Literacy Studies

Perspectives from Cognitive Neurosciences, Linguistics, Psychology and Education

Volume 23

Ronit Levie • Amalia Bar-On
Orit Ashkenazi • Elitzur Dattner • Gilad Brandes
Editors

Developing Language and Literacy

Studies in Honor of Dorit Diskin Ravid

Springer

Editors
Ronit Levie
Department of Communication Disorders
Tel Aviv University
Tel Aviv, Israel

Amalia Bar-On
Department of Communication Disorders
Tel Aviv University
Tel Aviv, Israel

Orit Ashkenazi
Department of Communication Disorders
Hadassah Academic College
Jerusalem, Israel

Elitzur Dattner
Department of Communication Disorders
Tel Aviv University
Tel Aviv, Israel

Gilad Brandes
School of Education
Tel Aviv University
Tel Aviv, Israel

ISSN 2214-000X ISSN 2214-0018 (electronic)
Literacy Studies
ISBN 978-3-030-99890-5 ISBN 978-3-030-99891-2 (eBook)
https://doi.org/10.1007/978-3-030-99891-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Part III
Linguistic Literacy

Finger Movements and Eye Movements During Adults' Silent and Oral Reading

Davide Crepaldi, Marcello Ferro, Claudia Marzi, Andrea Nadalini, Vito Pirrelli, and Loukia Taxitari

Abstract Using a common tablet and a web application, we can record the finger movements of a reader that is concurrently reading and finger-pointing a text displayed on the tablet touchscreen. In a preliminary analysis of "finger-tracking" data of early-graders we showed that finger movements can replicate established reading effects observed in more controlled settings. Here, we analyse and discuss reading evidence collected by (i) tracking the finger movements of adults reading a short essay displayed on a tablet touchscreen, and (ii) tracking the eye movements of adults reading a comparable text displayed on the screen of a computer. Texts in the two conditions were controlled for linguistic complexity and page layout. In addition, we tested adults' comprehension in both silent and oral reading, by asking them multiple-choice questions after reading each text. We show and discuss the reading evidence that the two (optical and tactile) protocols provide, and to what extent they show comparable effects. We conclude with some remarks on the importance of ecology and portability of protocols for large-scale collection of naturalistic reading data.

Keywords Reading · Eye-tracking · Finger-tracking · Data collection · Task ecology · Oculomotor coordination · Natural reading behaviour · Chunking

D. Crepaldi
Cognitive Neuroscience, SISSA Trieste Italy, Trieste, Italy
e-mail: davide.crepaldi@sissa.it

M. Ferro · C. Marzi · A. Nadalini · V. Pirrelli (✉) · L. Taxitari
Institute for Computational Linguistics, CNR Pisa Italy, Pisa, Italy
e-mail: marcello.ferro@ilc.cnr.it; claudia.marzi@ilc.cnr.it; andrea.nadalini@ilc.cnr.it; vito.pirrelli@ilc.cnr.it; loukia.taxitari@ilc.cnr.it

R. Levie et al. (eds.), *Developing Language and Literacy*, Literacy Studies 23,
https://doi.org/10.1007/978-3-030-99891-2_17

1 Introduction

Reading evidence can be collected and analysed in a number of ways, depending on the specific interests of the investigators and the range of theoretical and practical issues they intend to address. The technological advances of the last three decades have provided better and more sophisticated methods of reading research, which have greatly improved data collection and analysis, while contributing to broader and more detailed experimental and educational models of reading.

In the cognitive literature, a wide variety of experimental tasks that involve reading assessment have been proposed and tested in research labs. Among these tasks, eye-tracking and self-paced reading have probably established themselves as the most widely-accepted methods. In turn, experimental methods have been assessed for their ability to give detailed indications about the main perceptual, attentional, oculomotor, cognitive and linguistic processes that make reading possible, and their incremental integration in online text processing.

In education, protocols for reading assessment have focused on the basic outcomes of reading, and what is required for them to optimally interact in real practice, i.e. given specific instructional tasks and objectives. This perspective is more functional than explanatory, geared towards understanding how knowledge of basic reading skills can be used to help students learn more effectively through instructional readings. For example, based on the "Simple View of Reading" (Hoover & Gough, 1990), the two tasks of text decoding and comprehension are assessed to understand how they can be scored either independently or jointly, to evaluate a reader's proficiency and recommend dedicated training for reading improvement.

Cognitive and educational approaches to reading assessment strike us as highly complementary. It is only to be expected that a better understanding of the cognitive processes underlying reading would lead to more effective research-based intervention approaches to maximize reading performance. Conversely, a large screening of the school population for reading assessment at scale would deliver massive naturalistic data to reading researchers for quantitative analysis and modelling. Larger-scale studies could thus be conducted, paving the way to more generalizable results than in the past.

In spite of their huge potential for synergy, however, much remains to be done for the two perspectives to be integrated into a common observational framework. One of the main reasons why we are still far from achieving such a level of interdisciplinary continuity is the lack of an appropriate technological infrastructure for data collection and harvesting (Jamshidifarsani et al., 2019). Eye-tracking technology is constantly improving in sophistication and adaptivity (Jarodzka et al., 2021). Yet, measuring eye movements during natural reading in fairly unconstrained settings like schooling activities in the classroom remains a challenging task. Experimental research requires careful selection of input stimuli, which need be classified along a number of linguistic dimensions (involving orthographic, phonological, morphological, lexical, syntactic and pragmatic knowledge) and cognitive parameters (including working memory and executive functions) to investigate readers' language skills either in isolation or in interaction.[1] However, controlling for all these parameters in protocols for language education and intervention can be very hard, especially when students' reading performance is assessed on real instructional texts, as opposed to words or sentences presented in isolation. In sum, boosting synergy between experimental and educational reading research requires timely delivery of behavioural data that are fine-grained, robust and scalable: a long-awaited desideratum (Chzhen et al., 2018). In this paper, we describe a new method for collecting reading data with a common tablet connected to a server equipped with Artificial Intelligence and Natural Language Processing technology. The method is based on *ReadLet* (Ferro et al., 2018b), a web application that uses a tablet touchscreen to capture the reading behaviour of a subject by recording the speed of her finger pointing to a text as she reads it. We report the results of an experiment where "finger-tracking" data are analysed against eye-tracking data in adults' reading trials. Our method proves to be able to offer reading data that are robust, scalable and remarkably congruent with more established evidence. Structural and dynamic aspects that are specific of the finger-tracking evidence are also discussed.

2 Oculomotor Coordination in Visual Decoding

Recent experimental evidence in visual perception analysis shows that eye movements and finger movements are strongly congruent when a subject is asked to visually explore an image. Lio et al. (2019) recorded the eye movements of subjects while they are viewing an image displayed on a computer screen. In a separate experiment, the authors then invited the same subjects to explore a (different) blurred image, displayed on a touchscreen, by moving their fingers on the display. Blurred picture areas were intended to simulate parafoveal vision. However, the same areas were automatically shown in high resolution (corresponding to the subject's foveal vision), as soon as the subject touched a point on the screen located immediately below the blurred area. No single subject explored the same image in the two modes, but the same images were explored by different subjects either optically or haptically. The experiment proved that the subjects' image-exploring patterns in the two modalities strongly correlate at both individual and group levels. Synchronized recordings of eye movements and hand motor activities are reported from other domains such as piano playing (Furneaux & Land, 1999; Truitt

[1] Laubrock and Kliegl (2015) estimate that there are more than 50 word properties that could account for variance in fixation duration during reading (ranging from lexical neighbourhood frequency, bigram and trigram frequency and orthographic-phonological consistency to lexical frequency, length, subjective familiarity, concreteness and age of acquisition). Similarly, there are a number of sentence level variables that relate to eye movement control, going from subject-initial and object-initial constructions, to main or subordinate clauses, passive or active clauses etc. (see also Balota et al. (2004)).

et al., 1997), handwriting (Alamargot et al., 2007) and typewriting (Butsch, 1932; Inhoff et al., 1986; Inhoff & Wang, 1992; Inhoff & Gordon, 1997). Although these data are fairly heterogeneous and only indirectly related to reading, they converge into highlighting the basic need to coordinate fast eye movements and the much slower motor system of the hands. In particular, the average time the hand lags behind the eyes, or Eye-Hand Span, is fairly constant at around one second, if measured in units of time, but increases with expertise if measured in information units (e.g. letters or musical notes). This is in line with models of working memory that measure the working memory capacity in terms of the execution time of effector-driven processes (i.e. the individual articulatory rate of a speaker in Baddeley's phonological loop (Baddeley, 2007)) rather than processing units.

This evidence is in line with Laubrock & Kliegl's (2015) dynamic investigation of reading dual response costs, when the reader is engaged in simultaneously executing oculomotor and articulatory movements, either overt (in oral reading) or covert ones (in silent reading). During reading, eye movements provide sequential information to the short-term orthographic input buffer, where this information decays very quickly. Buffering is necessary because articulation (either covert or overt) is just too slow to keep the pace of both visual decoding and grapheme-to-phoneme conversion. However, due to short-term memory decay and the buffer's limited capacity, orthographic information cannot be retained indefinitely. Since the voice proceeds at a fairly linear pace, most of the adjustment has to be performed by the oculomotor system, which can either reduce its pace to stop the span between visual decoding and articulation (or Eye-Voice-Span) from growing out of control, or can retrace its steps backwards with a regressive saccade, to refresh items in the orthographic short-term buffer. Incidentally, similar buffering mechanisms (involving morphological processing at the word level) have been shown to be operational also in other types of dual task, such as handwriting (Kandel et al., 2008) and typing (Ferro et al., 2016; Gagné & Spalding, 2016).

Another familiar context which exploits the synergistic behaviour of the ocular system and another slower motor system, is when children learn to read using the finger of their dominant hand to point the letters of written words as they read a connected text. Despite the undoubtedly different dynamics of the two types of text exploration, finger-pointing a word during reading helps children learn to look at print, and supports critical early reading behaviours: directional movement, attention focus, and voice-print match (Mesmer & Lake, 2010; Uhry, 2002). In previous work (Marzi et al., 2020; Taxitari et al., 2021), we presented a preliminary analysis of finger-tracking data of early-graders that were engaged in concurrently reading and finger-pointing a short text displayed on a tablet touchscreen. The evidence showed that finger-tracking can replicate established eye-tracking benchmark effects, such as stable correlation values between tracking time and word frequency (negative correlation), and tracking time and word length (positive correlation). Besides, a comparative analysis of typically and atypically developing children (Marzi et al., 2020) shows different development patterns, with typical readers developmentally becoming less sensitive to word frequency effects, unlike atypical readers, who appear to remain sensitive to lexical frequency for much longer. The effect, also observed by Zoccolotti et al. (2009), is interpreted as showing that the orthographic lexicon of typically developing readers makes room for increasingly rarer and longer words with age. Conversely, atypical readers do not seem to be able to update their orthographic lexicon with rarer and longer words at the same pace as typically developing readers do.

In the present contribution, we examine the correlation of finger-tracking and eye-tracking times during adults' silent and oral reading. Our objective is to show to what extent reading movement patterns and their speed are congruent across the optical and tactile protocols in the two (silent and oral) modalities. In what follows, we will first recall some known features of different reading protocols to discuss their connection with natural reading and finger-tracking. After that, the methodology of a comparative finger-tracking and eye-tracking experiment with adult readers is described in some detail. Results are illustrated and discussed in the ensuing section and, finally, some methodological remarks are reported in the conclusions. Overall, the evidence shows that finger-tracking offers a novel, minimally invasive, inexpensive and nonetheless highly informative way to assess and understand natural reading across different proficiency and age levels. In the end, the protocol's portability and robustness have the potential to bridge the current gap between cognitive and educational research on reading.

3 Natural Reading and Reading Tasks

In a typical eye-tracking experiment, entire sentences are presented one at a time, and participants are asked to read each sentence, either silently or aloud, at their normal reading speed. During the task, the location and duration of their eye fixations are recorded, which makes it possible to assess participants' reading patterns as they evolve during the process of sentence reading. As participants are given very few or no restrictions on how reading should take place, the task allows for a number of reading strategies to be observed. Reading can take place very carefully, with readers making their way through the sentences at a regular pace, virtually word by word. Alternatively, they can decide to skim through the sentences to get the gist of their content, which may be enough for them to correctly answer a few simple multiple-choice comprehension questions.

At the beginning of a typical self-paced reading session, a sentence is displayed on a computer screen as a series of dashes or hash marks, each covering a single character in the sentence. Upon a button press by the experiment participant, the first word of the sentence is shown on the screen. When the participant is ready to view the next word, (s)he presses the button again, thus reverting the current word to dummy marks, and unmasking the immediately ensuing word in the sentence. The participant proceeds in this way until the last word of the sentence is shown. In this case, the only dependent variable is the time the reader takes to push the button in recognition of the word currently displayed. The reading task is certainly less natural than in the eye-tracking protocol. In particular, it reduces the number of possible reading strategies. As only one word at a time is displayed, self-paced reading forces explicit fixation of those (mainly functional) words that are often skipped in natural reading, and it does not permit regressive eye movements. If, on the one hand, this protocol allows for a much narrower interindividual variability in terms of processing strategies, a variety of reading strategies are nonetheless available (Witzel et al., 2012), as the reader may decide to integrate each unmasked word online, or rather buffer it in working memory and postpone sentence integration to a later stage.

More recently, new word-by-word reading techniques (like the maze task) have been proposed as a way to tap into the reader's processing strategies by placing even stricter limits on their freedom. For example, at each button press, two words rather than a single one are displayed, which provide two alternative continuations of the sentence. The reader has to decide on the most sensible alternative (see Gallant and Libben (2020) for a recent adaptation of the task).

As nicely put in a recent contribution by Libben et al. (2021) "[...] a key challenge in the design of psycholinguistic research on lexical processing is to create experiments that have ecological validity and at the same time are sufficiently controlled so that specific variables and hypotheses regarding their effects can be examined." Accordingly, classical reading tasks can mostly be evaluated according to two dominant parameters: (i) whether the task allows investigators to collect evidence of online processing ease/difficulty for the reader, and (ii) whether the reading task is modelled in a "natural" way.

The first parameter is of paramount importance, as reading patterns are a primary source of information of text processing and comprehension operations. Ideally, a reading protocol should provide detailed information of this kind through indication of reading time differences across texts and readers. From this perspective, it is important that the protocol can precisely show where the processing difficulty arises in the text. This explains why self-paced reading and its variants still enjoy considerable popularity among investigators. Due to its preeminent focus on one single parameter at a time, self-paced reading places tight constraints on subjective reading strategies, to provide fairly local, consistent and comparable data points.

However, on our second parameter, self-paced reading fares much worse. There is little that is natural in placing rigid restrictions on the subject's ability to "look ahead" in the reading direction, while processing a word in context. Although context manipulation can reveal a lot of the reader's processing strategies, at least some self-paced data can reflect strategic choices that respond to specific aspects of the task (e.g. pending phrase structures can be "closed" prematurely, to provide fast integration of the current word in the preceding context), rather than reflecting natural reading behaviour.

Eye-tracking seems to suffer something of a mirror-image problem. On the one hand, it places very few if any restrictions on the subject's reading strategy, thereby capturing a natural reading behaviour. On the other hand, it provides a wealth of behavioural patterns and measures (forward saccades, regressions, fixations, refixations and word skippings) that portray the reader's ability to process several words in parallel in its full complexity and interindividual variability. In some cases, however, this makes it difficult to control for specific context-sensitive behavioural patterns, as with the controversial case of parafoveal-on-foveal effects (Brothers et al., 2017), whose investigation may require considerable online manipulation of the textual context (for example, Rayner's (1975) boundary technique: Angele et al. (2013), Dare and Shillcock (2013)).

A finger-tracking experiment consists in recording the movements of the dominant hand's index finger of a subject reading a text displayed on a tablet touchscreen (Ferro et al., 2018a). In the task, the reader is instructed to point each word in the text as she reads it, as is common practice of beginning readers. During reading, the tablet can record the sliding movements of the finger captured by the tablet touchscreen, as well as the voice of the reader (when reading aloud is requested) captured by a built-in microphone. Both acoustic and haptic recordings are continuous in time, while recorded finger movements are also continuous in space: i.e. they tend to cover text letters, punctuation marks and even blanks evenly, with a limited number of orthographic units being skipped. This dynamic is in sharp contrast with the succession of discontinuous individual movements that typically characterize the eyes during reading, which is better described as a series of more or less long fixations that are traversed "in jumps", i.e. by alternating fixated with nonfixated words. Notwithstanding these dynamic differences, both eye and finger movements can be aligned with the time of a reading session and the line(s) and words of the text being read. In particular, the touch screen technology of a current good-quality commercial tablet is able to capture finger movements with a sampling rate in the 60–120 Hz range, approximately corresponding to 12–24 touch events per syllable when reading at a speed of 5 syllables per second. The tablet can thus map a continuous sliding movement on the touchscreen into a discrete series of densely distributed "touch events", each located in the screen area. At any given point in time, an algorithm can thus precisely reconstruct where on the screen the reader's finger is pointing to. This series of discretized events are ultimately mapped onto the text lines, in much the same way a sequence of fixations is projected onto a sequence of words. This allows researchers to observe which letter is pointed to by the reader's finger at any moment during reading.

To illustrate, Fig. 1 (top) shows the visual rendering of a typical eye-tracking record of a short text paragraph after alignment. It is useful to compare the figure with the corresponding record of finger-tracking data for the same text (Fig. 1, bottom). In both renderings, the tracking time is represented through a horizontal bar below each paragraph line. The bar's false-colours code time in milliseconds, with the coding scale being depicted along the vertical bar on the right-hand side of the text. Note that finger-tracking covers text lines in a spatially denser and more continuous way than eye-tracking does. The finger tends to slow down at the end of each line, and moves to the beginning of an ensuing line in one jump, to resume tracking the new line. Note that words, and even single letters within words, are tracked at different speeds. In contrast, the eye typically jumps from one word to another, fixating single words between successive jumps. A word can be fixated more or less quickly, and more than once, whereas some words are completely skipped.

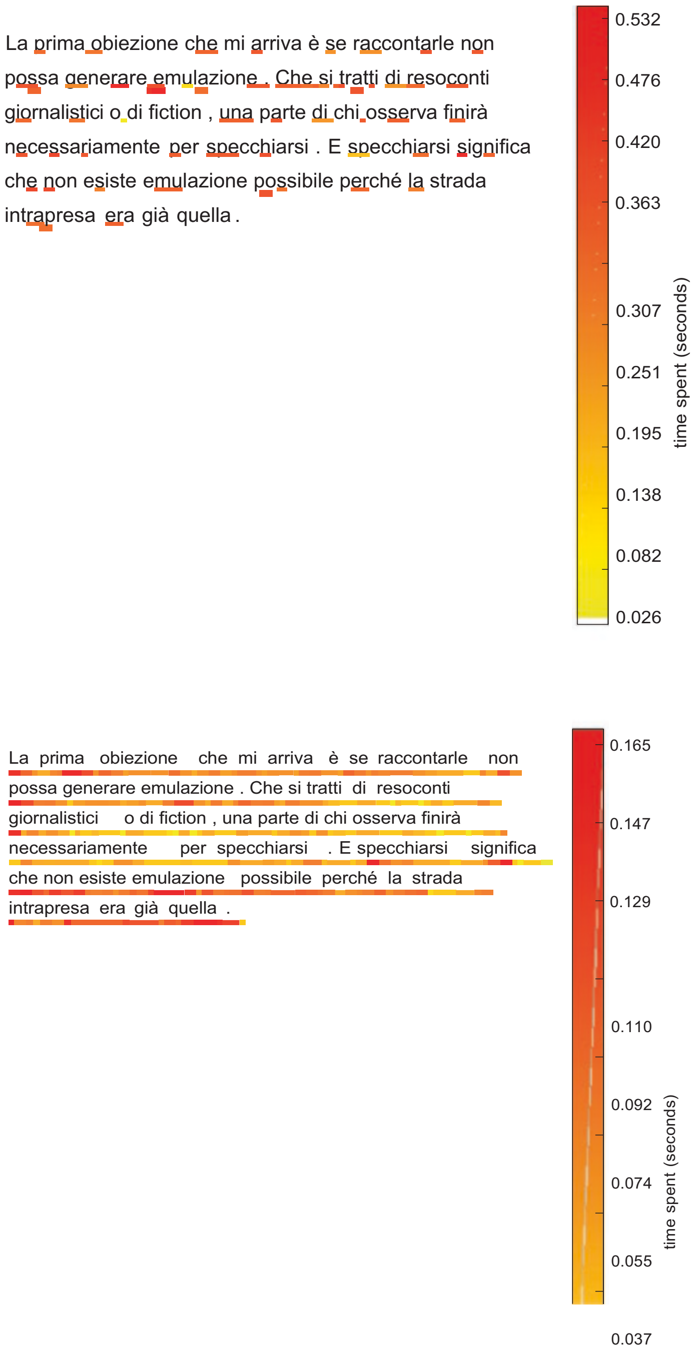


Fig. 1 Visual rendering of an eye-tracking record (top) and a finger-tracking record (bottom) of the first page of one of the Roberto Saviano's journal articles used for the adults' reading trials

Although the two tracking records are differently scaled and differently distributed across the text, it makes sense to align them at the word level for comparison. As a first approximation, the time taken to read the word w_i in the text can be calculated as its total fixation time in the eye-tracking record and its total tracking time in the finger-tracking record. This inevitably neglects important pieces of information that are provided by the two signals (for example, both of them keep track of regressions). Likewise, we are levelling out significant differences between the two protocols: for example, finger-tracking, unlike eye-tracking, allows researchers to examine differences in reading pace at various points within the word. Nonetheless, a word-level comparison provides a useful starting point to understand more about how the two time series of signals correlate in the natural reading of a connected text.

In what follows, we will thus focus on some descriptive statistics of the reading time data offered by the two protocols, and will compare these data across different hierarchical levels of linguistic units: starting from words, to include non-recursive phrase constituents (or "chunks") and full sentences. Ultimately, this preliminary investigation will also allow us to validate finger-tracking against a challenging, well-established benchmark technology in reading research such as eye-tracking, and better understand the similarities and differences between the two experimental protocols in adults' reading.

4 The Experiment

The study was approved by the CNR Research Ethics and Integrity Committee and funded by the Ministry of University and Research through the PRIN grant 2017W8HFRX.

4.1 Participants

Fifty-Six young adults (27 female, 29 male, mean age = 27, age range = 18–39) were recruited for the experiment. All participants were Italian native speakers, with normal or corrected-to-normal vision, and without any known learning or reading difficulties. Two participants were left-handed. Participants gave written informed consent for their involvement. 22 experimental sessions were conducted at the CNR Research Area in Pisa, and 34 experimental sessions in SISSA, Trieste.

4.2 Design

The current study adopted a 2 (tracking protocol: eye-tracking vs. finger-tracking) by 2 (reading condition: silent vs. aloud) Latin square, fully counterbalanced design. Accordingly, each participant was asked to read in four experimental conditions, combining reading mode and tracking protocol. This resulted in a total of four reading sessions per participant, all conducted the same day. Order of presentation of the tracking protocol, i.e. tablet vs. eye-tracker, was counterbalanced across participants. Reading conditions, i.e. silent reading vs. reading aloud, alternated for each participant to avoid fatigue effects, and the order was counterbalanced across participants. The presentation of the different reading texts for each experimental condition was also counterbalanced across participants, so that text passages were equally distributed across experimental conditions.

4.3 *Materials*

Test reading materials consisted of 8 short Italian texts, each taking up two tablet screen pages. They were extracted from either Roberto Saviano's tabloid news articles, or Lamberto Maffei's popular neuroscience book *Elogio della parola* ('In praise of words') (2018). All texts were automatically PoS-tagged and shallow-parsed in word chunks. For PoS-tagging, we used the coarse-grained level of the ISST-TANL morpho-syntactic tagset (Dell'Orletta et al., 2007), output in CoNLL format. Chunking (Abney, 1991; Federici et al., 1996) defines a non-recursive level of phrasal text segmentation, where lexical heads are always the rightmost word token in the chunk, as illustrated in Example (1). With a few exceptions, functional words are mostly incorporated as pre-head word units within a lexical chunk, and this provides linguistically principled ways to explore the correlation between acoustic, prosodic and syntactic cues in reading a connected text (Pate & Goldwater, 2011).

Example 1
[...] [with his legendary **sword**]$_{P\_C}$ [the **king**]$_{N\_C}$ [has been **ousting**] $_{FV\_C}$ [his **enemies**]$_{N\_C}$ [from the **realm**]$_{P\_C}$

At each reading session (i.e. for each combination of reading mode and tracking type), the subject was required to read two texts: one by Saviano, and the other by Maffei. A single multiple-choice question was asked soon after the reading of each text. On average, the total amount of text being read at each session comprises 557.5 words (range: 524–586) and 24.5 sentences (range: 23–30). Sentences are, on average, 22.75 word tokens long (range: 1–90). A summary of the lexical and linguistic features of our reading texts is presented in Table 1.

4.4 *Procedure*

Overall, the entire experiment, consisting of four reading sessions, lasted around 30 min per participant. At each reading session, participants were asked to read two texts of two screen pages each, either silently or aloud, and their session was either

Table 1 Distributional features of lexical and phrasal stimuli

	Type	#	Mean	Range
POS types		12		
Word tokens		2230		
Word types		1024		
Word token length (by letters)			5.14	1–16
Chunk types		14		
Sentence length (by tokens)			22.75	1–90
Session text length (by tokens)			557.50	524–586

eye-tracked or finger-tracked. During the reading session, participants were instructed to put on a pair of wireless noise-cancelling headphones with a retractable microphone (BlueParrott S450-XT for the tablet sessions, Razer Nari Essentials Gaming Headset for the eye-tracking ones).

4.4.1 Eye-Tracking Protocol

Participants sat in front of a desk at about 60 cm from a 24″ DELL computer screen. An eye-tracker was placed below the screen, on the same desk. Participants used a five-button response box to proceed from one page to the next one, pressing a central green button. The same toolbox was used to answer the questions, by pressing one of the remaining buttons, numbered from 1 to 4. Eye movements were recorded via an Eyelink Portable Duo eyetracker (SR Research, Canada), which supports head-free eye-tracking with a reported accuracy of 0.25° to 0.50°. Only the right eye of each participant was tracked at a 500 Hz sampling rate. Texts were visualized in Arial (25pt), in black against a white background. Stimulus presentation and eye movements recording were handled with Matlab Psychtoolbox (Brainard, 1997; Kleiner et al., 2007). Compared to the tablet protocol, here the font size was increased to adjust for the larger participant-screen distance than the participant-tablet distance.

Participants were firstly instructed to read two passages on the computer screen while trying to keep as still as possible, without moving their head. Before the actual experiment started, a nine-point-calibration procedure was conducted until the average error was below 0.5° of visual angle. No chin-rest was used during the experiment in either reading modes. A practice reading excerpt from the Italian translation of a Harry Potter's novel was presented as training. After training and calibration, the actual texts were shown, with each passage followed by two multiple choice questions. Upon page changing, a drift correction was carried out to correct for small movements, and a stable fixation on a centrally located target was required before proceeding further.

4.4.2 Finger-Tracking Protocol

Participants sat in front of a desk on which there was a tablet placed on a tablet stand, resting on an anti-slip place mat to prevent accidental tablet displacement during finger-tracking. The tablets used for the study were 10.1 inches Samsung TAB A SM-T510N (1.8 GHz Octa-Core, 3 GB RAM, 64 GB eMMC, Android 10). The screen of the tablet was 14.9cm × 24.5cm with a resolution 1920 × 1200 pixels. The text was presented in Arial (21.25pt). For each text, the same word bounding box coordinates used by the tablet touchscreen were then used to define the text layout on the computer screen used in the eye-tracking sessions.

Before starting a real experimental session, participants were instructed to use the tip of the index finger of their dominant hand to point the text words displayed on the tablet while reading them. An excerpt from the Italian *Pinocchio* novel by Carlo Collodi was used for a brief practice session. Each participant read two passages, of two pages each, one silently and one aloud, one after the other. Each passage was followed by a single 4-choice question.

4.5 Data Processing and Measures

4.5.1 Eye-Tracking

Out of all 112 eye-tracked reading sessions, 24 were excluded from the analyses due to technical malfunctions during data acquisition or excessive noise artifacts in the fixation patterns (e.g., horizontal transposition; 19 sessions). This left us with an effective sample of 88 reading sessions (44 aloud, 44 silent) to be included in the analysis dataset.

A first automatic data trimming was conducted using an R script. Any individual fixation falling more than 60 pixels out of each text bounding box was excluded from the dataset (1.65% data loss). In addition, as the drift correction before each recording was done on a fixation point located in the center of the screen, early fixations falling below the Y-coordinate of the first line of the text were further excluded. Secondly, for each participant, a double visual inspection of fixation patterns was performed. Due to the fact that readers, prior to reading a text and after its completion, normally look at the text in an unpredictable manner, such fixations were manually excluded.

All these steps were taken to ensure optimal performance of the post-hoc drift correction algorithm, which was not designed to spot and exclude the aforementioned fixations. In particular, we used the “warp” algorithm developed by Carr et al. (2021), which was shown to achieve a very good performance. After the vertical drift correction, each fixation was assigned to the corresponding word if falling within its corresponding bounding box.

As a measure of word reading performance, Total Fixation Time (hereafter TFT) was used. It consists of the total time spent fixating a word, including regressions to the word, i.e. second-time fixations. For higher-order linguistic units such as chunks or sentences, the corresponding TFT was calculated as a summation of the TFTs of all words the higher-order unit spans over.

4.5.2 Finger-Tracking

For the tablet sessions, automatic text-to-finger alignment was enforced using a convolutional algorithm finding the closest match between text lines and touch event sequences. The bounding box of each character in the text was then used to calculate the finger-tracking time of the corresponding letter (or Letter Tracking Time, LTT). For each uninterrupted time series of touch events falling within a letter bounding box, LTT is then equal to the difference between the last time tick and the first time tick in the series of touch events.

As a measure of reading performance for any text unit (e.g. a word or a chunk), its Total Tracking Time (TTT) was then used. TTT consists of the total time needed to finger-track the text unit, including possible regressions to the unit (i.e. second-time tracking), calculated as a summation of the LTTs of all letters the text unit spans over.

5 Data Analysis

This section provides an analysis of reading data in both reading types (silent and oral) and for the two tracking protocols (finger-tracking and eye-tracking). As our main objective is to understand how finger-tracking data relate to eye-tracking data in text reading, the main focus will be on comparing total tracking times and total fixation times across the two reading modalities. Accordingly, we start with comparatively assessing how our eye-tracking data pattern across silent and oral reading; we then focus on finger-tracking data, to see how they differently behave in the two reading modes. Eye-tracking and finger-tracking data in each reading modality will then be meaningfully compared and discussed against this background.

5.1 Eye-Tracking

Eye movement patterns are known to differ between silent and oral reading both spatially (in terms of the number of fixated words and saccade amplitude) and temporally (in terms of fixation durations).

Our data are in line with the observation that parafoveal information significantly benefits silent reading more than oral reading in terms of a reduction in both fixation duration (measured by total word fixation times, see Fig. 2) and number of fixations (see Table 2). All in all, eye movement patterns in oral reading can be described

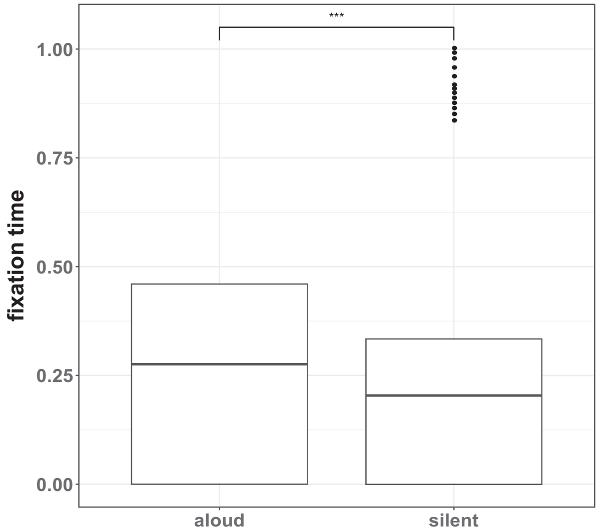


Fig. 2 Box plot distribution of word TFTs (in seconds) in adults' oral and silent reading

Table 2 Distribution of eye movement patterns across adults' oral and silent reading

	Oral		Silent	
	#	%	#	%
Single saccades (same line)	24,612		20,892	
Regressions	6365	26%	4786	23%
Fixations	25,608		22,363	
Backward fixations	9203	36%	6632	30%
Nonfixated words	5479	25%	6990	31%

as more "sequential" than those in silent reading: i.e. the former take smaller steps (i.e. shorter saccades, Fig. 3), make significantly longer fixations ($t(32116) = 31.17$, $p < .001$: Fig. 2), skip fewer words and, finally, make regressions more often (Table 2).

Such a difference is emphasized when we look at the way word TFTs are distributed across chunks of different length. Figure 4 shows a clear gap in the distribution of silent and oral word fixation times, which decreases with longer syntactic chunks: from one word chunks (chunk length = 1) to multiple word chunks (chunk length = 2, 3, 4).

The reason for such a decreasing gap becomes apparent when we look at the number of fixated words in chunks of different size, in oral (Table 3) and silent reading (Table 4). In silent reading, words are skipped more often than in oral reading. In fact, in silent reading, only 15% of 4-word chunks are fully fixated (i.e. no word in the chunk is skipped) against 24% in oral reading, and 6% of 4-word chunks are fixated on a single word only, against 1% of 4-word chunks in oral reading. Besides,

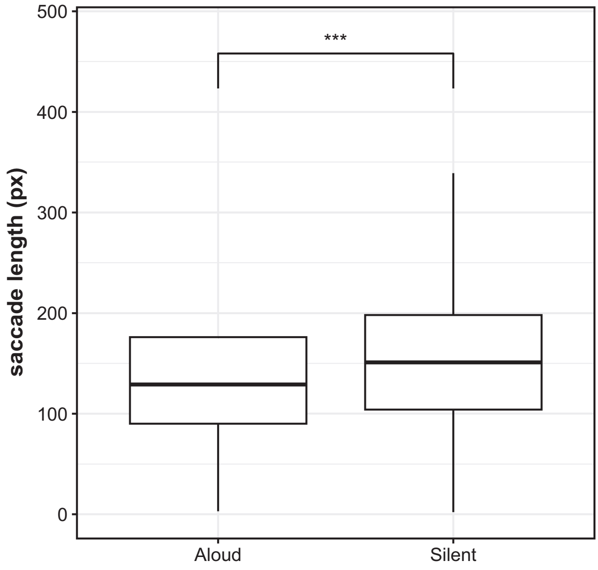
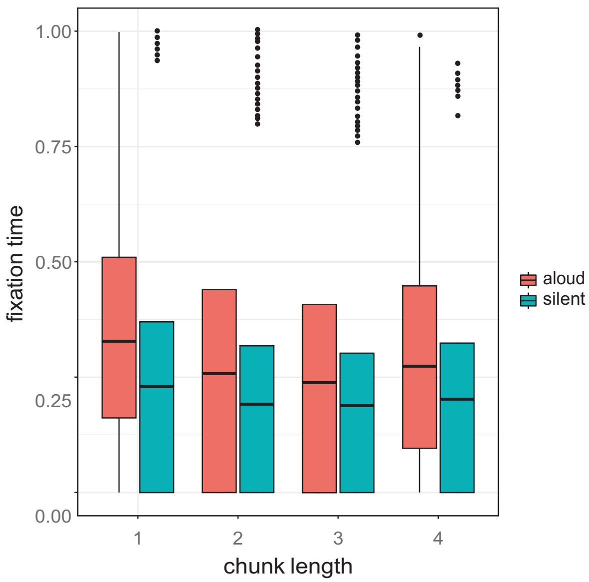


Fig. 3 Box plot distribution of saccade amplitudes (in pixels) for adults' oral and silent reading

Fig. 4 Box plots of word TFTs for chunks of increasing length in adults' oral (red boxes) and silent (cyan boxes) reading



in both oral and silent reading the percentage of fully-fixated chunks decreases with the length of the chunk (see Tables 3 and 4).

A non-linear regression model predicting word fixation time by word position in chunks, for chunks of different token length, sheds light on this pattern of data (Fig. 5). Word fixation time increases as subjects read more of a chunk, and

Table 3 Oral reading: distribution of nonfixated (#0) and fixated words in chunks of increasing length

	Ixated words in oral reading				
Chunk length in words	#0	#1	#2	#3	#4
1	1397 (19%)	5725 (81%)	0	0	0
2	131 (3%)	2584 (51%)	2348 (46%)	0	
3	17 (1%)	163 (12%)	730 (56%)	400 (31%)	
4	0	1 (1%)	33 (17%)	110 (58%)	45 (24%)

Table 4 Silent reading: distribution of nonfixated (#0) and fixated words in chunks of increasing length

	Ixated words in silent reading				
Chunk length in words	#0	#1	#2	#3	#4
1	1801 (25%)	5524 (75%)	0	0	0
2	195 (4%)	3183 (61%)	1883 (36%)	0	0
3	15 (1%)	294 (22%)	734 (54%)	315 (23%)	0
4	0	11 (6%)	57 (28%)	102 (51%)	30 (15%)

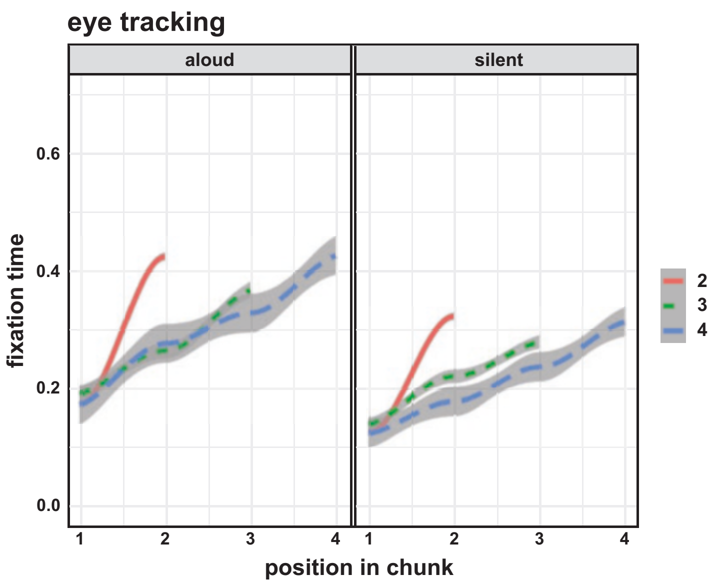


Fig. 5 Regression plots (*ggplot*) of interaction effects between chunk lengths in number of words and word position in the chunk for adults' oral (left) and silent reading (right)

culminates on the chunk's head, as shown by the ascending regression curves by chunk position. This appears to reflect the structure of a chunk. First, the chunk's head is the most prominent syntactic and semantic unit in the chunk, and plays a pivotal role in text processing and comprehension. Secondly, functional units in the chunk tend to take a more peripheral (chunk initial) position than lexical units, and are more likely to be skipped than words that are placed closer to the chunk's head. Incidentally, the processing impact of the chunk structure is similar in both silent and oral reading, as shown by the similar ascending lines in both regression plots.

5.2 Finger-Tracking

Word tokens are systematically and continuously finger-tracked in both silent and oral reading, with no significant difference in the (negligible) number of skipped (i.e. nonfixated) tokens (Table 5). Likewise, tokens are tracked more than once only occasionally in either mode, suggesting that adults' finger movements do not appear to follow eyes' regressive saccades (which are nonetheless frequent in both silent and oral adults' reading, as shown in Table 2), thus exhibiting a mostly one-way, forward-moving trajectory. Like with eye-tracking, silent reading significantly speeds up finger-tracking times ($t(59625) = 30$, $p < .001$: Fig. 6).

Plots of tracking time regressed on chunk length and chunk position (Fig. 7) highlight an interesting interaction between word tracking time and chunk length (measured by the number of word tokens within the chunk). In both oral and silent reading, we observe a clear effect of chunk structure on reading pace. The characteristically step-wise ascending patterns in the plots, with a similar, steepest increase in the tracking time of the head for chunks of different length, highlight, once more, the pivotal role of the head in chunk processing and the different processing loads associated with chunk-initial and intermediate units. This pattern is strongly reminiscent of what we observed for fixation data in the plots of Fig. 4.

We expect such a processing sensitivity to the internal structure of the chunk to have an effect on the average word tracking time for chunks of different length. In longer chunks more functional words are likely to be finger-tracked more quickly than the chunk's head. This expectation is confirmed by the box plots of Fig. 8, whose word tracking times in oral and silent across chunks of different length are strikingly similar to the patterns we observed for word fixation times in Fig. 4.

5.3 Finger-Tracking vs. Eye-Tracking

So far, we have observed patterns of reading behaviour across silent and oral reading, and have shown how they are recorded (fairly consistently) with eye-tracking and finger-tracking. Silent reading takes less time than oral reading, lexical heads increase processing demands and take longer to read, and longer chunks tend to speed up average word reading. In this section, we directly compare evidence in the two tracking protocols.

A cursory look at the density plots of both TTTs (total finger-tracking times) and TFTs (total fixation times) in adults' aloud reading (Fig. 9) shows, unsurprisingly, that most of the nonfixated words fall within the 1–4 letter range (as shown by the steep peak centred on $x = 0$, of the red density curve in the top left panel). Similarly,

Table 5 Distribution of finger movement patterns across adults' oral and silent reading

	Oral		Silent	
	#	%	#	%
Total tokens	31,059		30,037	
Tracked tokens	30,649	98.7%	29,637	98.7%
Nontracked tokens	410	1.3%	400	1.3%
Tracked more than once	529	1.7%	654	2.2%

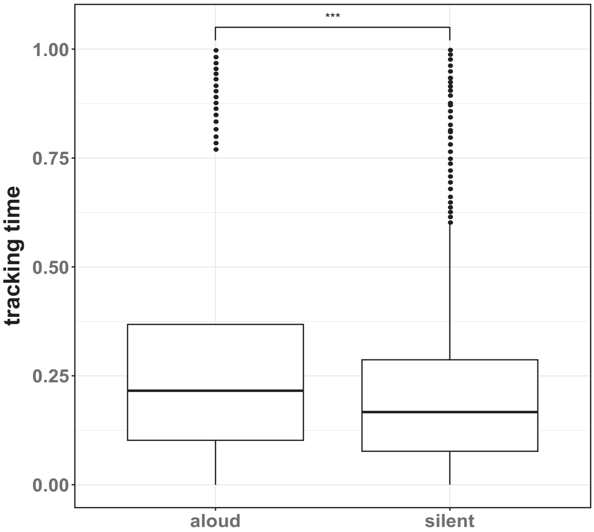


Fig. 6 Box plot distribution of finger-tracking times (in seconds) in adults' oral and silent reading of words

the few non-tracked words fall within the same range (light-blue density line). The effect of nonfixated tokens becomes negligible for longer words, where the two curves appear to exhibit a similar (unimodal) shape.

Table 6 shows correlation scores (Spearman *p*) between TTTs and TFTs in both aloud and silent reading of word tokens. Correlations are given for aggregate data, and by major lexical parts of speech. In each column, scores outside the parentheses are calculated after including all nonfixated tokens, while scores within parentheses are calculated after discarding nonfixated words. It is worth noting that scores between parentheses are comparatively smaller than the scores that are outside the parentheses. Since nonfixated words are mostly short, and finger-tracking times strongly correlate with word length (Spearman *p* .78, against .34 for fixation times), most nonfixated words are finger-tracked quickly, thus raising the correlation across word classes when they are included. As words are skipped more often in silent reading than in oral reading, the gap between correlation scores calculated after including nonfixated words and correlation scores calculated after discarding non-fixated words is larger for silent reading. For the same reason, function words

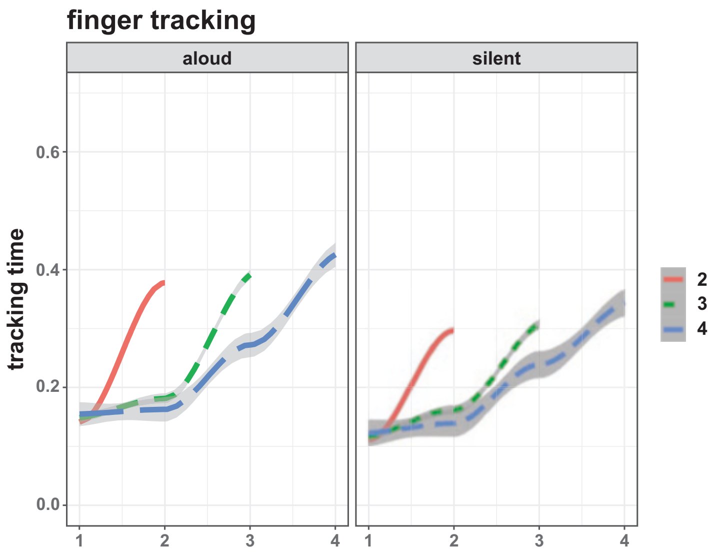


Fig. 7 Regression plots (*ggplot*) of interaction effects between chunk lengths in number of words and word position in the chunk for both adults' oral and silent reading

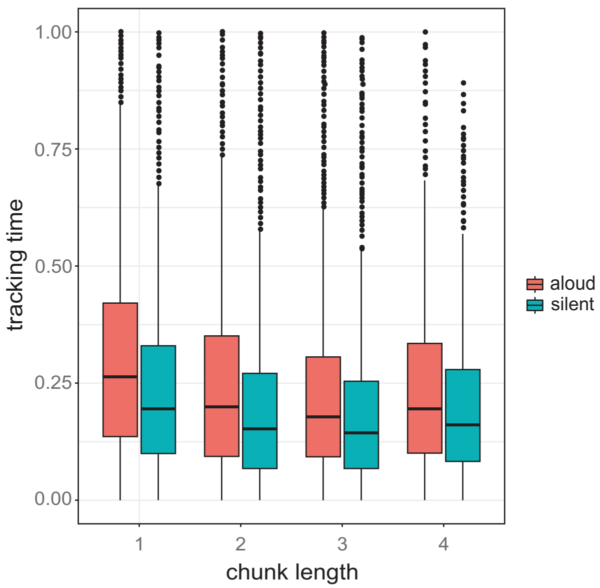


Fig. 8 Box plots of word TTTs for chunks of increasing length in adults' oral (red boxes) and silent (cyan boxes) reading

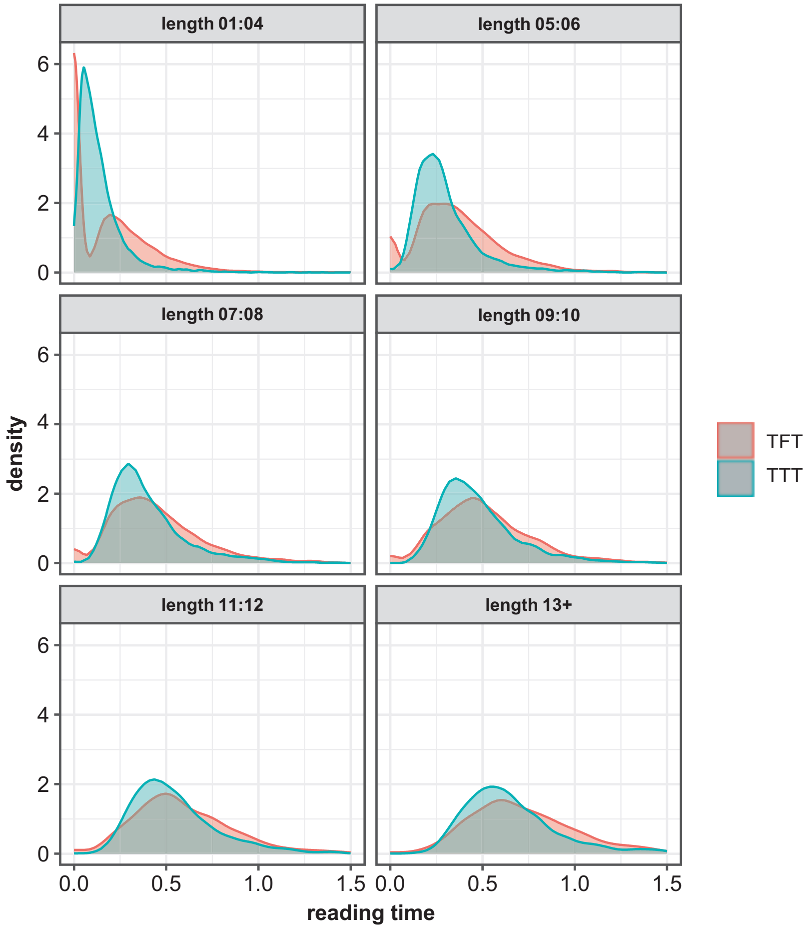


Fig. 9 Density plots of TFTs (red) vs. TTTs (light-blue) for tokens of increasing length in adults' aloud reading

present much larger gaps than lexical words (for silent reading, the *p* value is nearly halved without nonfixated words).

6 Discussion

Finger movements and eye movements reveal surprisingly consistent patterns of reading behaviour, with high correlations between total fixation times and total finger-tracking times at the level of individual words, chunks and sentences. Two major trends are worth reporting in this connection.

Table 6 Spearman *p*'s between TTTs and TFTs in aloud and silent reading of word tokens

Type	Aloud[a]	Silent[a]
ALL	.820 (.670)	.815 (.571)
LEXICAL TOKENS	.684 (.626)	.656 (.502)
FUNCTIONAL TOKENS	.709 (.505)	.716 (.365)
ADJECTIVES	.666 (.626)	.644 (.506)
ADVERBS	.723 (.649)	.701 (.484)
NOUNS	.577 (.550)	.550 (.469)
VERBS	.787 (.687)	.736 (.527)

[a]Scores in parentheses are calculated after discarding nonfixated words

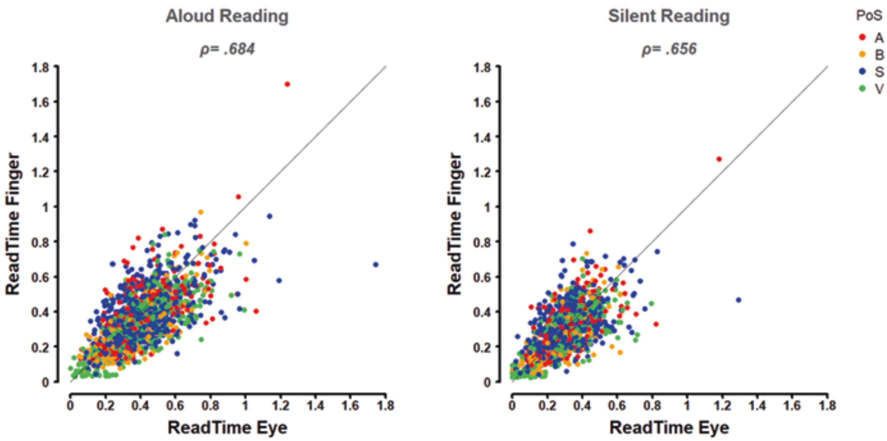


Fig. 10 Scatter plot of TTTs vs. TFTs in adults' aloud/silent reading of adjectives (A), adverbs (B), nouns (S) and verbs (V)

Firstly, Spearman *p* correlation values grow with embedding levels of linguistic units, ranging, in oral reading, from .617 on the word level (Fig. 10), to .787 on the chunk level (Fig. 11) and .99 on the sentential level (Fig. 12). This means that although the two signals are not perfectly synchronised with either letters or words, they nonetheless tend to be in step at the end of major linguistic units (i.e. chunks or sentences). Such an increase in time correlation can be an effect of noise reduction in time aligned data. In fact, the likelihood of finding occasionally misaligned data (e.g. a word that is wrongly assigned the fixation of an immediately preceding or an immediately ensuing word) goes down when text units are chunked together and the alignment window is widened. However, such a significant increase in correlation cannot be the sheer effect of noise reduction. In addition, we conjecture that finger movements, eye movements and sound articulation tend to keep in step at the end of major linguistic units such as chunks or sentences, where reading can naturally pause to give room for text integration and comprehension.Secondly, fixation and tracking times are more highly correlated in oral reading than in silent reading. We

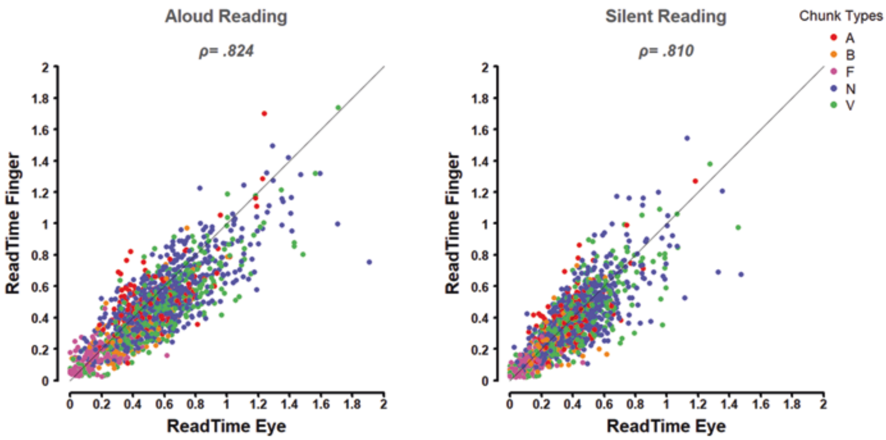


Fig. 11 Scatter plot of TTTs vs. TFTs in adults' aloud/silent reading of adjectival (A), adverbial (B), functional (F), nominal (N) and verbal (V) chunks

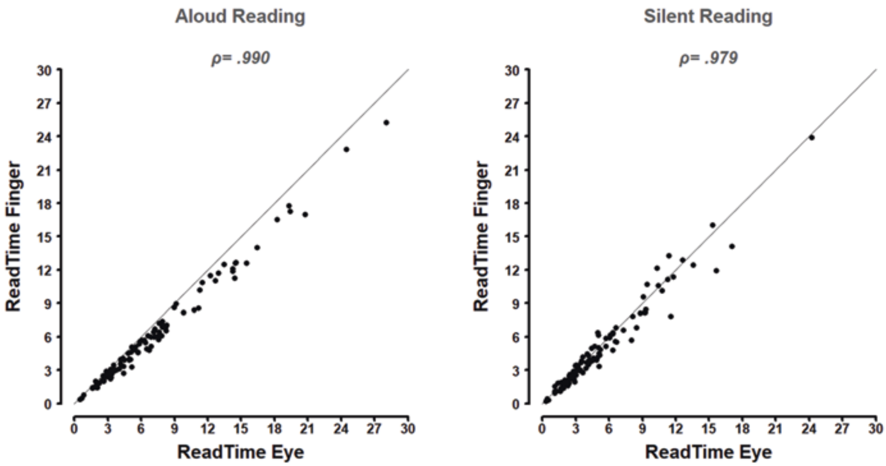


Fig. 12 Scatter plot of TTTs vs. TFTs in adults' aloud/silent reading of sentences

conjecture that this is due to the sequential and continuous dynamic of finger movements, which is more consistent with an oral reading strategy (more and longer fixations, shorter saccades) than a silent reading strategy (fewer and shorter fixations, longer saccades).

Overall, the evidence presented here suggests a nontrivial interaction between eye and finger movements. During reading, eye movements and finger movements are initiated independently and at different speeds, with a likely later start of the finger. The eyes have the main role of processing the text and initiating decoding and articulation (either covert or overt). The finger plays the subsidiary role of pace making and place holding at the end of major syntactic units. This function is more compatible with a more rapid and somewhat inertial series of forward movements.

Both TFTs and TTTs appear to be modulated and constrained by overt articulation (for oral reading) or covert articulation (for silent reading). This is proved by the influence of reading mode on both TFTs and TTTs. In the eye-tracking literature, there is considerable evidence of the influence on the Eye-Voice-Span (EVS) on adult readers' oculomotor planning. Laubrock and Kliegl (2015) offer a compelling working memory interpretation of this influence. To prevent words from exceeding the memory capacity of the orthographic input buffer and thus being skipped in reading, the oculomotor system reduces its pace to stop EVS from exceeding working memory capacity, or refreshes old items in working memory with regressive saccades. Since full articulation is slower than covert articulation, the slowing down effect of keeping EVS to a manageable amplitude is stronger in oral reading. As to the text-pointing finger, a slightly different way to achieve the same goal is pursued. According to our evidence, the finger does not appear to follow regressive eye saccades; it rather slows down its pace at natural linguistic boundaries, such as the end of a chunk or a sentence, where eye and finger movements are eventually kept in sync.

To check the impact of word length on word position in chunk (for chunks of different length) in the two reading modalities – oral and silent – and across the two experimental protocols – finger and eye-tracking – we ran four Generalized Additive Models (or *GAMs*) with word's position in chunk, word length and chunk length as independent variables predicting fixation and tracking times, including subjects as random effects. All models are fairly robust (see parametric coefficients for eye-tracking models in Tables 7 and 8, and for finger-tracking models in Tables 9 and 10).

Table 7 *GAM* fitted to TFT in aloud reading, using word position in chunk, chunk length in number of words, and word length as fixed effects, with subjects as random effects

	Estimate	ST.E	T Value	PR(>|T|)
Intercept	.276	.0083	33.21	<2e-16 ***
Position in chunk	−.029	.0029	−10.01	<2e-16 ***
Chunk length	−.005	.0023	−2.23	.02 *
Word length	.0326	.0005	61.63	<2e-16 ***
Subjects	<2e-16 ***			
Dev. explained	53.2%			

Table 8 *GAM* fitted to TFT in silent reading, using word position in chunk, chunk length in number of words, and word length as fixed effects, with subjects as random effects

	Estimate	ST.E	T Value	PR(>|T|)
Intercept	.249	.0114	21.73	<2e-16 ***
Position in chunk	−.013	.0027	−4.88	<1e-6 ***
Chunk length	−.004	.002	−2.11	.03 *
Word length	.0186	.0005	37.77	<2e-16 ***
Subjects	<2e-16 ***			
Dev.. explained	47.5%			

As expected, word length slows down reading. The effect is more prominent in oral reading than in silent reading (consistent with the higher cost of concurrent overt articulation), and affects finger-tracking more than eye-tracking. In fact, finger movements are continuous through space: they do not skip words and do not benefit from parafoveal vision as much as eye movements do. Chunk length (measured by the number of words making up a chunk) is an accelerating factor. This is also shown by the regression plots of Figs. 5 and 7, where differences in slope across chunks of different lengths are always significant for finger-tracking data only. Note, finally, that the position of a word within a chunk turns out to have opposite effects in eye-tracking (where it speeds up fixation times),[2] and finger-tracking (where it slows down finger movements). We interpret the speeding-up effect of word position on fixation times as the compounded benefit of parafoveal (anticipatory) information and incremental chunk processing. Later words in a chunk are more likely to be predicted and easier to be integrated than earlier words, due to the concurrent availability of reader's topdown (syntactic) expectations and bottom-up (parafoveal) information. That no such effects are observed in finger-tracking seems to suggest that finger movements are less sensitive to complex top-down and bottom-up processing effects, and confirms that finger-tracking is more compatible with a sequential reading strategy.

Table 9 *GAM* fitted to TTT in aloud reading, using word position in chunk, chunk length in number of words, and word length as fixed effects, with subjects as random effects

	Estimate	ST.E	T Value	PR(>|T|)
Intercept	.0575	.0055	10.40	<2e-16 ***
Position in chunk	.0189	.002	9.07	<2e-16 ***
Chunk length	−.0312	.002	−19.29	<2e-16 ***
Word length	.0489	.000	137.65	<2e-16 ***
Subjects	<2e-16 ***			
Dev.. explained	47.7%			

Table 10 *GAM* fitted to TTT in silent reading, using word position in chunk, chunk length in number of words, and word length as fixed effects, with subjects as random effects

	Estimate	ST.E	T Value	PR(>|T|)
Intercept	.0243	0095.	2.56	.01 *
Position in chunk	.0125	.0017	7.144	<9e-13 ***
Chunk length	−.0156	.0014	−11.46	<2e-16 ***
Word length	.0395	.000	131.59	<2e-16 ***
Subjects	<2e-16 ***			
Dev.. explained	50.4%			

[2] The effect is not shown in the regression plot of Fig. 5, whose model does not include word length as an independent variable.

To sum up, in adults' reading the finger appears to play the twofold role of pace maker and place holder. It slides smoothly across the text units until it reaches a suitable syntactic boundary where EVS is checked and kept down to a manageable span with either longer fixations, or regressive saccades. This explains why tracking times are, on average, shorter than the corresponding fixation times. Since finger-tracking is mainly guided by syntactic and prosodic information, while playing little role in actual decoding, its pace is modulated by natural syntactic joints, as well as stress and pauses within prosodic domains. This account explains why the correlation between tracking times and fixation times increases with levels of text analysis: from words to chunks, to sentences. Ultimately, it is at the level of the sentence, arguably the largest EVS-checking domain, that the two measures correlate nearly perfectly, as the end of the sentence is where articulation and eye movements are eventually synchronized. This also provides some reason why the correlation of tracking times and fixation times is higher in oral reading than in silent reading. As observed above, in silent reading eye movements are less sequential, parafoveal vision plays a more extensive and effective role, and saccades are longer. When skipping a long stretch of nonfixated words widens EVS, the ensuing fixation gets longer to compensate for it, and because of heavier processing demands (Kliegl et al. (2006)). The finger, as we saw, typically does not skip words: nonetheless, its pace must be fast enough, on both nonfixated and fixated words, to catch up with the eyes, and this reduces the correlation between tracking times and fixation times at the word level. If our interpretation of the role of text pointing in adults' reading goes in the right direction, then finger-tracking data are not only more compatible with a sequential reading strategy, but they even favour such a strategy, by serially pacing fixations. In the end, although text pointing in adults' reading no longer plays the role of direction/attention controller that is observed in child reading, nonetheless it may force the eyes to behave in a more "oral" reading mode, whether articulation is overt or not. Clearly, to prove this, one needs to eye-track and finger-track a reader at the same time, which is something we intend to do in future trials.

7 Conclusions

Of late, finger and eye movements have been found to provide highly congruent dynamic patterns during exploration of images that are displayed on a computer's touchscreen (Lio et al., 2019). This is not surprising *per se*, and relates to previous work on the synergistic behaviour of fingers and eyes in tasks requiring synchronization of fast eye movements and the slower motor system controlling the fine coordination of finger movements (e.g., Furneaux & Land, 1999; Inhoff & Gordon, 1997). To our knowledge, however, no one has so far investigated the concurrent dynamics of adults' eyes and fingers in a highly demanding cognitive task such as reading. Albeit preliminary, the evidence reported here shows, for the first time, that the two patterns strongly correlate in adults' reading. This is in line with previous evidence (Marzi et al., 2020; Taxitari et al., 2021) reporting a somewhat more expected correlation in child's reading, where text finger-pointing is known to help beginning readers control directional movement, attention focus and voice-print match (Mesmer & Lake, 2010; Uhry, 2002). Here, we focused on the role of text pointing in adults' reading, where it has arguably lost its directional or attentional role.

Adults' finger-tracking times and eye fixation times appear to correlate highly in both oral and silent reading. Their correlation is far from being perfect at the word level, as the two dynamics appear to fulfill different functions and follow a different pace. In adults' reading, the eyes take the leading role. They are responsible for processing the written text, and filling in the orthographic buffer with appropriately encoded time series of letters, ready to be converted into sounds. Although they are autonomous from articulation, fixation times are paced by the need to keep the Eye-Voice-Span to a manageable amplitude. Due to the limited capacity of the orthographic short-term working memory, the eyes' processing speed is bounded by the articulatory rate of the reader. At any given point in time, each word in the orthographic buffer must be read aloud before it is replaced by other upcoming words. In this dynamic, major linguistic units such as syntactic chunks and sentences can play the role of EVS-checking domains. At the end of each chunk, a natural reading pause helps keep articulation and visual processing in step. It looks like finger-tracking in adults' reading plays the role of marking these supralexical domains. Incidentally, this does not mean that the pace of finger-tracking reflects only supralexical levels of processing. Some laboratory evidence not discussed here shows that the finger-tracking speed is also modulated by word stress and word structure patterns, particularly in longer words. Nonetheless, it is at the level of larger linguistic structures that finger-tracking times are more closely related to fixation times, at least in adults' reading.

Our results highlight the usefulness of finger-tracking data as a proxy of more established reading evidence, like eye-fixation data, which nonetheless appears to require more sophisticated and invasive protocols for data collection. We showed that an integration of Artificial Intelligence and Natural Language Processing technologies, exposed as web services on a cloud-based architecture, and a simple web application running on a common tablet, can go a long way toward collecting rich behavioral data that have so far been confined to highly controlled laboratory settings. In addition, reading data collected with finger-tracking are remarkably naturalistic, because they involve full text reading on a very simple and friendly device like a tablet.

We believe that finger-tracking has the potential to offer novel opportunities for reading research, by complementing existing evidence and protocols with new data, which are particularly interesting to investigate from a developmental perspective. For example, we conjecture that a turning point in reading development may involve a radical change in the use of text pointing during reading: for a mature reader, the finger stops playing the role of attention and direction tracker to acquire the subsidiary role of a pace maker and a marker of EVS-checking domains.

The great potential of mobile information technology and cloud computing for huge data collection and analysis makes the finger-tracking methodology especially suitable for extensive reading assessment activities in primary schools. The computing architecture described here supports highly parallel and distributed processes of data acquisition, which can be delivered in real time to research, clinical and education centers as terminals for data modeling and quantitative analysis. Large-scale studies can be conducted, paving the way to more generalizable results than ever in the past. In addition, the possibility to take single-subject measurements on more occasions and in different settings makes finger-tracking evidence suitable not only for group studies, but also for individual diagnostic purposes and large developmental studies.

A recent piece of (neuro)cognitive literature has raised serious concerns regarding the detrimental effects of digital technology on reading and cognitive development (Baron, 2015; Carr, 2020; Greenfield, 2015; Maffei, 2018; Wolf, 2018). However, evidence that digital reading interferes with learning and cognitive development is still inconclusive, and mostly based on internet text materials and digital books which are not optimally enhanced for educational purposes (Kong et al., 2018; Clinton, 2019; Furenes et al., 2021). Although we are ready to acknowledge that educational digital editing is still in its infancy, and much more effort should be put into redesigning present digital formats for child reading, we believe that current advances in information technology (e.g., machine learning, natural language processing and artificial intelligence, but also portable devices and cloud computing) may enable new forms of reader-book interactivity and content adaptivity, based on a detailed assessment of the child's reading profile. In the near future, assistive digital technologies may compare well with current adult reading mediation (at home or in the classroom), boosting emergent readers' motivation and self-confidence, and helping educators assess and address specific reading difficulties. At the same time, this will provide massive, naturalistic data for quantitative analysis and modeling in reading research, thereby advancing our understanding of reading and cognitive development at an unprecedented rate.

Acknowledgements The Italian National Strategic Research Grant (PRIN) 2017W8HFRX *ReadLet: reading to understand. An ICT driven, large-scale investigation of early grade children's reading strategies* (2019–2023) is gratefully acknowledged. Authors are listed in alphabetical order. Their contribution to the present work, according to the CRediT taxonomy, is as follows: Conceptualization: DC, MF, CM, VP; Methodology: MF, CM, AN, VP, LT; Software: MF; Validation: MF, CM, AN; Formal analysis: CM, AN; Investigation: MF, AN, LT; Resources: AN, LT; Data curation: MF, AN, LT; Writing – original draft preparation: CM, AN, VP, LT; Writing – review and editing: CM, VP, DC; Visualization: CM, AN; Funding acquisition: DC, VP.

References

Abney, S. P. (1991). Parsing by chunks. In R. C. Berwick (Ed.), *Principle-based parsing* (pp. 257–278). Springer.

Alamargot, D., Dansac, C., Chesnet, D., & Fayol, M. (2007). Parallel processing before and after pauses: A combined analysis of graphomotor and eye movements during procedural text production. *Studies in Writing, 20*, 13.

Angele, B., Tran, R., & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 526.

Baddeley, A. (2007). *Working memory, thought, and action* (Vol. 45). Oxford University Press.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283–316.

Baron, N. S. (2015). *Words on screen: The fate of reading in a digital world*. Oxford University Press.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.

Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language, 96*, 9–22.

Butsch, R. L. C. (1932). Eye movements and the eye-hand span in typewriting. *Journal of Educational Psychology, 23*(2), 104.

Carr, J. W., Pescuma, V. N., Furlan, M., Ktori, M., & Crepaldi, D. (2021). Algorithms for the automated correction of vertical drift in eye-tracking data. *Behavior Research Methods*, 1–24.

Yekaterina Chzhen, Anna Gromada, Gwyther Rees, Jose Cuesta, and Zlata Bruckauf. An unfair start: Inequality in children's education in rich countries. Technical Report 15, UNICEF Office of Research, , 2018.

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading, 42*(2), 288–325.

Dare, N., & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *Quarterly Journal of Experimental Psychology, 66*(3), 487–504.

Dell'Orletta, F., Federico, M., Lenci, A., Montemagni, S., & Pirrelli, V. (2007). Maximum entropy for italian pos tagging. *Intelligenza Artificiale, 4*(2).

Federici, S., Montemagni, S., & Pirrelli, V. (1996). Shallow parsing and text chunking: A view on underspecification in syntax. In *Proceedings of ESSLLI'96 workshop on robust parsing* (pp. 35–44).

Ferro, M., Cardillo, F. A., Pirrelli, V., Gagné, C. L., & Spalding, T. L. (2016). Written word production and lexical self-organisation: Evidence from english (pseudo) compounds. In P. Basile, A. Corazza, F. Cutugno, S. Montemagni, M. Nissim, V. Patti, G. Semeraro, & R. Sprugnoli (Eds.), *Proceedings of the third Italian conference on computational linguistics, CLiC-it 2016* (pp. 156–151).

Ferro, M., Cappa, C., Giulivi, S., Marzi, C., Cardillo, F. A., & Pirrelli, V. (2018a). ReadLet: An ICT platform for the assessment of reading efficiency in early graders. In *11th International Conference on the Mental Lexicon* (p. 61).

Ferro, M., Cappa, C., Giulivi, S., Marzi, C., Nahli, O., Cardillo, F. A., & Pirrelli, V. (2018b). Readlet: Reading for understanding. In *Proceedings of 5th IEEE congress on information science & technology (IEEE CiST'18).*

Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research*, 0034654321998074.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye–hand span during musical sight-reading. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 266*(1436), 2435–2440.

Gagné, C. L., & Spalding, T. L. (2016). Effects of morphology and semantic transparency on typing latencies in english compound and pseudocompound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(9), 1489.

Gallant, J., & Libben, G. (2020). Can the maze task be even more amazing?: Adapting the maze task to advance psycholinguistic experimentation. *The Mental Lexicon, 15*(2), 366–383.

Greenfield, S. (2015). *Mind change: How digital technologies are leaving their mark on our brains*. Random House Incorporated.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160.

Inhoff, A. W., & Gordon, A. M. (1997). Eye movements and eye-hand coordination during typing. *Current Directions in Psychological Science, 6*(6), 153–157.

Inhoff, A. W., & Wang, J. (1992). Encoding of text, manual movement planning, and eye-hand coordination during copytyping. *Journal of Experimental Psychology: Human Perception and Performance, 18*(2), 437.

Inhoff, A. W., Morris, R., & Calabrese, J. (1986). Eye movements in skilled transcription typing. *Bulletin of the Psychonomic Society, 24*(2), 113–114.

Jamshidifarsani, H., Garbaya, S., Lim, T., Blazevic, P., & Ritchie, J. M. (2019). Technology-based reading intervention programs for elementary grades: An analytical review. *Computers & Education, 128*, 427–451.

Jarodzka, H., Skuballa, I., & Gruber, H. (2021). Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educational Psychology Review, 33*(1), 1–10.

Kandel, S., Álvarez, C. J., & Vallée, N. (2008). Morphemes also serve as processing units in handwriting production. In *Neuropsychology and cognition of language behavioral, neuropsychological and neuroimaging studies of spoken and written language* (pp. 87–100). Research Signpost.

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). *What's new in psychtoolbox-3?*

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.

Kong, Y., Seo, Y. S., & Zhai, L. (2018). Comparison of reading performance on screen and on paper: A meta-analysis. *Computers & Education, 123*, 138–149.

Laubrock, J., & Kliegl, R. (2015). The eye-voice span during reading aloud. *Frontiers in Psychology, 6*, 1432.

Libben, G., Gallant, J., & Dressler, W. U. (2021). Textual effects in compound processing: A window on words in the world. *Frontiers in Communication, 6*, 42. https://doi.org/10.3389/fcomm.2021.646454. URL https://www.frontiersin.org/article/10.3389/fcomm.2021.646454

Lio, G., Fadda, R., Doneddu, G., Duhamel, J.-R., & Sirigu, A. (2019). Digit-tracking as a new tactile interface for visual perception analysis. *Nature Communications, 10*(5392), 1–13.

Lamberto Maffei. Elogio della parola. , 2018.

Marzi, C., Rodella, A., Nadalini, A., Taxitari, L., & Pirrelli, V. (2020). Does finger-tracking point to child reading strategies? In J. Monti, F. Dell'Orletta, & F. Tamburini (Eds.), *Proceedings of 7th Italian conference on computational linguistics* (Vol. 2769).

Mesmer, H. A. E., & Lake, K. (2010). The role of syllable awareness and syllable- controlled text in the development of finger-point reading. *Reading Psychology, 31*(2), 176–201.

Nicholas Carr. (2020). *The shallows: What the internet is doing to our brains*. WW Norton & Company.

Pate, J. K., & Goldwater, S. (2011). Unsupervised syntactic chunking with acoustic cues: Computational models for prosodic bootstrapping. In *Proceedings of the 2nd workshop on cognitive modeling and computational linguistics* (pp. 20–29).

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

Taxitari, L., Cappa, C., Ferro, M., Marzi, C., Nadalini, A., & Pirrelli, V. (2021). Using mobile technology for reading assessment. In *Proceedings of 6th IEEE congress on information science & technology (IEEE CiST'20)*.

Truitt, F. E., Clifton, C., Pollatsek, A., & Rayner, K. (1997). The perceptual span and the eye-hand span in sight reading music. *Visual Cognition, 4*(2), 143–161.

Uhry, J. K. (2002). Finger-point reading in kindergarten: The role of phonemic awareness, one-to-one correspondence, and rapid serial naming. *Scientific Studies of Reading, 6*(4), 319–342.

Witzel, N., Witzel, J., & Forster, K. (2012). Comparisons of online reading paradigms: Eye tracking, moving-window, and maze. *Journal of Psycholinguistic Research, 41*(2), 105–128.

Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. Harper.

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), 965–992.