semantic relations and 3) integration to semantic representation [3]. Of particular interest here are stages 2 and 3.

Several factors affect eye movements during semantic processing. Predictability, for example, is a major determinant of fixation duration during reading [4]. In addition, the thematic relationships between words modulate the gaze behaviour. For example, when a word clearly violates the sentence context, it will result in a longer first-pass gaze duration (sum of fixation durations during the first pass of the target word) [5,6]. In earlier research merely implausible or improbable word in relation to thematic rules was not reflected in the first-pass measures but rather in the total fixation duration (sum of all fixation durations during the trial allocated to the target word), which indicates that implausibility is processed at a later stage than outright theme violation in the form of a semantic anomaly [5]. However more recent research with larger sample sizes and more suitable statistical approaches has shown that plausibility is capable of modulating first pass measures[7,8] and this includes influences from parafoveal processing[9,10].

A systematic way of inducing ERP effects reflecting semantic processing is to set up a semantic context with a sentence and then disrupt it by inserting a word that is anomalous with the established context. Consequently, this anomalous word violates the set of thematic rules established by the other words and gives rise to N400, a centroparietal brain response with negative polarity typically reaching its maximum amplitude in RSVP experiments at about 400ms post onset of the anomalous word [11]. N400 is thought to result from difficulty in integrating the meaning of the violating word with the context of the preceding sentence [11] or to reflect indexing of the goodness of fit of the currently processed word with the context of the sentence [12]. Both of these suggested functions are similar to stage 2 of the model described above [3]. Other factors besides congruency have also been shown to influence N400. For example word frequency affects the amplitude of N400 in low-context situations where prior context has not been defined, for example, at the beginning of the sentence. Thus, in order to avoid confounds from word frequency and to maximise semantic incongruity, the manipulation should be placed at the end of the sentence. Further, the presentation parameters of RSVP have direct consequences to N400. Faster presentation rates diminish amplitudes and shorten the latency of N400 [13], which emphasises the uncertainty regarding the ecological validity of RSVP experiments with slow presentation rates.

Generally, N400 experiments have manipulated only the semantic aspects without paying attention to low-level features, such as word length or visuo-spatial similarities of anomalous and plausible words. A fairly recent experiment [14] tested whether the deviation of a single letter from a probable word was sufficient to cause N400 –i.e. does the anomalous orthographic word neighbour of the predicted word cause a similar N400 as an anomalous word without a similar word body? N400 was found to be diminished for anomalous word neighbours compared to anomalous words without orthographic resemblance to the correct word, which was interpreted to indicate facilitated integration of words with similar word bodies. This effect was also present for pseudoword neighbours containing no semantic information, thus providing strong evidence that a set of orthographic features was pre-activated by the preceding context. As these findings were found in a fairly slow-paced (250ms stimulus duration + 250ms blank screen) RSVP design, the question remains as to whether a one-letter deviation from a plausible word would elicit similar effects during free reading.

Recently P600, a parietal positive ERP reaching maximal amplitude at 600ms that has been classically defined as a response to syntax violations [15], has also been shown to be elicited by semantic anomalies [3,16,17]. Similarly to N400, the P600 effects seem to diminish with faster presentation rates [18]. Curiously, P600 has been found to be related to regressive eye movements in an FRP study investigating the effects of free reading on brain activity during reading comprehension [1]. It was also found to be absent in the RSVP version of the experiment. This led the authors to conclude that coupling between P600 and regressions indicates attempts to recover or re-organise confusing content. It has also been argued that rather than encountering anomalies in a specific aspect of language, such as semantics, syntax or grammar, P600 reflects an integration phase or combinatory reprocessing of the elements of the sentence [3,19–21] or even a conscious-level perception of the whole sentence-level semantic anomaly [12].

The neural generators of N400 seem to be rather widespread [22,23]. Studies of individuals with brain lesions, studies with intracranial recordings and studies with magnetoencephalographic data have implicated contributions of the left and right temporal lobe, with a left hemispheric predominance [24]. The larger left than right hemispheric response can also be observed in a study comparing typically reading adults with dyslexic adults [25]. A recent study using a beamformer analysis of the magnetic equivalents of N400 (N400m) and P600 (P600m) found left superior temporal and posterior frontal regions to underlie N400m and distributed activation of bilateral frontal, posterior temporal and parietal regions P600m [26]. Further, recordings with functional magnetic resonance imaging (fMRI) suggest that the processing of semantic anomalies occurs in mid portions of the superior temporal region and insular cortex, both bilaterally, whereas syntactic processing is associated with the anterior portion of the left superior temporal sulcus and left posterior frontal operculum [27]. In positron emission tomography (PET) studies, the role of the angular gyrus has also been implicated in semantic processing [28,29]. It needs to be noted here that different measuring and analysis methods as well as different experimental designs have different sensitivities and can thus lead to quite distinct location results. Different results do not render conclusions right or wrong but rather reflect different aspects of the process, which are reflected in different signals.

In our current study, we investigated semantic processing and how orthographic similarity between anomalous and plausible words affects semantic processing during free reading in 12–13-year-old children using FRPs (co-registered eye tracking and EEG time-locked to fixation onset). We employed a free sentence reading paradigm to examine the effects of a semantic anomaly in two conditions: unrelated anomalous words and anomalous orthographic word neighbours of plausible words as the target words appearing at the end of the sentence. We were especially interested in whether a deviation of a single letter from a plausible word would elicit complementary effects on gaze behaviour in relation to brain activation during free reading. Based on previous studies [14], the anomalous orthographic word neighbours were expected to produce reduced N400 compared to the unrelated anomalous words. As the fixation duration and N400 effects are associated with each other [2] and react to the same manipulations, it was expected that the fixation duration effect of the anomalous word neighbour would be similarly attenuated (similar to the N400 effect) when compared to unrelated anomalous words. Further, since P600 is thought to reflect an integration phase in sentence comprehension [3], we expected that unrelated anomalous words would generate stronger P600 responses than anomalous neighbours of a plausible word. We also considered the possibility that the semantic responses would be delayed in the case of the anomalous word neighbours of the plausible words.

Methodological considerations when studying brain activity during reading

RSVP (rapid serial visual presentation) is most common stimulus presentation procedure for studying brain function during reading. In RSVP, each word in a sentence (or other text material) is usually presented on the screen individually one word at a time in a sequence to the participant's fovea, and the brain activity is time-locked to each word (screen) onset. Arguably, the elicitation of ERPs by individual words makes RSVP an attractive approach to study reading-related brain processes. Indeed experiments utilising the RSVP technique have formed the foundation for the knowledge about brain function during word recognition and higher-level semantic processing. However, the way in which RSVP oversimplifies reading into a static linear process decreases its ecological validity considerably. From eye-movement studies we know that reading consists of multiple dynamic processes, such as visual intake during fixations, the lengths of which are modulated by the processing demands, skipping of words because of their predictability and visual characteristics and frequent gaze regresses to previous parts of the text to re-read material that are ambiguous or hard to understand [4]. Generally, in RSVP experiments presentation times are static, and there is no chance of skipping words or returning back in the text. However, during normal text reading, the parafoveal information of the words surrounding the fixated words is also processed to some extent [4], which is not possible in the standard implementation of RSVP. Alternative versions of RSVP have been proposed to counteract these issues: self-paced RSVP, where [30] the presentation rate of the words is controlled by the participant, and RSVP with flankers [31], where the sentence 'slides' over the participants' foveal field of vision. However, these modified versions also have some problems. Self-paced RSVP arguably requires conscious monitoring of the presentation rate by the participant while eye movements during reading do not have such requirement, as they are largely automatised. Meanwhile, RSVP with flankers requires suppression of the eye movements towards the flankers, which is a cognitive requirement that is not present during natural reading situations. Furthermore it has been shown that pre-saccadic attention prepares the visual system for the next retinal input [32,33]. As pre-saccadic attention is tightly related to impending saccade execution, its role cannot be fulfilled if saccades are not made. Recent findings illustrate that volition in attention allocation and saccade generation towards words facilitate word recognition in a way that is not present in RSVP [34]

Co-registering EEG and eye tracking represent an ecologically valid alternative for studying reading. This is achieved through analysing the EEG signal based on selected gaze behaviours, for example, fixation on a target word. The combination of co-registration and a behaviour-based analysis approach allows the participant to proceed in the reading task at his or her own pace, with the ability to regress back in the text as well as to skip words. From this participant-initiated reading behaviour we can extract FRPs (fixation-related potentials), electrical brain responses that are very similar to visually evoked potentials from more traditional ERP experiments [34,35]. Indeed, studies comparing the RSVP and FRP methodologies have shown that brain activity during naturalistic FRP experiments differs drastically from RSVP equivalents [1,34], which further emphasises the need for brain research during reading with naturalistic experimental designs.

When performing reading studies with the FRP method, two things need to be considered. First, the ocular artefacts and extra-ocular muscle artefacts feature prominently in the co-registered EEG signal [35]. Generally, these artefacts are stronger than the signal of interest and thus need to be removed from the data. This can be achieved with blind-source separation methods such as ICA and temporal selection methods [36,37]. Temporal selection exploits the dependence between the ocular artefact independent component time course and the time course of the optical eye-movement record [37]. Recent research has shown that artefact signals associated with FRPs can be effectively managed [1,35,38,39]. Second, a more complex issue relates to the fact that during reading fixations occur in relatively fast succession, which results in additional spatiotemporal mixing of the scalp-recorded EEG/ERP signals, especially in the parts of the signal that exceed the duration of the fixation that is used as the time-locking event [40]. As cognitive processes have an evident influence on eye movements [4], changes in gaze pattern due to differences in conditions reflect the mixing of brain activity from different sources in the latter parts of the averaged epoch [40]. We propose to untangle this issue by analysing the time courses of these underlying activity patterns by separating them with blind-source separation methods, such as ICA [36,41].

Even though identifying and removing artefactual signals with ICA from the signal of interest is the typical use in contemporary EEG analysis [42], ICA can also be applied to separate brain signals from each other. This can be of particular interest for late latency components, which typically have multiple neural sources, or when there are multiple overlapping sensory responses. The underlying assumptions of ICA place certain restrictions on the nature of the sources detected. First, a signal arising from a spatial source is assumed to be temporally independent from other spatial sources [36,41]. Second, the spatial source is assumed to be in a fixed location throughout the duration of the measurement of the data of interest [43]. As independent components are defined as spatially fixed sources that change in activity through time, several interesting possibilities arise. First, there is the option to forward project an independent component's activity to a scalp activity and localise it [43,44]. Second, data from several experiments from the same subject can be inserted into the same decomposition to determine whether different experiments share psychological processes and underlying neural mechanisms [42]. Third, a decomposition analysis (run) can also include data from multiple subjects, and thus the obtained group solutions can be applied to all of the subjects [45]. Hence, the application of ICA beyond artefact cleaning is a promising approach for FRP experiments on natural reading, as the stationarity of sources provides a solution to the ever-present spatio-temporal overlap of activity from previous and consequent fixations and their potential confounds to the scalp signal. Running ICA across a sample of participants also reduces measurement-related error variation (e.g. due to random individual differences in arousal state) and should thus improve the results of ICA decomposition when looking for shared brain activity sources across the sample. In the current study, we employed a group ICA procedure to disentangle spatially overlapping processes during sentence reading and when encountering semantic anomalies in the sentences.

Methods

Participants

A total of 66 typically reading elementary school students in grade 6 (from 12.0 to 13.5 years; 36 female) were recruited as part of the eSeek!–Internet and Learning Difficulties: A

Multidisciplinary Approach for Understanding Reading in New Media–project. In agreement with the Declaration of Helsinki, the study was approved by the Ethical committee of the University of Jyväskylä. Written consent was acquired from the parents of the participants, and the children had the opportunity to discontinue the experiment at any time. All participants reported normal or corrected vision, no history of learning difficulties or neurological abnormalities and were rated higher than the weakest 13% of the whole eSeek population (whole norming population being 542 children of the same age group) in reading fluency. Reading fluency was estimated with three tests and reduced to a single factor with principal axis factoring with PROMAX rotation using the IBM SPSS 24 statistics programme (IBM Inc.). The three tests were the *Word Identification Test*–a subtest of the standardised Finnish reading test ALLU [46], the *Word Chain Test* [47] and the *Oral Pseudoword Reading Test* [48].

Apparatus

The experiment was administered using Experiment Builder (1.10.1630) software running on a Dell Precision T5500 workstation. Eye movements were recorded with a table-mounted Eyelink 1000 eye tracker with a 2000 Hz upgrade (SR Research Ltd). Both eyes were recorded at 1000 Hz. The EEG was recorded with a NeurOne amplifier (Bittium) at a 1000 Hz sampling rate with a 128 channel electrode net (Electrical Geodesics Inc.) using Ag/AgCl electrodes. The synchrony between the eye movements and EEG measures was established with a combination of triggering Ethernet messages and transistor to transistor logic (TTL) pulses, both originating from the workstation running the experiment. The stability of the synchronisation between the eye-tracker recording and EEG was checked by comparing the time differences in the trial onset and offset messages in both data streams. The participants leaned on a chinrest while their eye movements were recorded, with a distance of 60 cm from the participants' eyes, and the EEG was recorded simultaneously. The participants' responses were recorded into the EEG event stream and into an individual response file on the workstation running the experiment. The experiments were carried out in a dimly lit and soundproofed room at the laboratory facilities of the University of Jyväskylä.

Stimulus procedure

Before each trial, a black dot appeared on the left side of the screen at the vertical level of the participant's eyes, and the participant was instructed to remain fixated on it while the experimenter approved the fixation. The black dots also served as confirmation for the validity of the calibration. If the fixation on the dot differed from the calibration by more than 1 degree, the experiment software alerted the experimenter, and the calibration was redone. After the experimenter accepted the fixation, the fixation dot disappeared, and the sentence appeared. The participants were instructed to read the sentence as quickly as possible and then judge whether it was sensible or not by pressing the left button for 'yes' and the right for 'no' on a two-button response box using the right index and middle fingers. There was no time limit for responding, and the sentence disappeared only after the response. The experiment was divided into four blocks, between which the eye tracker was recalibrated and the quality of the EEG maintained.

Materials

The stimuli consisted of 200 sentences with a median length of six words (ranging from 5 to 9). On the screen, one letter subtended 0.4 degrees visual angle. The sentences were divided into three categories: 100 plausible sentences, 50 sentences where the target word was severely anomalous to preceding sentence context and 50 sentences where the target word was severely anomalous to the preceding context but was an orthographic word neighbour of a plausible

1) Museon vetonaula oli hauras muumio

2) Lihaksien kasvattamiseen vaaditaan rankka sormus

3) Yöllä kissan saaliiksi päätyi hidas hiili

Fig 1. Example stimuli. The part of the sentence that deviates from the context is highlighted in red. Sentence translations: 1) plausible–'Main attraction of the museum was the fragile mummy'; 2) unrelated anomalous–'For building up muscle mass one needs an intense ring'; 3) anomalous word neighbour–'During the night the cat caught a slow coal' ('hiili', coal, being a neighbour of 'hiiri', a mouse).

https://doi.org/10.1371/journal.pone.0209741.g001

word. The categories were balanced in this way to make task response (yes/no) probability 50%. Examples of the sentences are presented in Fig 1. The frequency of the target words and the previous words was controlled for (2x3 repeated measures ANOVA: previous word vs. target word $p>0.05$, condition $p>0.05$, condition * previous word vs. target word $p>0.05$). The frequencies of the target and previous words (Table 1) were extracted from a newspaper corpus [49]. A norming study was completed with an independent sample (N = 10) to estimate the cloze probability of the target words and plausibility of the whole sentences. Also the plausible ending of the sentences in the anomalous word neighbour condition was included in the norming study. Cloze was estimated with a standard cloze task while plausibility was estimated using a seven-point Likert-type scale, where 1 corresponded to 'not at all plausible' and 7 to 'highly plausible'. Unrelated anomalous words did not significantly differ from anomalous word neighbours in either cloze or plausibility (independent samples t-test, $p>0.05$). The plausible condition did not significantly differ from the plausible ending of anomalous word neighbour sentence in cloze or plausibility ($p>0.05$). These norms are presented in Table 1.

EEG and eye-tracking data preprocessing

Data were preprocessed using Eeglab 13.3.2 [36] with the EYE-EEG extension [35]. The saccades were detected from the gaze location data with a median velocity-based algorithm [50], using 6 standard deviations from the median velocity as a threshold for a saccade (minimum duration of 4ms; if the two saccades were less than 50ms apart from each other, they were merged into a single saccade).

Table 1. Target word and pre-target word frequencies and sentence characteristics.

Condition	Previous Word (SD)	Target Word (SD)
PLA	7.42 (21.58)	7.51 (20.86)
URA	4.98 (6.95)	3.32 (3.59)
AWN	3.62 (4.03)	4.59 (7.27)
	Cloze probability	Plausibility
PLA	41% (36%)	6.76 (0.35)
URA	0 (0)	1.20 (0.23)
AWN	0 (0)	1.26 (0.45)
P-AWN	51% (39%)	6.86 (0.18)

Frequencies (mean incidence in 1 million words) of target and preceding words (with standard deviations).

Plausibility: 1 = Not at all plausible, 7 = Highly plausible

Note. PLA = Plausible, URA = Unrelated anomalous, AWN = Anomalous Word Neighbour, P-AWN = Plausible Word Neighbour of the used anomalous Word Neighbour.

https://doi.org/10.1371/journal.pone.0209741.t001

EEG data were off-line filtered, with a high-pass filter of 0.5 Hz and 20 Hz as the low-pass filter. Electric manifestations of eye movements (ocular artefacts) were modelled with ICA, with a PCA (principal component analysis) reduction of the 128 channels to 100 principal components prior to the ICA training. The components were selected to correspond with the optical recording of eye movements using the temporal covariance criterion [37] of 1.1. Selected independent components (ICs) were pruned out of the EEG data.

After the ocular artefact pruning, the EEG epochs with the fixations of interest were selected. These included the first fixation on the target word and the previous fixation. If the previous fixation was not on the word preceding the target word, the whole trial was discarded to make the baseline segments of the FRP similar. Epochs for averaging were -100ms to 900ms, time-locked to the fixation onset. Trials with both correct and incorrect answers were kept. Fixations and their corresponding epochs were discarded if the EEG during the averaging epoch differed by more than 5 standard deviations from the mean on any channel or if the trial ended within the 900ms after first fixation on the target word to avoid confounding activity from trial offset.

Segmented single trial data (-100ms to 900ms) were subjected to a second run of ICA to determine shared fixation-related components. Prior to ICA, the data for each subject were standardised (basic z-score conversion) so that individuals with strong voltages would not drive the ICA decomposition. The data were downsampled to 250 Hz to make computational requirements feasible, and the group matrix was whitened to 30 principal components with PCA. Extended Infomax, an ICA algorithm sensitive to sub-Gaussian distributions [51], was used because it has been shown to produce good results with non-simulated group EEG data [52]. The resulting group-level ICA weights were then applied to the 1000 Hz individual data, and the resulting IC activities were used for statistical inference. It is important to note here that our application of group-level ICA does not attempt to reconstruct topographies of individual subject's ICs but applies the group-level weights obtained from the population-mixing matrix directly to the individual subject data. Ten of the 30 components were determined by visual inspection to be such that they could be generated by a cortical source by displaying a dipolar pattern. Pattern was judged to be dipolar if it had either 1-pole uniform field on several adjacent channels or 2-pole field on several adjacent channels and nearby polarity reversal. Components judged to be noise were pruned out of the data if they displayed known artefact characteristics (eye movements, single channel pop-out artefacts). Topographies of all components derived with the group ICA are presented in Fig 2. Table 2 contains the PVAF (percentage of variance accounted for) values of each component calculated with the EEGLAB [36] function *compvar*, PVAF is a measure that describes the amount of variance an independent component explains.

The average quantity of fixations and the corresponding epochs for averaging for each condition were as follows: plausible 83, unrelated anomalous 41 and word neighbour anomalous 41 fixations.

Statistical analysis of the eye-tracking and response accuracy behavioural variables

The behavioural variables were analysed in MATLAB. The effects of the conditions on response accuracy were analysed using a Wilcoxon sign-rank test, from which we report Z-values, P-values and effect sizes [53]. Based on the eye-movement variables, the previous fixation duration (PFD), the first fixation duration, gaze duration and immediate re-fixations were analysed. Because trials that were ended with a participant response within 900ms of the first fixation onset were discarded, the re-fixation measure here is a binary measure of regression to

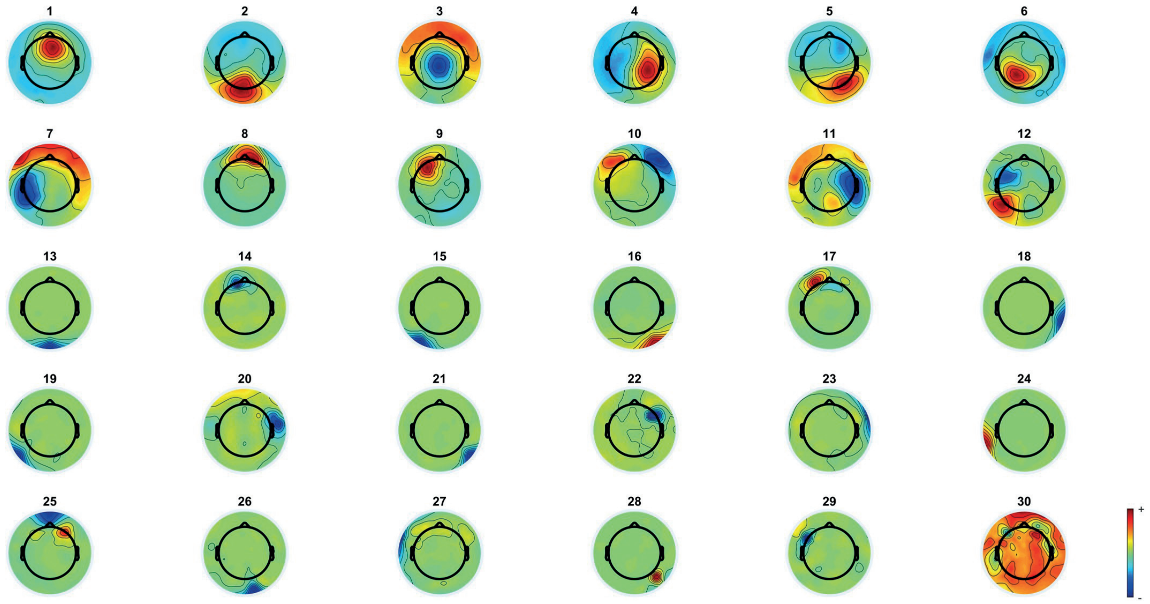


Fig 2. The group-level ICA estimated spatial filters. Components 1, 2, 3, 4, 5, 6, 7, 9, 11 and 12 were retained in the data because they were judged to have a dipolar field structure and thus were likely be generated by brain tissue. Components 8 and 10 were pruned out of the data since they were judged to be likely to have been caused by eye movements. Components 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 and 30 were pruned out of the data because they were judged to be random pop-out artefacts and mechanical artefacts mostly constrained to few channels.

https://doi.org/10.1371/journal.pone.0209741.g002

Table 2. Descriptive values of the percentage of variance accounted for (PVAF) by the group ICA components.

	Mean PVAF	Sd	Median PVAF		Mean PVAF	Sd	Median pvaf
All IC	**82.52%**	**8.08%**	**85.44%**	**Cleaned scalp (*)**	**51.75%**	**20.69%**	**58.4%**
IC 1 *	7.97%	5.19%	8.28%	IC 16	1.48%	3.3%	0.63%
IC 2 *	6.93%	4.94%	6.82%	IC 17	1.42%	2.8%	0.74%
IC 3 *	7.24%	5.05%	6.97%	IC 18	1.5%	5.07%	0.54%
IC 4 *	6.76%	4.03%	6.41%	IC 19	1.25%	2.08%	0.61%
IC 5 *	5.74%	3.25%	5.97%	IC 20	1.04%	0.74%	0.88%
IC 6 *	3.98%	2.24%	3.96%	IC 21	1.17%	3.29%	0.37%
IC 7 *	4.21%	5.01%	3.44%	IC 22	1.05%	3.24%	0.46%
IC 8	4.53%	6.63%	2.8%	IC 23	1.02%	1.05%	0.64%
IC 9 *	3.9%	4.08%	3%	IC 24	1.29%	5.36%	0.32%
IC 10	3.35%	3.28%	2.57%	IC 25	1.03%	1.4%	0.63%
IC 11 *	2.57%	1.14%	2.59%	IC 26	1.15%	3.95%	0.25%
IC 12 *	2.46%	6.52%	1.52%	IC 27	0.88%	0.88%	0.56%
IC 13	1.94%	3.76%	0.8%	IC 28	1.56%	10.65%	0.18%
IC 14	1.7%	4.3%	0.9%	IC 29	0.89%	1.92%	0.35%
IC 15	1.76%	4%	0.52%	IC 30	0.76%	1.1%	0.35%

PVAF (percentage of variance accounted for) values for each component.

Note.

* Denotes components retained in the data for statistical analysis

https://doi.org/10.1371/journal.pone.0209741.t002

Table 3. Eye-movement results.

Condition	PFD (SD)	FFD (SD)	GD (SD)	RFP (SD)
PLA	269ms (120ms)	296ms (158ms)	415ms (292ms)	31% (46%)
URA	272ms (115ms)	338ms (201ms)	571ms (410ms)	44% (50%)
AWN	273ms (118ms)	336ms (208ms)	588ms (468ms)	44% (50%)

Means and standard deviations of the previous fixation duration (PFD), the first fixation duration (FFD), gaze duration (GD) and re-fixation probability (RFP).
Note. PLA = Plausible, URA = Unrelated anomalous and AWN = Anomalous word neighbor.

https://doi.org/10.1371/journal.pone.0209741.t003

earlier parts of the sentence vs. immediate re-fixation within the target word. The eye-movement variables were analysed with the linear mixed effects (LME) models (for FFD and GD) and generalized mixed effects (GLME) models (for refixations), which is the standard procedure in eye-movement research to control for random effects from the individual subjects and items (*fitlme* & *fitglme* functions of the MATLAB statistics and machine learning toolbox). In the LME results, we use the plausible (PLA) condition as a reference condition, against which we compare unrelated anomalous (URA) and anomalous word neighbour (AWN). The estimated random structure was simple intercept structure for trials and subjects. Possible differences between URA and AWN are explored by comparing whether their effect parameters are different (*coefTest* function of the MATLAB statistics and machine learning toolbox). The mean values of the eye-movement variables are presented in Table 3 and also in Fig 3.

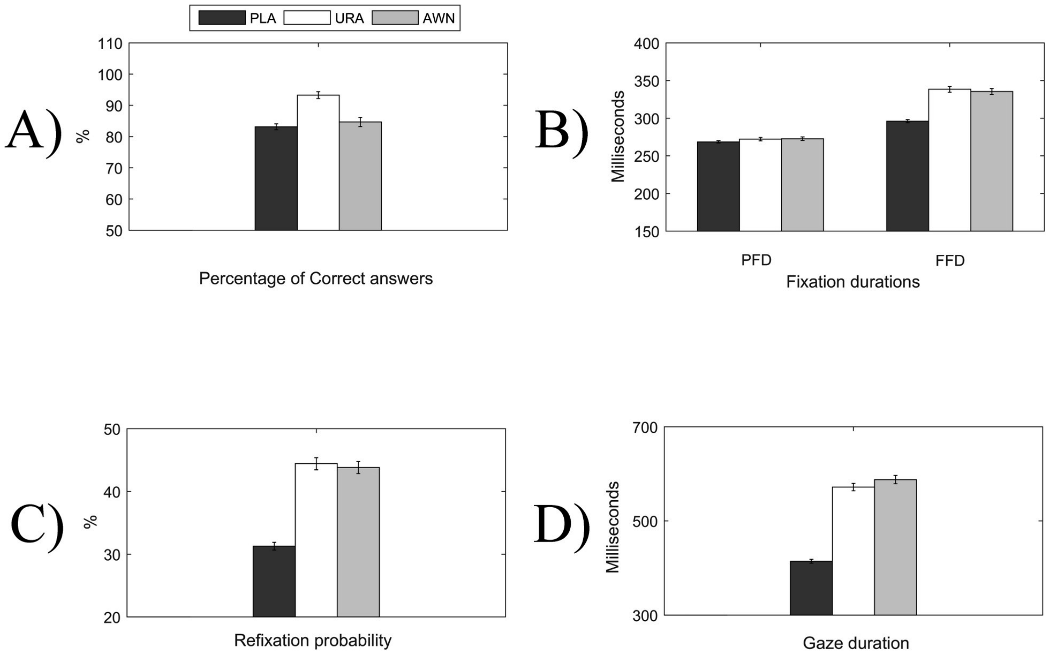


Fig 3. Behavioural measures. Bar graph presentation of the behavioural measures; exact values are presented in Table 3. A) Percentage of correct answers, B) Fixation durations of the previous fixation (PFD) and the first fixation (FFD) and C) Re-fixation probability (RFP) after the first fixation on the target word in a semantic sensibility judgement task in a sentence reading context of 12- to 13.5-year-old children (N = 66). The error bars denote 1 standard error. PLA = Plausible, URA = Unrelated anomalous and AWN = Anomalous word neighbour.

https://doi.org/10.1371/journal.pone.0209741.g003

Statistical analysis of the first fixation time-locked FRPs

FRPs were analysed with nonparametric cluster-based pairwise permutation statistics in Besa statistics 2.0 (for the method description, see[54]), which provides a solution to the multiple comparison problem present in multi-channel EEG recordings. Nonparametric cluster-based permutation tests have two essential steps. First, the desired test (here normal t-test) is run over all channels (when done at the scalp level) and time-points. Then, values over a certain significance (here the typical p<0.05 threshold) are clustered based on clustering criteria. Here, these criteria were temporal adjacency (significant samples were consecutive) and an electrode distance of 3cm (a significant sample in the electrode was part of a cluster if the distance from another significant sample in another electrode was less than 3cm). All t-values of the clusters (each sample, each electrode) are summed to form the cluster test statistic, which is used to estimate test significance. Second, the distribution to estimate this test statistic is generated by randomly re-assigning the condition labels in each average and running the test and clustering procedure again and storing the results in the permutation distribution. When the real observed sum-t probability in contrast to permutation distribution is smaller than 0.05, then the observed cluster is considered to be statistically significant. We used 10,000 permutations to define the permutation distribution. With three conditions, the pairwise condition contrasts for the permutation statistics were as follows: unrelated anomalous vs. plausible, word neighbour anomalous vs. plausible and unrelated anomalous vs. anomalous word neighbour condition. For the sake of simplicity, when we use the terms negative or positive in the results section, they describe the amplitude difference between the first response and the second response in the paired comparison.

Results

Response accuracy

The mean response accuracy (percentage of sentences correctly identified as sensible/insensible) was 83.34% (SD = 7.27) for the plausible condition, 93.4% (SD = 8.82) for the unrelated anomalous condition and 84.86% (SD = 12.02) for the anomalous word neighbour condition. Unrelated anomalous sentences had a significantly higher response accuracy (RA) than plausible (Z = 5.525, p<0.000001, r = 0.481) or word neighbour anomalous (Z = 6.414, p<0.000001, r = 0.558) sentences. There was no difference in response accuracy between plausible and anomalous word neighbour sentences (Z = 1.028, p = 0.303, r = 0.09).

Eye-movement results

Fixation durations. There were no significant effects of condition between plausible (PLA), unrelated anomalous (URA) and anomalous word neighbour (AWN) in PFDs (intercept: beta = 269.77, SE = 5.70, t = 47.31, p<0.00001, CI = 258.6–280.95; URA: beta = 3.62, SE = 4.12; t = 0.88, p = 0.37865, CI = -4.44–11.70 AWN: beta = 3.69, SE = 4.11; t = 0.89, p = 0.37042, CI = -4.38–11.76 and URA vs. AWN (p = 0.9892)). Analysis of FFD (intercept: beta = 295.49, SE = 7.29, t = 6.71, p<0.00001, CI = 281.19–309.78) showed that FFD was longer for URA (beta = 44.613, SE = 6.65, t = 6.71,p<0.00001, CI = 31.58–57.64) and AWN (beta = 40.98, SE = 6.65, t = 6.16, p<0.00001, CI = 27.94–54.01) target words than for PLA target words. There was no significant difference between the FFD for the anomalous word neighbour and unrelated anomalous target words (p = 0.6361).

First-pass gaze duration. Pattern of results for the first-pass gaze duration was the same as for the first fixation duration (FFD) (Intercept: beta = 411.17, SE = 21.50, t = 19.13, p<0.00001, CI = 369.04–453.30; URA: beta = 164.74, SE = 17.54, t = 9.39, p<0.00001, CI = 130.35–199.13; AWN: beta = 177.45, SE = 17.55, t = 10.11, p<0.00001, CI = 143.05–211.84; and URA vs. AWN (p = 0.5308).

Re-fixation probability (RFP). RFP was higher after the first fixation on the unrelated anomalous target word (Intercept: beta = -0.9227, SE = 0.1075, t = -8.58, p<0.00001, CI = -1.1334– -0.7119, URA: beta = 0.6697, SE = 0.0894, t = 7.49, p<0.00001, CI = 0.4945–0.84497), and the RFP was higher after the first fixation on the anomalous word neighbour (AWN: beta = 0.62827, SE = 0.0895, t = 7.02, p<0.00001, CI = 0.4529–0.80365) than after the first fixation on the plausible target word. There was no difference in the RFP after the first fixation between the unrelated anomalous and anomalous word neighbour (p = 0.6858).

Semantic anomaly and word neighbour effects in FRPs. The first FRPs (Fig 4) showed several statistically significant clusters where the FRPs differed between the conditions (Fig 5). The FRPs differed between unrelated anomalous and plausible conditions, which was evident in a negative cluster in the central frontal scalp area (p < 0.0001) spanning from 138ms to 900ms, which remained stable throughout presence of the cluster, and a positive parietal occipital cluster (p<0.001) with a duration of 180ms to 900ms. The cluster was initially observed in the occipital fringe channels, from which it moved to central parietal locations (roughly 450ms onwards). Anomalous word neighbour and plausible conditions were also significantly different, as there was a fronto-central negative cluster from 97ms to 900ms (p<0.00001) that remained on the frontal sites for the duration of its presence and a positive cluster in the posterior occipital scalp area from 154ms to 900ms (p<0.001 which was initially constrained to occipital fringe channels whence it moved to central parietal sites (roughly 550ms onwards). Furthermore, unrelated anomalous and anomalous word neighbour were significantly different conditions, as we detected one cluster in the time window from 457ms to 698ms, which was positive in the centro-parietal regions (p <0.05).

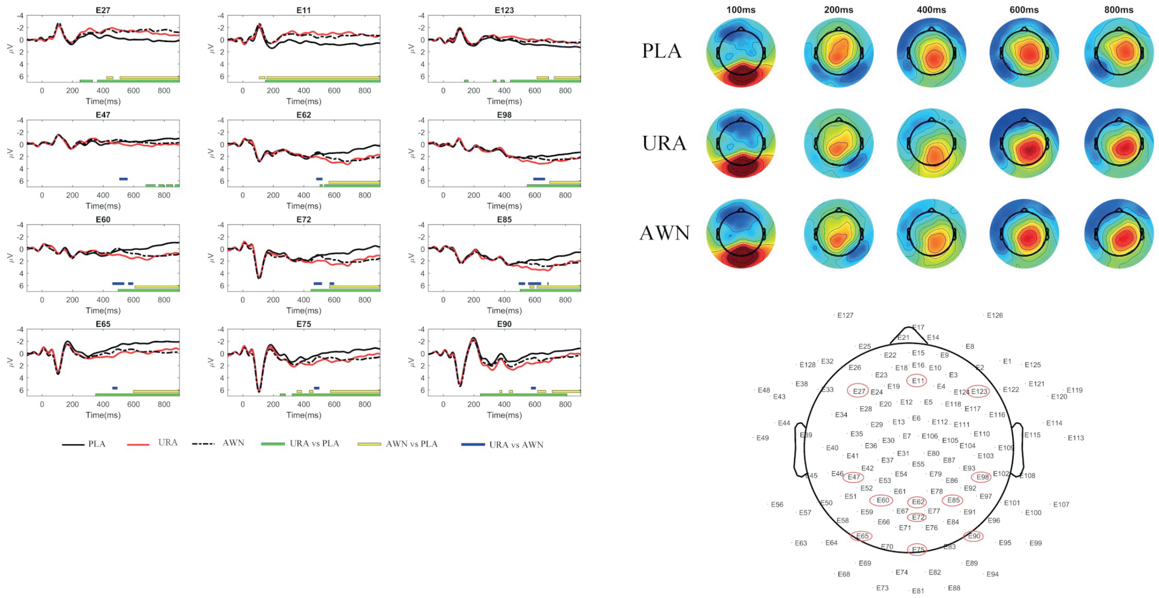


Fig 4. FRPs to the target words. Left Panel: FRPs from the first fixation on the target word on selected channels (E11, E27, E47, E60, E62, E65, E72, E75, E85, E90, E98 and E123 of the GSN-Hydrocel 128-channel cap (Electrical Geodesics Ltd)). Coloured bars under each channel highlight time-points with significant differences that belong to a cluster in the nonparametric permutation test. Right Panel: the associated topographies (all channels) over time in 12- to 13-5-year-old children (N = 66). The zero time-point is the onset of the first fixation on the target word.

https://doi.org/10.1371/journal.pone.0209741.g004

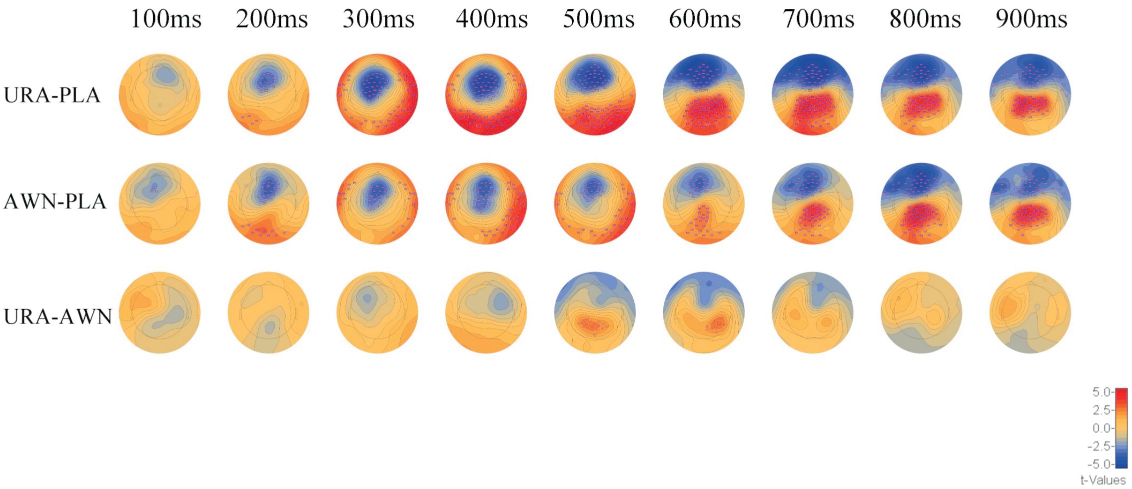


Fig 5. Statistical differences in the FRP amplitude for the first fixation on the target word. The scale depicts differences between the first element of the pair to the second element of the pair in values of the test statistic t. Rectangles signify that the sensor displays a statistically significant difference and belongs to a cluster. The colour of the stars within the squares signifies membership in a specific cluster. PLA = Plausible, URA = Unrelated anomalous and AWN = Anomalous word neighbour.

https://doi.org/10.1371/journal.pone.0209741.g005

Interim discussion

The behavioural results suggest clear and strong effects of semantic anomaly, as the FFDs, GDs and RFPs on the target words all show similar results with increased reading time to anomalous target words. Response accuracy was highest for the unrelated anomalous condition than for the responses for the plausible and anomalous word neighbour conditions. It is important to note, however, that the PFDs in the pre-target word do not show any effects of the semantic manipulation. This effectively shows that the processing shares the same trajectory until the first fixation on the target word. The topographical analysis of the FRPs time-locked to the first fixation onset show a systematic pattern of early frontal negativity (at ca 100–900ms) and late parietal positivity (at ca 500–900ms) for both of the anomalous target word categories. The differences between conditions that we observe are long and widespread. This may, in part, also be due to continuous reading being analogous of the RSVP with fast stimulus rates; fast presentation rates have been shown to diminish N400 and P600 amplitudes in adults [18]. The relatively late parietal positivity can fairly reliably be identified as P600. The P600 difference between anomalous word neighbour and plausible conditions was later than for the unrelated anomalous vs. plausible words. This may suggest that semantic processing of the word neighbour anomalous words of plausible words is delayed in relation to the semantic processing of unrelated anomalous words.

The frontal negativity observed for both anomalous conditions begins quite early and lasts through the remainder of the FRP waveform. One possibility is that this pattern is related to the age difference of our subjects compared to most of the N400 literature reporting adult responses. However, support for the age-related shift of the N400 topography in the previous literature is scarce and relatively old [55]. Thus, the argument for developmental differences is suggested here and would require more systematic developmental study with a traditional RSVP experimentation style. Another possibility is that this pattern is due to the difference between RSVP and FRP measurement techniques, although this is not supported by findings from previous research comparing RSVP and FRP in semantic processing [1]. A reasonable line of thought could also be to relate the current results to the evidence of early contributions of the dorsolateral frontal cortex to visual processing [56]. These contributions have been assumed to be of a top-down nature [57,58], and thus one cannot rule out the possibility that these effects are in fact also modulating the fixation durations or saccade targeting (refixation or regression) either directly or indirectly. The possible contribution of saccadic control is supported by the fact that the difference topography is directly on the top of structures that include frontal eye fields (FEFs) [56]. However this cannot be determined by topographical analysis alone. Thus, this claim would gain more support from source analysis and the interpretation would be constrained by providing information regarding the source time behaviour and location estimates.

Significant temporo-spatial mixing can be observed from the difference topographies. For example, the polarity flip of the frontal negativity (the early positivity in the fringe channels) seems to be mixing with the later parietal positivity. Temporo-spatial mixing makes determining the time of the onset of effects difficult, as is evident for example, for the emergence time of the P600 effect, which does not have any distinct or clearly identifiable onset. The spatial aspects of mixing can be counteracted by analysing the time course of the components from the group ICA procedure.

Group ICA analysis methods

Statistical analysis of the source activity

Source waveforms were statistically analysed in MATLAB utilising a method similar to the nonparametric cluster-based permutation approach used for the scalp FRPs above [54]. Condition effects were examined with pairwise comparisons using a Wilcoxon sign-rank test sample by sample. The Wilcoxon test was chosen due to the assumption that considering each time-point of the independent component responses to be normally distributed is unreasonable. The multiple comparisons problem was dealt with as follows. Consequent samples (minimum two) with significant differences were defined to be part of the same difference between the signals, and the duration of this difference is what we base our correction on. These ranges consisting of significant differences are later referred to as consequent samples clusters. After this, we employed a randomisation approach, where we flipped the label of the condition within each given subject randomly, re-ran the Wilcoxon sign-rank test sample by sample and stored the length of each cluster that was observed in the permutation distribution. This randomisation step was then repeated 10000 times. If the randomisation iteration resulted in a solution where there was not a single sample with a significant difference, that iteration was entered as zero into the permutation distribution. After the iterations, the permutation distribution contained the distribution of durations of significant differences when the data were randomised. As the null-observation clusters and the true-observation clusters are based on length and contain no negative values, the correction is one-tailed. Those true-observations that were longer than the 95^{th} percentile of the permutation distribution were accepted for further analysis, and those that were shorter than the 95^{th} percentile of the permutation were rejected. In addition to reporting the p-value of the cluster in relation to the permutation distribution, we also report Wilcoxon statistics and effect sizes for the mean of the cluster [53].

Source localisation

In order to visualise the sources behind the group level ICs, we forward projected components with significant differences after the correction procedure and then used cortically restricted

LORETA implemented in Brain Electrical Source Analysis (BESA, Besa GmbH, Gräfelfing, German) using an age-appropriate template MRI for 12-year-olds [59,60].

Group ICA results

The statistical analysis of source activity indicated significant differences between conditions in components 1, 2, 3, 4, 6, and 7. These components were then subjected to source analyses, the results of which are presented in Table 4 and Fig 6. The statistical results are presented in Fig 6 and in Table 5.

Discussion regarding the semantic anomaly effects in the independent components and their localisation

Group ICA analysis revealed that the differences observed in the scalp had six underlying topographical sources, which differed significantly in their time course in relation to the experimental manipulations. Four of those sources have a plausible contribution to the pattern of early frontal negativity (component 1, Fig 6D) and late posterior positivity (components 4, 3 and 6, Fig 6C, 6B and 6A respectively). Component 7 (Fig 6F), albeit with a brief effect, will be discussed as well as the source location and effect timeframe, which fit well to previous research on semantic processing. One of the components (component 2, Fig 6E) seems to be the manifestation of the early visual response P1 and thus not in the focus of this article. Here, we discuss the frontal source and then the parietal and temporal source observations.

Frontal projecting component

Frontal negativity (Fig 5), which dominates the difference between conditions in scalp waveforms, seems to be generated by a single IC component (component 1, Fig 6D). Source localisation analyses with cortically restricted LORETA suggest that it is localised in the right posterior frontal middle gyrus, in close proximity to Brodman areas 8 and 6. In PET [61] and some fMRI experiments [62,63], this location has been implicated as an FEF, an area that is associated with many features related to voluntary eye movements, including maintaining fixation, releasing fixation, triggering of eye movements, saccade amplitude and velocity [56] and inhibition of return [64]. In addition, contributions to higher cognitive functions have been proposed, for example, retaining saccade target locations over a couple of seconds, which is essentially a form of working memory [65], and deployment of both overt and covert spatial attention [66]. Transcranial magnetic stimulation of the left or right site of FEF causes saccade latencies to the contralateral direction be altered [56], and cortical electrical stimulation conjugates eye movements to the contralateral direction [67]. Our manipulation is associated with a difference in the next saccade target (refixation vs. regression), which in most cases is in the

Table 4. Source locations of the independent components.

Component	Location
IC 1	Prefrontal cortex, Right
IC 2	Occipital cortex
IC 3	Parietal cortex, Central
IC 4	Parietal cortex, Right
IC 6	Parietal cortex, Left
IC 7	Left temporal cortex

Table 4 summarises the source locations of the independent components that contained significant time course differences after minimum duration correction.

https://doi.org/10.1371/journal.pone.0209741.t004

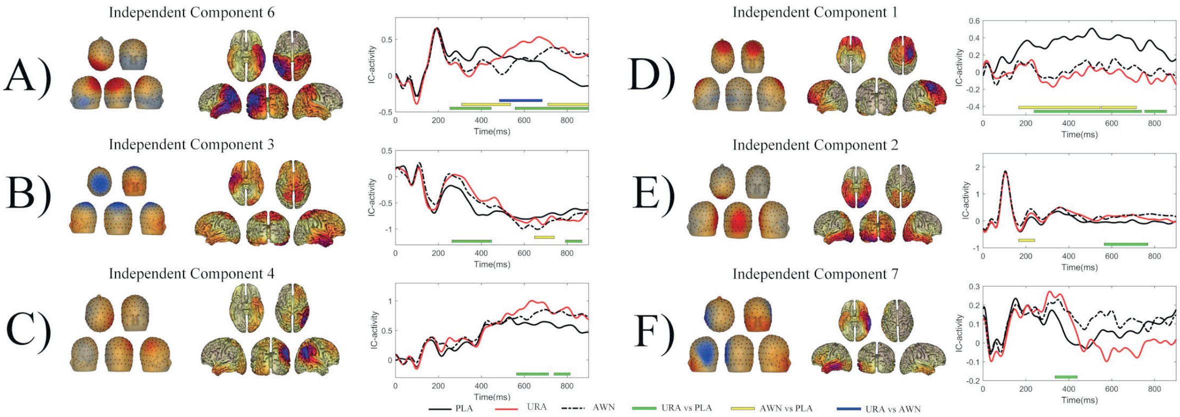


Fig 6. Condition effects in independent components. Topographies of the individual IC spatial filters are shown on the left side of the sub-figure. Red denotes a positive spatial filter weight and blue a negative spatial filter weight. Please note that depending on the activity of the component polarity on the scalp, the sensors change accordingly (for example, in sub-figure A, component 6 is contributing positive voltage to the scalp; however, roughly between 200ms to 500ms anomalous conditions have a more negative–or 'less positive'–contribution to the measured scalp activity). Cortically restricted LORETA solutions of the individual components are presented in the middle. Temporal evolutions of the components in each condition are shown on the right side of picture (solid black line denoting plausible condition, solid red line unrelated anomalous condition and dashed black line indicating anomalous word neighbour condition); bars under the waveforms denote statistically significant clusters of observations after minimum duration correction (green: unrelated anomalous vs. plausible, yellow: anomalous word neighbour vs. plausible and blue: unrelated anomalous vs. anomalous word neighbour). Acronyms: PLA = Plausible, URA = Unrelated anomalous and AWN = Anomalous word neighbour.

https://doi.org/10.1371/journal.pone.0209741.g006

Table 5. Statistical results from the independent component comparisons.

Component	Contrast	Cluster onset (ms)	Cluster offset (ms)	Cluster duration (ms)	Cluster duration p	Cluster mean Z	Cluster mean p	Cluster mean R
1	URA-PLA	238	738	501	<0.0001	-4.168	<0.0001	-0.363
	URA-PLA	754	854	101	<0.05	-3.542	<0.001	-0.308
	AWN-PLA	167	547	379	<0.001	-3.600	<0.001	-0.313
	AWN-PLA	551	714	164	<0.01	-2.610	<0.01	-0.227
2	URA-PLA	565	767	203	<0.01	2.661	<0.01	0.232
	AWN-PLA	157	273	117	<0.05	2.986	<0.01	0.260
3	URA-PLA	263	447	185	<0.01	2.891	<0.01	0.252
	URA-PLA	792	869	77	<0.05	-2.495	<0.01	-0.217
	AWN-PLA	648	739	92	<0.05	-2.571	<0.01	-0.224
4	URA-PLA	564	709	149	<0.01	3.338	<0.001	0.291
	URA-PLA	739	813	75	<0.05	2.693	<0.01	0.234
6	URA-PLA	254	445	192	<0.01	-2.776	<0.01	-0.241
	URA-PLA	558	899	342	<0.001	4.124	<0.0001	0.359
	AWN-PLA	309	535	227	<0.01	-3.229	<0.001	-0.281
	AWN-PLA	710	899	190	<0.01	3.804	<0.001	0.331
	URA-AWN	484	683	200	<0.01	3.804	<0.01	0.255
7	URA-PLA	342	434	93	<0.05	2.967	<0.01	0.258

Table 5 summarises statistically significant condition differences in the independent components. Cluster onset is the first time-point where the cluster is present while the cluster offset is the last. Cluster duration is the amount of time-points belonging to the cluster, and the cluster duration p is the p-value category of that cluster duration. Cluster mean Z, P and R are Wilcoxon sign-rank test parameters when the contrast is tested over the average of the cluster time window.

Note. PLA = Plausible, URA = Unrelated anomalous and AWN = Anomalous word neighbour.

https://doi.org/10.1371/journal.pone.0209741.t005

earlier parts of the sentence, contralateral of the right FEF. Localisation of the early frontal activity to the right FEF, the pattern of saccade targeting differences and previous findings on associated functions of FEF together strongly suggest that the early frontal activity difference that we observe is associated with the first fixation behavioural effects (FFD, refixation probability) that we observed.

Parietal and temporal projecting components

In the time behaviour of the independent components projecting to the parietal scalp, a biphasic N400/P600-like pattern can be observed. The biphasic N400/P600 pattern appears in two components (components 3 and 6, Fig 6B and 6A, respectively) for both the anomalous conditions. One component (component 4, Fig 6C) only contributes to the P600 effect of the unrelated anomalous condition. As a whole, the parietal brain source structure (in the LORETA analysis) is highly reminiscent of the beamforming results in recent research on semantic and syntactic processing [26], where N400m was localised to left superior temporal and posterior frontal regions and P600m was found to have a plurality of sources, including the bilateral frontal, posterior temporal and parietal regions. In our data, the pattern was similar, with the two components exhibiting a biphasic N400/P600 pattern localised in the left parietal and left posterior temporal areas (component 6) and in the central parietal areas, spreading to the right anterior temporal regions (component 3). Component 4 contributing only to the P600 pattern of the unrelated anomalous condition was localised in the right parietal regions. We will next discuss each of these parts of the parietal pattern in detail.

Component 4 was localised in the right parietal cortex in the approximate area of the right angular gyrus (AG), in the vicinity of Brodman areas 39 and 40 (Fig 5C). AG has been most consistently associated with semantic processing. This is especially the case for the left AG area, although less strong yet equally consistent findings of right AG activation in semantic processing exist as well [68]. Right AG activity has also been found to be associated with cognitive conflict without semantic constraints. This was apparent in research comparing a flanker, stroop and sentence plausibility comprehension tasks, where it was found that the left AG would react to the semantic conflict only while the right AG would respond to all three types of conflict equally [69]. Inhibition of inappropriate responses during no/go tasks has also been associated with the right AG [70]. Thus, in our case, where independent component 4 is contributing only to the P600 pattern of the unrelated anomalous condition, a right AG origin might be related to the capacity of the right AG to process cognitive conflict in several modalities. In the case of the unrelated anomalous condition, the conflict might be gross enough to also be processed in the right AG in the late P600 phase, although the detection of single letter deviation is not strong enough to elicit a significantly different response in the right AG, which is not linguistically specialised.

The independent component 3 topography shows an archetypical centro-parietal N400/P600 complex, and the time course of the component complements this observation (Fig 6B). Unrelated anomalous words significantly modulate both N400 and P600 patterns, whereas anomalous word neighbours modulate only the P600 pattern. The cortical LORETA model shows a pattern of two distinguishable loci of activation: the bilateral superior parietal cortex and right anterior temporal cortex. However, one needs to be particularly cautious when interpreting distributed source model solutions implicating anterior temporal regions, as noise in the frontal and face region scalp sensors tends to migrate to these locations, with a gradual increase of source strength towards the temporal pole being a specific sign of possible artefactual contamination (as stated in discussion with Monto. S. on February 2, 2018, a University researcher at the Centre of Interdisciplinary Brain research at University of Jyväskylä). This is not, however, the case in our data. The locus of the temporal activity is located at the anterior part of the middle temporal gyrus rather than at the anterior pole. If our localisation on this part was actual rather than artefactual, our effects in component 3 could reflect semantic features of the word, as anterior temporal lobes have been argued to act as semantic storage hubs [71]. Anterior temporal lobe contribution to very precise semantic features is particularly evident in semantic dementia, in which bilateral degradation of anterior temporal structures is typical [72,73]. The other areas implicated in our LORETA source localisation solution of component 3 were the bilateral superior parietal areas, which were also implicated in recent research on semantic violations [26].

Component 6 (Fig 6A) contains a highly interesting temporal pattern of activation. First, it displays a contribution to the biphasic N400/P600 for both anomalous conditions. Unlike the other components, component 6 also displays significant differences between the anomalous conditions, and this is essentially because there is delay in the onset of the N400/P600 pattern for the word neighbour anomalous condition. The difference that contributes more negative voltage for the word neighbour anomalous ending than for the plausible ending begins roughly 50ms later than the negative difference in responses between the unrelated anomalous and plausible conditions. Furthermore, the onset of a positive difference between the word neighbour anomalous and plausible conditions occurs roughly 150ms later than the difference between the unrelated anomalous and plausible words. Second, component 6 localises in the LORETA analysis to a broad patch of the left temporal and parietal cortex. Given their heavy involvement in reading and semantic processing in the current experiment, these regions are expected to show activation. For example, the left AG, which is one of the areas implicated in our cortical LORETA model of component 6, is largely involved in all aspects of semantic processing that require concept retrieval and concept integration [68]. Moreover, the left AG has been claimed to provide semantic constraints during language comprehension [74] and to engage when semantic associations are made [75].

The left temporal lobe was also part of the cortical LORETA model of component 6 (Fig 5A). In previous research, the left temporal lobe has been found to be a major contributor to N400 effects in brain lesion, intracranial and MEG studies [24]. These findings converge with fMRI studies, which commonly find middle temporal gyrus activation associated with semantic context manipulations [24]. Overall, the source pattern of component 6 is very consistent with previous research on semantic processing.

In addition, component 7 (Fig 6F) displays a significant difference between URA and PLA around 400ms, and the LORETA model implicates the left temporal lobe as a potential source for activity reflected in the component. However the significance pattern and the waveform of the component make it difficult to interpret how component 7 relates to the pattern observed at component 6. It may well be that the short-lived difference reflects sensitivity to large-scale anomalies but a lack of sensitivity to recognise anomalies of one letter.

The temporal pattern and LORETA solution of the forward-projected independent components converge on an interpretation that the semantic processing of anomalous word neighbours is delayed in relation to the semantic processing of unrelated anomalous words. It could be that the assignment of the semantic role of the anomalous word neighbour becomes delayed because sentence context would favour the semantic role of the plausible word over the anomalous word. The negative contributing cluster is longer for the contrast between the anomalous word neighbour and the plausible word than between the unrelated anomalous word and the plausible word. This could indicate that the assignment of a semantic role or fitting the anomalous word neighbour to the context is more laborious, as the semantic system receives interference from the plausible word that is just one letter away. The delay of onset and longer duration in the earlier negative-contributing phase builds up to a greater delay at the onset of the positive contributing phase. It could be that difficulty in the assignment of the semantic role results in a heavily delayed onset of the integration processes, reflected by the P600 that independent component 6 seems to be contributing to in the later parts of the waveform.

Summary discussion

Our findings show that when reading sentences with a semantic manipulation of the last word, two types of semantic anomalies (the unrelated anomalous words and the anomalous word neighbours of plausible words) produce similar effects on the FFDs, GDs and the RFP patterns in 12–13.5-year-old children in the grade 6 The FRPs for the unrelated anomalous words share a highly similar morphology with the anomalous word neighbours up to about 500ms after the fixation onset on the target word. The most prominent difference between FRPs for the anomalous and plausible words is the frontal negativity from as early as 100ms to the end of the epoch. The brain source analysis of this negative activation suggested the FEF as a likely source. FEF is an area that has a close relationship with eye-movement execution and contributes to a variety of visuo-spatial attention functions [56]. A parietal positivity, which can be identified as a P600 response, starts to emerge at about 500ms, but only for the contrast between unrelated anomalous and plausible sentence endings. In the responses for the contrast between word neighbour anomalous and plausible sentence endings, this parietal positivity can be observed later, starting at around 700ms. P600 response has been linked to the integration of sentence meaning [3] and a conscious perception of semantic anomalies on the whole sentence level [12]. The time behaviour of the underlying independent components revealed that this delay is also present in an earlier time window around 250ms (see Fig 6A). The source analysis of the components projecting to the parietal scalp areas revealed bilateral contributions from the angular gyri and temporal lobes, both related to semantic processing [24].

The response accuracy was similar to the plausible endings and the anomalous orthographic word neighbours, and it was clearly highest for the unrelated anomalous sentence ending. We propose that the effects of the FFD, GD, RFP and frontal negativity stem from encountering anomalous visual features in relation to sentence context and that the deviation of a single letter from the plausible word is sufficient to elicit these effects. The effects in P600 amplitude and timing, in components projecting to parietal scalp and in response accuracy suggest that the semantic processing of anomalous word neighbour endings is slower than the processing of unrelated semantic anomalies.

FFDs and GDs were longer and re-fixation probability was higher for both the word neighbour anomalous and unrelated anomalous target words than for the plausible target words. Longer GDs for semantic anomalies are typical [6,76], but also FFD effects in relation to semantic manipulations are found in more recent research (see [1,7–10]). The lack of an FFD effect between the unrelated anomalous and word neighbour target words shows that a deviation of a single letter from a plausible word was sufficient to modulate the fixation durations to a similar extent as the anomalous target word without orthographic similarity. In principle, one could expect that processes reflected by the early occipital components P1 and N1 could precede these behavioural effects. However, the absence of P1 and N1 modulations in the current experiment is reasonable since the common modifiers of P1, such as low-level visual features including stimulus size [77], and of N1, such as word frequency [78,79], word length [78] and lexicality [79], were controlled for.

Instead we found early frontal negativity that separates the anomalous conditions from the plausible condition and precedes the gaze behaviour effects. The LORETA source analysis of the frontal independent component resembling the frontal negativity pinpointed the right FEF, an area implicated in eye-movement control [56], as the source of the response, which strongly suggests that the patterns of gaze behaviour and frontal activity are interrelated. FEF also receives fast connections from the primary visual cortex, so it is possible that the effects we observed between conditions are visual rather than semantic and may be related to pre-saccadic parafoveal prediction of post-saccadic foveal input, and thus they may be an effect of visual rather than semantic processing [32,33,80].

In summary, complementary evidence from the FFDs, RFP, GD and frontal negativity suggest that there is a mechanism in the pre-300ms time period that is able to differentiate read words from plausible words on a single-letter level. In all of these variables, the unrelated anomalous and word neighbour anomalous target words differed significantly from the plausible target words.

The effects of semantic context and expectation are systematically found at later latencies, from 200ms onwards from the stimulus onset, and they are reflected in the extensively studied N400 component [11]. Severe semantic anomalies have also been found to elicit P600 responses [3,16]. We can identify P600-like modulation in our scalp waveforms, but the archetypical N400 with central-parietal topography is missing from the scalp waveform. However, analysing the underlying components separately, we can identify two components that project more negative voltages to the parietal scalp between 200–500ms in anomalous conditions and also localise to known N400 generator locations in our LORETA models. Thus, it seems that the frontal negativity that we observe in the scalp waveform is hiding the N400 modulations.

Our N400 results are quite different from those presented in similar RSVP designs in English [14], where N400 was found to be diminished for word neighbours. Authors interpreted that the recognition facilitation from the context spreads to the orthographic information of the neighbouring words. The differences in our results and those of previous RSVP results could stem from various differences in the experimental setups and participants populations–one obvious difference being the utilisation of the FRP technique on our part, which effectively introduces differences in timing between words, parafoveal previews including pre-saccadic attention and volition into the mix. Another plausible reason is the age difference between our participants and those utilised in previous research [14].

In previous research, P600 has been connected to the integrative phase of sentence comprehension [3], response to the impossibility of the sentence [16,81] and syntax anomalies [15]. It has been suggested, however, that rather than signifying the encountering of an anomaly, the P600 effect would reflect re-attending to the complete sentence in an attempt to revise the initial parse of the sentence [3,19–21] or a conscious perception of the semantic anomaly at the whole sentence level [12]. The occurrence of P600 in the anomalous conditions could thus indicate that encountering clearly anomalous words triggers a re-evaluation process for the whole sentence. However, for the word neighbour anomalous condition, P600 seems to have significantly delayed onset, which seems to begin in the earlier N400-contributing phase and accumulate to a more substantial delay in the P600-contributing phase. The accumulated delay of the P600 effect in the neighbour condition could imply that the comprehension process does not orient on the preceding sentence immediately for anomalous word neighbours. The re-evaluation of the preceding sentence could be inhibited by the context, that is, assigning and integrating a semantic role of a plausible word instead of the read anomalous word neighbour target word. The comprehension system might be inclined to consider the deviating letter of the word neighbour as a typographical error and not re-orient the comprehension process immediately to the preceding context. The sentences with neighbour manipulations are mostly considered nonsense at some later point, as indicated by the relatively high response accuracy.

Conclusion

Based on the data pattern, we arrive at two broad conclusions. First, the observed FFD, RFP, GD and frontal parts of the fixation-related brain response form a pattern that implicates that a divergence of a single letter from a more expected plausible sentence is sufficient to significantly alter gaze behaviour. It is important to note that the difference in frontal activity precedes the difference in gaze behaviour. However, we do not think that these effects are directly related to semantic processing but rather reflect encountering anomalous visual features in relation to sentence context. This is supported by our source localisation of the component resembling the frontal activity in the FEF, a brain area that in addition to making a direct contribution to eye movements also contributes to a variety of visuo-spatial attention functions and can thus be considered a part of the visual processing hierarchy. Further, the very early onset of these frontal effects supports the notion that these effects are related to visual rather than semantic processing.

The second data pattern relates more directly to semantic processing of the anomalous words. The delayed P600 of the scalp waveform for the word neighbour anomalous condition indicates that it takes longer for a single-letter word anomaly to reach the integration stage and perhaps conscious detection compared to a full-word anomaly. The underlying component waveforms indicate that this delay in fact begins earlier in the processing, in the time window corresponding to the archetypical N400. We thus argue that the delay associated with the word neighbour anomalous condition is a result of the difficulty in assigning the semantic role of the word, as it is competing against the semantic role of the plausible word that is just one letter away. This competition is eventually resolved correctly, as can be observed from the delayed but significant onset of P600 for the word neighbour anomalous ending and the relatively high response accuracy for this condition.

Limitations, boundary conditions and discussion

There are several methodological limitations in the current study. First, to fully control the contributions of parafoveal processing to the process under study, we could have used parafoveal masking of the target word until it was foveated. The hardware that we utilise certainly makes this possible. Forcing parafoveal prediction error with the parafoveal mask would have allowed us to probe whether the pattern that we observed in the frontal scalp and in component 1 was truly a result of a prediction error based on pre-saccadic attention. Then again, it could be argued that including parafoveal masking in our experiment would require us to go through all combinations of mask type with counterbalancing, resulting in nine conditions instead of three, which would have increased the requirements related to the experiment length and population size radically to achieve required signal quality and statistical power. Nevertheless, such a study would be particularly interesting for testing the hypothesis about pre-saccadic attention generating predictions in FEF of the next retinal input during reading.

Second, there is the crucial issue regarding the temporal overlap of activation from previous and subsequent fixations, which effectively creates a situation where one cannot be exactly certain of the degree to which other fixations besides the fixation of interest are contributing to the pattern observed on the scalp. In some studies, this is resolved by a procedure called 'fixation matching', where FRPs are matched to eliminate the possible effects of diverging gaze paths. This means that from the condition that has a shorter fixation duration on average, the longer fixations are kept in the FRP, and from the condition that has longer fixations on average, the shorter fixations are selected for the FRP. However, this creates a psychologically biased dataset, as fixation durations and other gaze events are thought to reflect cognitive operations and their processing costs [35]. We thus adopt the position that 'fixation matching' compromises the ecological validity of FRP studies, and thus the overlap issue should be addressed with signal processing techniques. We have addressed this issue through a group-level ICA procedure and a separate analysis of the unmixed components. Examination of component 2 (see Fig 6E) highlights the first clear-cut feature that we can observe in the waveforms, the P1/λ-response. This component localises to the primary visual cortex and reflects the initial processing of the low-level features of the foveated stimuli. Thus, if we expect the subsequent fixations to confound the conditions differently, we should observe signs of it in the later parts of the time course of component 2. Of course, this applies mainly to the timing distribution differences between conditions, not to the semantic or other features to which the primary visual cortex is insensitive. From the statistical results of component 2, we see that there is a late difference between the unrelated anomalous and plausible words between 562ms and 774ms. As it is highly unlikely that this difference would reflect semantic processing, we conclude that it is a manifestation of the difference in the fixation timing of the subsequent fixations between conditions. It would take roughly 100ms (from the first positive peak of component 2 to the first positive peak of component 6 and component 4) for effects from this difference in the primary visual cortex to influence the P600 results. All of our results begin prior to this time-point of 660ms. Hence, if the confounding influence is there, it is very minor. If the activity of component 2 would suggest otherwise, the temporal overlap could be further investigated with recently developed signal processing techniques. For example ADJAR (Adjacent Response algorithm)[82,83] or GLM-based linear deconvolution[84] would be suitable (with concrete adjustments to pre-processing pipeline).

Third, we did not rotate sentence frames across conditions, so each target word was always associated with a specific sentence. Therefore, we cannot eliminate the possibility that our results could be affected by random differences between the sentences rather than by our manipulation within these sentences. However, the behavioural results from PFDs argue against this possibility. If the difference between sentences is driving our effects, it should also be present in the PFDs. We do not see any modulation that would signify confounds from the sentences preceding the target words.

Fourth, the target word was sentence final. This is generally not done in eye-movement studies in order to avoid confounds from wrap-up effects that are associated with the final sentence words. However a recent review of ERP studies illustrates that this assumption of a ‘general wrap up’ process at the sentence end is not supported by empirical evidence and suggests that psycholinguistic research has been damaged by the last word target word avoidance dogma [85]. Moreover, given that our focus is on late responses related to semantic processing, where the context provided by the whole sentence is crucial, we think that our placement of the target word is justified. Semantic anomaly manipulation might cause different reactions if placed in earlier parts of the sentence, which would be an interesting research subject.

Fifth, there is confound in the cloze probability of the plausible condition and the plausible word neighbour of the anomalous word neighbour and thus differences between the plausible and two other conditions can partially be due difference in predictability. However the semantic anomaly in the stimuli is clear and thus we consider it to be the driving influence behind the effects. It is important to note that this confound does not impact comparisons between unrelated anomalous and anomalous word neighbour conditions and thus does not impact the main finding of delayed processing of the anomaly in anomalous word neighbour condition.

Finally we studied 12–13.5-year-old children while typical cognitive neuroscience experiments employ university students around 20 years of age. Our sample is a sub-population of the eSeek-project (eSeek! Internet and Learning Difficulties–Multidisciplinary approach for understanding information seeking in new media), and this article essentially describes how school children process semantic information. The pattern of results might change due to developmental effects if this experiment were to be conducted with a representative adult sample.

Acknowledgments

This research was funded by grant 274022 from the Academy of Finland to Paavo H.T. Leppänen. The authors would like to thank all of the research colleagues, research assistants and students involved in the research project.

Author Contributions

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

Data curation: Otto Loberg.

Formal analysis: Otto Loberg.

Funding acquisition: Paavo H. T. Leppänen.

Investigation: Otto Loberg.

Methodology: Otto Loberg, Jarmo A. Hämäläinen.

Project administration: Paavo H. T. Leppänen.

Supervision: Jarkko Hautala, Jarmo A. Hämäläinen, Paavo H. T. Leppänen.

Visualization: Otto Loberg.

Writing – original draft: Otto Loberg, Jarkko Hautala, Jarmo A. Hämäläinen, Paavo H. T. Leppänen.

Writing – review & editing: Otto Loberg, Jarkko Hautala, Jarmo A. Hämäläinen, Paavo H. T. Leppänen.

References

1. Metzner P, von der Malsburg T, Vasishth V, Rösler F. The Importance of Reading Naturally: Evidence from Combined Recordings of Eye Movements and Electric Brain Potentials. 2016; https://doi.org/10.1111/cogs.12384 PMID: 27307404
2. Dambacher M, Kliegl R. Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. Brain Res. 2007; 1155: 147–162. https://doi.org/10.1016/j.brainres.2007.04.027 PMID: 17499223
3. Friederici AD, Weissenborn J. Mapping sentence form onto meaning: The syntax-semantic interface. Brain Research. 2007. pp. 50–58. https://doi.org/10.1016/j.brainres.2006.08.038 PMID: 16956590
4. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull. 1997; 124: 372–422. https://doi.org/10.1037/0033-2909.124.3.372
5. Rayner K, Warren T, Juhasz BJ, Liversedge SP. The Effect of Plausibility on Eye Movements in Reading. J Exp Psychol Learn Mem Cogn. 2004; 30: 1290–1301. https://doi.org/10.1037/0278-7393.30.6.1290 PMID: 15521805
6. Warren T, McConnell K, Rayner K. Effects of context on eye movements when reading about possible and impossible events. J Exp Psychol Learn Mem Cogn. 2008; 34: 1001–1010. https://doi.org/10.1037/0278-7393.34.4.1001 PMID: 18605885
7. Abbott MJ, Staub A. The effect of plausibility on eye movements in reading: Testing E-Z Reader's null predictions. J Mem Lang. 2015; 85: 76–87. https://doi.org/10.1016/j.jml.2015.07.002
8. Matsuki K, Chow T, Hare M, Elman JL, Scheepers C, McRae K. Event-based plausibility immediately influences on-line language comprehension. J Exp Psychol Learn Mem Cogn. NIH Public Access; 2011; 37: 913–34. https://doi.org/10.1037/a0022964 PMID: 21517222
9. Veldre A, Andrews S. Is Semantic Preview Benefit Due to Relatedness or Plausibility? J Exp Psychol Hum Percept Perform. 2016; 42: 939–952. https://doi.org/10.1037/xhp0000200 PMID: 26752734
10. Veldre A, Andrews S. Parafoveal preview benefit in sentence reading: Independent effects of plausibility and orthographic relatedness. Psychon Bull Rev. 2017; 24: 519–528. https://doi.org/10.3758/s13423-016-1120-8 PMID: 27418260

11. Kutas M, Federmeier KD. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol. 2011; 62: 621–47. https://doi.org/10.1146/annurev.psych.093008.131123 PMID: 20809790
12. Sanford AJ, Leuthold H, Bohan J, Sanford AJS. Anomalies at the borderline of awareness: an ERP study. J Cogn Neurosci. 2011; 23: 514–523. https://doi.org/10.1162/jocn.2009.21370 PMID: 19925201
13. Rossell SL, Price C, Nobre AC. The anatomy and time course of semantic priming investigated by fMRI and ERPs. Neuropsychologia. 2003; 41: 550–564. https://doi.org/10.1016/S0028-3932(02)00181-1 PMID: 12638581
14. Laszlo S, Federmeier KD. A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. J Mem Lang. 2009; 61: 326–338. https://doi.org/10.1016/j.jml.2009.06.004 PMID: 20161064
15. Osterhout L, Holcomb PJ. Event-related brain potentials elicited by syntactic anomaly. J Mem Lang. 1992; 31: 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z
16. DeLong KA, Quante L, Kutas M. Predictability, plausibility, and two late ERP positivities during written sentence comprehension. Neuropsychologia. 2014; 61: 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016 PMID: 24953958
17. Van Petten C, Luka BJ. Prediction during language comprehension: Benefits, costs, and ERP components. International Journal of Psychophysiology. 2012. pp. 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015 PMID: 22019481
18. Tanner D, Grey S, van Hell JG. Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. Psychophysiology. 2016;0. https://doi.org/10.1111/psyp.12788 PMID: 27859315
19. Friederici AD, Hahne A, Mecklinger A. Temporal structure of syntactic parsing: early and late event-related brain potential effects. JExpPsycholLearnMemCogn. 1996; 22: 1219–1248.
20. Hahne A, Friederici AD. Electrophysiological Evidence for Two Steps in Syntactic Analysis: Early Automatic and Late Controlled Processes. J Cogn Neurosci. 1996; 11: 194–205.
21. O'Rourke PL, Van Petten C. Morphological agreement at a distance: Dissociation between early and late components of the event-related brain potential. Brain Res. Elsevier B.V.; 2011; 1392: 62–79. https://doi.org/10.1016/j.brainres.2011.03.071 PMID: 21466794
22. Nobre AC, McCarthy G. Language-Related ERPs: Scalp Distributions and Modulation by Word Type and Semantic Priming. J Cogn Neurosci. 1994; 6: 233–255. https://doi.org/10.1162/jocn.1994.6.3.233 PMID: 23964974
23. Nobre AC, Mccarthy G. Language-Related Field Potentials in the Anterior-Medial Temporal Lobe: II. Effects of Word Type and Semantic Priming. J Neurosci. 1995; 15: 1090–98. PMID: 7869085
24. Petten C Van Luka BJ. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain Lang. 2006; 97: 279–293. https://doi.org/10.1016/j.bandl.2005.11.003 PMID: 16343606
25. Helenius P, Salmelin R, Service E, Connolly JF. Semantic cortical activation in dyslexic readers. J Cogn Neurosci. 1999; 11: 535–550. https://doi.org/10.1162/089892999563599 PMID: 10511642
26. Kielar A, Panamsky L, Links KA, Meltzer JA. Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. Neuroimage. 2015; 105: 507–524. https://doi.org/10.1016/j.neuroimage.2014.11.016 PMID: 25463470
27. Friederici AD, Rüschemeyer S-A, Hahne A, Fiebach CJ. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb Cortex. 2003; 13: 170–177. https://doi.org/10.1093/cercor/13.2.170 PMID: 12507948
28. Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RS. Functional anatomy of a common semantic system for words and pictures. Nature. 1996. pp. 254–256. https://doi.org/10.1038/383254a0 PMID: 8805700
29. Price C. A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage. 2012; 62: 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062 PMID: 22584224
30. Ditman T, Holcomb PJ, Kuperberg GR. An investigation of concurrent ERP and self-paced reading methodologies. Psychophysiology. Blackwell Publishing Inc; 2007; 44: 927–935. https://doi.org/10.1111/j.1469-8986.2007.00593.x PMID: 17850242
31. Barber HA, van der Meij M, Kutas M. An electrophysiological analysis of contextual and temporal constraints on parafoveal word processing. Psychophysiology. 2013; 50: 48–59. https://doi.org/10.1111/j.1469-8986.2012.01489.x PMID: 23153323
32. Li H-H, Barbot A, Carrasco M. Saccade Preparation Reshapes Sensory Tuning. Curr Biol. Elsevier Ltd; 2016; 26: 1564–1570. https://doi.org/10.1016/j.cub.2016.04.028 PMID: 27265397

33. Zirnsak M, Moore T. Saccades and shifting receptive fields: anticipating consequences or selecting targets? Trends Cogn Sci. 2014; 18: 621–628. https://doi.org/10.1016/j.tics.2014.10.002 PMID: 25455690

34. Kornrumpf B, Niefind F, Sommer W, Dimigen O. Neural correlates of word recognition: A systematic comparison of natural reading and RSVP. J Cogn Neurosci. 2016; 1–35.

35. Dimigen O, Sommer W, Hohlfeld A, Jacobs AM, Kliegl R. Co-Registration of Eye Movements and EEG in Natural Reading: Analyses & Review. Rev J Exp Psychol Gen. 2011; 140: 552–572. https://doi.org/10.1037/a0023885 PMID: 21744985

36. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004; 134: 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009 PMID: 15102499

37. Plöchl M, Ossandón JP, König P. Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. Front Hum Neurosci. 2012; 6: 278. https://doi.org/10.3389/fnhum.2012.00278 PMID: 23087632

38. Dimigen O, Kliegl R, Sommer W. Trans-saccadic parafoveal preview benefits in fluent reading: A study with fixation-related brain potentials. Neuroimage. 2012; 62: 381–393. https://doi.org/10.1016/j.neuroimage.2012.04.006 PMID: 22521255

39. Kretzschmar F, Bornkessel-Schlesewsky I, Schlesewsky M. Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. Neuroreport. 2009; 20: 1613–1618. https://doi.org/10.1097/WNR.0b013e328332c4f4 PMID: 19884865

40. Frey A, Ionescu G, Lemaire B, López-Orozco F, Baccino T, Guérin-Dugué A. Decision-making in information seeking on texts: an eye-fixation-related potentials investigation. Front Syst Neurosci. 2013; 7: 1–22. https://doi.org/10.3389/fnsys.2013.00001

41. Makeig S, Bell AJ, Jung T-P, Sejnowski TJ. Independent Component Analysis of Electroencephalographic Data. Adv Neural Inf Process Syst. 1996; 8: 145–151. https://doi.org/10.1109/ICOSP.2002.1180091

42. Wessel JR. Testing Multiple Psychological Processes for Common Neural Mechanisms Using EEG and Independent Component Analysis. Brain Topogr. Springer US; 2016; 1–11. https://doi.org/10.1007/s10548-015-0457-z

43. Onton J, Westerfield M, Townsend J, Makeig S. Imaging human EEG dynamics using independent component analysis. Neurosci Biobehav Rev. 2006; 30: 808–822. https://doi.org/10.1016/j.neubiorev.2006.06.007 PMID: 16904745

44. Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, et al. Dynamic brain sources of visual evoked responses. Science (80-). 2002; 295: 690–694. https://doi.org/10.1126/science.1066168 PMID: 11809976

45. Lio G, Boulinguez P. How Does Sensor-Space Group Blind Source Separation Face Inter-individual Neuroanatomical Variability? Insights from a Simulation Study Based on the PALS-B12 Atlas. Brain Topogr. Springer US; 2016; 1–14. https://doi.org/10.1007/s10548-015-0457-z

46. Lindeman J. ALLU-Ala-asteen lukutesti [ALLU-Reading Test for Primary School]. University of Turku, Finland. The Center for Learning Research.; 1998. https://doi.org/10.1016/S0022-3913(03)00368-8

47. Nevala J, Lyytinen H. Sanaketjutesti [Word chain test]. Jyväskylä: Niilo Mäki Institute; 2000.

48. Eklund K, Torppa M, Aro M, Leppanen PHT, Lyytinen H. Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. J Educ Psychol. 2015; 107: 126–140. https://doi.org/10.1037/a0037121

49. Language Bank of Finland. Frequency lexicon of the Finnish newspaper language [Internet]. 2007. Available: http://urn.fi/urn:nbn:fi:lb-201405271

50. Engbert R, Mergenthaler K. Microsaccades are triggered by low retinal image slip. Proc Natl Acad Sci. National Acad Sciences; 2006; 103: 7192–7197. https://doi.org/10.1073/pnas.0509557103 PMID: 16632611

51. Lee T-W, Girolami M, Sejnowski TJ. Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub—Gaussian and Super—Gaussian Sources. Neural Comput. 1999; 11: 409–433. https://doi.org/10.1162/089976699300016719

52. Bridwell DA, Rachakonda S, Silva RF, Pearlson GD, Calhoun VD. Spatiospectral Decomposition of Multi-subject EEG: Evaluating Blind Source Separation Algorithms on Real and Realistic Simulated Data. Brain Topogr. Springer US; 2016; 1–15. https://doi.org/10.1007/s10548-015-0457-z

53. Pallant J. SPSS survival manual. 2007; https://doi.org/10.1046/j.1365-2648.2001.2027c.x

54. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 2007; 164: 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024 PMID: 17517438

55. Holcomb PJ, Coffey S, Neville H. Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. Dev Neuropsychol. 1992; 8: 203–241. https://doi.org/10.1080/87565649209540525

56. Vernet M, Quentin R, Chanes L, Mitsumasu A, Valero-Cabré A. Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. Front Integr Neurosci. Frontiers Media SA; 2014; 8: 66. https://doi.org/10.3389/fnint.2014.00066 PMID: 25202241

57. Ruff CC, Blankenburg F, Bjoertomt O, Bestmann S, Freeman E, Haynes JD, et al. Concurrent TMS-fMRI and Psychophysics Reveal Frontal Influences on Human Retinotopic Visual Cortex. Curr Biol. 2006; 16: 1479–1488. https://doi.org/10.1016/j.cub.2006.06.057 PMID: 16890523

58. Taylor PCJ, Nobre AC, Rushworth MFS. FEF TMS affects visual cortical activity. Cereb Cortex. 2007; 17: 391–399. https://doi.org/10.1093/cercor/bhj156 PMID: 16525126

59. Richards JE, Sanchez C, Phillips-Meek M, Xie W. A database of age-appropriate average MRI templates. Neuroimage. 2016; 124: 1254–1259. https://doi.org/10.1016/j.neuroimage.2015.04.055 PMID: 25941089

60. Richards JE, Xie W. Brains for All the Ages: Structural Neurodevelopment in Infants and Children from a Life-Span Perspective. Adv Child Dev Behav. 2015; 48: 1–52. https://doi.org/10.1016/bs.acdb.2014.11.001 PMID: 25735940

61. Paus T. Location and function of the human frontal eye-field: A selective review. Neuropsychologia. Pergamon; 1996; 34: 475–483. https://doi.org/10.1016/0028-3932(95)00134-4

62. Luna B, Thulborn KR, Strojwas MH, McCurtain BJ, Berman RA, Genovese CR, et al. Dorsal cortical regions subserving visually guided saccades in humans: an fMRI study. Cereb Cortex. Oxford University Press; 1998; 8: 40–47. https://doi.org/10.1093/cercor/8.1.40 PMID: 9510384

63. Petit L, Haxby J V. Functional Anatomy of Pursuit Eye Movements in Humans as Revealed by fMRI. J Neurophysiol. American Physiological SocietyBethesda, MD; 1999; 82: 463–471. https://doi.org/10.1152/jn.1999.82.1.463 PMID: 10400972

64. Lepsien J, Pollmann S. Covert reorienting and inhibition of return: an event-related fMRI study. J Cogn Neurosci. 2002; 14: 127–144. https://doi.org/10.1162/089892902317236795 PMID: 11970781

65. Armstrong KM, Chang MH, Moore T. Selection and maintenance of spatial information by frontal eye field neurons. J Neurosci. 2009; 29: 15621–9. https://doi.org/10.1523/JNEUROSCI.4465-09.2009 PMID: 20016076

66. Schall JD. On the role of frontal eye field in guiding attention and saccades. Vision Res. 2004; 44: 1453–1467. https://doi.org/10.1016/j.visres.2003.10.025 PMID: 15066404

67. Godoy J, Lüders H, Dinner DS, Morris HH, Wyllie E. Versive eye movements elicited by cortical stimulation of the human brain. Neurology. American Academy of Neurology; 1990; 40: 296–9. https://doi.org/10.1212/WNL.40.2.296 PMID: 2300252

68. Binder JR, Desai RH, Graves WW, Conant LL. FEATURE ARTICLE Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cereb Cortex December. 2009; 19: 2767–2796. https://doi.org/10.1093/cercor/bhp055 PMID: 19329570

69. Ye Z, Zhou X. Conflict control during sentence comprehension: fMRI evidence. Neuroimage. 2009; 48: 280–290. https://doi.org/10.1016/j.neuroimage.2009.06.032 PMID: 19540923

70. Nee DE, Wager TD, Jonides J. Interference resolution: insights from a meta-analysis of neuroimaging tasks. Cogn Affect Behav Neurosci. 2007; 7: 1–17. https://doi.org/10.3758/CABN.7.1.1 PMID: 17598730

71. Visser M, Jefferies E, Lambon Ralph M a. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. J Cogn Neurosci. 2010; 22: 1083–94. https://doi.org/10.1162/jocn.2009.21309 PMID: 19583477

72. Noonan KA, Jefferies E, Corbett F, Lambon Ralph MA. Elucidating the nature of deregulated semantic cognition in semantic aphasia: evidence for the roles of prefrontal and temporo-parietal cortices. J Cogn Neurosci. 2010; 22: 1597–613. https://doi.org/10.1162/jocn.2009.21289 PMID: 19580383

73. Jefferies E, Lambon Ralph MA. Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. Brain. Oxford University Press; 2006; 129: 2132–2147. https://doi.org/10.1093/brain/awl153 PMID: 16815878

74. Price C. The anatomy of language: a review of 100 fMRI studies published in 2009. Ann N Y Acad Sci. Blackwell Publishing Inc; 2010; 1191: 62–88. https://doi.org/10.1111/j.1749-6632.2010.05444.x PMID: 20392276

75. Price C, Mechelli A. Reading and reading disturbance [Internet]. Current Opinion in Neurobiology. 2005. pp. 231–238. https://doi.org/10.1016/j.conb.2005.03.003 PMID: 15831408

76. Rayner K, Ashby J, Pollatsek A, Reichle ED. The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. J Exp Psychol Hum Percept Perform. 2004; 30: 720–732. https://doi.org/10.1037/0096-1523.30.4.720 PMID: 15301620

77. Mahajan Y, Mcarthur G. Maturation of visual evoked potentials across adolescence. Brain Dev. 2012; 34: 655–666. https://doi.org/10.1016/j.braindev.2011.10.009 PMID: 22105091

78. Hauk O, Pulvermüller F. Effects of word length and frequency on the human event-related potential. Clin Neurophysiol. 2004; 115: 1090–1103. https://doi.org/10.1016/j.clinph.2003.12.020 PMID: 15066535

79. Sereno SC, Rayner K, Posner MI. Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. Cogn Neurosci. 1998; 9: 2195–2200. https://doi.org/10.1097/00001756-199807130-00009

80. Ehinger B V., Konig P, Ossandon JP. Predictions of Visual Content across Eye Movements and Their Modulation by Inferred Information. J Neurosci. 2015; 35: 7403–7413. https://doi.org/10.1523/JNEUROSCI.5114-14.2015 PMID: 25972169

81. Paczynski M, Kuperberg GR. Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. J Mem Lang. 2012; 67: 426–448. https://doi.org/10.1016/j.jml.2012.07.003 PMID: 23284226

82. Woldorff MG. Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. Psychophysiology. 1993; 30: 98–119. https://doi.org/10.1111/j.1469-8986.1993.tb03209.x PMID: 8416067

83. Kristensen E, Rivet B, Guérin-Dugué A. Estimation of overlapped Eye Fixation Related Potentials: The General Linear Model, a more flexible framework than the ADJAR algorithm. J Eye Mov Res. 2017; 10: 1–27. https://doi.org/10.16910/JEMR.10.1.7

84. Ehinger B V, Dimigen O. Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. bioRxiv. Cold Spring Harbor Laboratory; 2018; 360156. https://doi.org/10.1101/360156

85. Stowe LA, Kaan E, Sabourin L, Taylor RC. The sentence wrap-up dogma. Cognition. Elsevier; 2018; 176: 232–247. https://doi.org/10.1016/j.cognition.2018.03.011 PMID: 29609098

IV

INFLUENCE OF SPATIAL LOCATION OF AND TYPE OF DEVIATION FROM REAL WORDS ON SMALL SACCADES AND BRAIN ACTIVITY—A CO-REGISTERED EYE-TRACKING AND EEG STUDY

by

Otto Loberg, Jarkko Hautala, Jarmo A. Hämäläinen & Paavo H.T Leppänen,
2019

Submitted manuscript

Request a copy from the author.