

John Benjamins Publishing Company



This is a contribution from ML 18:3
© 2023. John Benjamins Publishing Company

This electronic file may not be altered in any way. The author(s) of this material is/are permitted to use this PDF file to generate printed copies to be used by way of offprints for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy at <https://benjamins.com/content/customers/rights>
For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

For further information, please contact rights@benjamins.nl or consult our website:
www.benjamins.com

Eye-voice and finger-voice spans in adults' oral reading of connected texts*

Implications for reading research and assessment


Andrea Nadalini¹, Claudia Marzi¹, Marcello Ferro¹,
Loukia Taxitari², Alessandro Lento,³ Davide Crepaldi,⁴ and
Vito Pirrelli¹

¹ Institute for Computational Linguistics “A. Zampolli” | ² Neapolis University | ³ Biomedical Campus University | ⁴ Scuola Internazionale Superiore di Studi Avanzati

The present paper investigates the interaction between eye movements, voice articulation and the movements of the index finger dynamically pointing to a text line in oral *finger-point reading* of Italian. During finger-point reading, the finger appears to be ahead of the voice most of the times, by a margin that is significantly modulated by the distribution of phrasal and prosodic units in the reading text. Eye movements replicate the same effects on a different time scale. The eye is ahead of both voice and finger by a wide margin (confirming evidence observed for English and German sentence reading), while showing a tendency to re-synchronise with voice articulation at the right edge of strong prosodic units (sentence boundaries). Our evidence suggests a multicomponent view of the time span between the eye/finger and the voice. The span is shown to be the dynamic outcome of an *optimally adaptive* reading strategy, resulting from the interaction between individual decoding skills, the reader's phonological buffer capacity, and the structural complexity of a reading text. Proficient readers modulate their span to compensate for the different timing between word fixation and word articulation, read faster, and dynamically adjust their processing window to the meaningful, prosodic units of a text.

Keywords: finger-point reading, eye-tracking, finger-tracking, eye-voice span, finger-voice span, eye-finger coordination, parallel processing, working memory, phonological buffer, adaptive reading

* The authors' contribution to the present work, according to the CRediT taxonomy, is as follows: Conceptualization & Methodology: AN, CM, LT, MF, VP; Validation, Formal Analysis & Visualization: AN, CM, MF; Data collection: AN, LT; Resources & Data curation: AL, AN, LT, MF; Software: MF; Writing – original draft preparation: AN, VP; Writing – review and editing: CM, DC, LT, VP; Supervision: CM, MF, VP; Funding acquisition: CM, DC, VP.

 Supplementary materials available from <https://doi.org/10.1075/ml.00025.nad.additional> | Published online: 12 March 2024

The Mental Lexicon 18:3 (2023), pp. 366–400. ISSN 1871-1340 | E-ISSN 1871-1375

© John Benjamins Publishing Company

1. Introduction

During reading development, a familiar situation occurs when a child is engaged in so-called “finger-point reading”, i.e. when she reads a text using the index finger of her dominant hand to point to the letters of written words while reading them out. This is known to help children learn to look at print, and support early basic skills such as directional movement, attention focus, and voice-print match (Uhry, 1999, 2002). Finger-point reading requires that learning readers be able to integrate three sources of information about print and speech: “their auditorily anchored awareness of syllables, their linguistic-conceptual knowledge of words, and their unfolding visuospatial understanding of printed words” (Mesmer & Williams, 2015, p. 486), built upon the visual and tactile exploration of the words' spatial dimension. In attaining an efficient synchronisation between word pointing and the onset of word articulation, the reader must resolve the competing information between the multiple syllables she hears and feels, and the individual words she sees on a printed page (Mesmer & Lake, 2010; Uhry, 1999, 2002).

Beyond this developmental evidence, however, comparatively little is known about the precise dynamic of finger movements in the visual exploration of printed letters, and the subsidiary role these movements play in visual reading, in coordination with eye movements and voice articulation. Recently, Lio, Fadda, Doneddu, Duhamel, and Sirigu (2019) studied the connection between eye movements and finger movements in the visual exploration of an image displayed on a computer touch screen. Presented with the blurred display of a picture, subjects were instructed to deblur the image by touching the screen area they wanted to inspect in full resolution. A strong correlation was observed between areas deblurred by touching the screen, and subjects' fixation patterns when a full resolution version of the same image was explored only visually. Spatial patterns of finger movements were found to be congruent with patterns of eye fixations on the same image, suggesting that tactile exploration of an image can be used as an ecological proxy of visual exploration.

In the present paper, we intend to investigate some fundamental properties of finger movements in finger-point reading, by comparing the independent tracking records of adults' finger and eye movements during the oral reading of a connected text. How do the two dynamics differ? What roles are they observed to play? Do finger movements closely approximate eye movements during reading? Do the former provide comparably accurate information about the processing strategies and online performance of a mature reader?

While unimodal aspects of reading have often been explored and investigated independently, much less work has been conducted to study their online inter-

action, also because of the technical difficulty with concurrently recording asynchronous time-series of signals. The *ReadLet* infrastructure for multimodal reading data collection (Ferro et al., 2018; Taxitari et al., 2021) was designed to meet the scientific desideratum of enabling interactive investigation of multisensory aspects of text reading in ecological contexts.

Using an ordinary tablet as a front-end, ReadLet makes it possible to record, store and remotely transmit the voice recording of an oral text reading session, together with the recording of the touch events caused by finger-pointing (hereafter referred to as “finger-tracking”). Offline, the sequence of touch events is time-aligned with the audio recording of aloud reading, and spatially-aligned with the position of words on the text page. This multimodal evidence is shown to provide a fine-grained time-stamped profile of a subject’s reading performance, whether she is an early (Marzi, Rodella, Nadalini, Taxitari, & Pirrelli, 2020) or an experienced reader (Crepaldi et al., 2022).

Unlike in Braille reading (Hughes, McClelland, & Henare, 2014; Nonaka, Ito, & Stoffregen, 2021), finger-pointing does not acquire written information in visual reading. Thus the question naturally arises of what role finger-pointing plays in reading. A straightforward way to address this question would require the concurrent recording of voice articulation, finger-tracking and eye-tracking data. Unfortunately, technical limitations have so far prevented the co-registration of eye and finger movements in the same reading trial. In this paper, we address the issue indirectly, using a *triangulation* approach to data analysis. First, we eye-tracked 55 adults engaged in orally reading connected texts displayed on a PC screen. The data were then used to compute the *Eye-Voice Span* (EVS; De Luca, Pontillo, Primativo, Spinelli, & Zoccolotti, 2013; Inhoff, Solomon, Radach, & Seymour, 2011; Laubrock & Kliegl, 2015; Silva, Reis, Casaca, Petersson, & Faísca, 2016), i.e. the distance between the eye and the voice in reading aloud. We also recorded and analysed the *Finger-Voice Span* (FVS), i.e. the distance between the finger and the voice, in finger-tracked oral reading sessions of the same texts used in the eye-tracked sessions. This allowed establishing the relative order in time and space of eye movements and finger movements, and understanding the independent coordination of eye-movements and finger-movements with the time course of oral reading. Before turning to this comparison, it is useful to focus on the literature on EVS in oral reading, and explore its connection with processing issues.

1.1 The eye-voice span

In reading aloud, eye fixations are known to be ahead of spoken words most of the times (e.g., Inhoff et al., 2011). EVS measures such a gap either in terms of the distance in characters between the currently articulated item and the currently

fixated one (spatial EVS) or in terms of time, i.e. how long it takes to start the articulation of an item after the item is fixated (temporal EVS). In turn, temporal EVS can be measured either from the *onset* of a word fixation to the onset of articulation of the same word (onset temporal EVS), or from the *offset* of a word fixation to the onset of its articulation (offset temporal EVS) (see Figure 1). In adult proficient readers, the eyes have been shown to be ahead of the voice by an average of 13–15 characters (spatial EVS) (Buswell, 1920), and about 500 ms (onset temporal EVS) (Inhoff et al., 2011; Laubrock & Kliegl, 2015). The spatial span of atypical populations was found to lag behind this average (De Luca et al., 2013). Hereafter, we will only be concerned with temporal measures of the eye-voice gap. Thus, we will talk of onset and offset EVS to mean onset and offset *temporal* EVS.

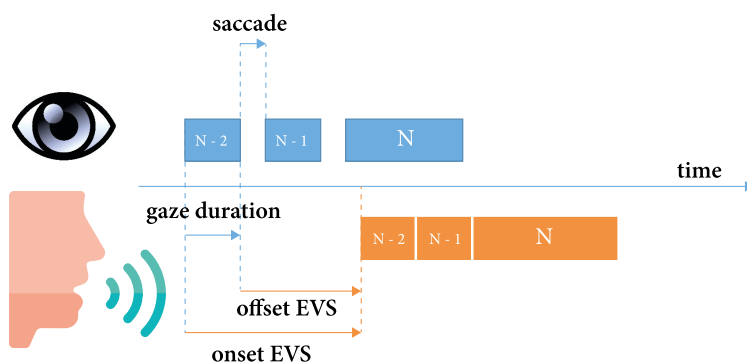


Figure 1. Temporal onset- and offset-EVS between eye-fixations and word articulation in multiple word reading (adapted from Silva et al. (2016)). Blue boxes represent consecutively fixated word tokens, and orange boxes represent the same tokens read aloud

Onset EVS has been considered as a way to measure the time taken by a reader to process a written word. Typically, such a time interval includes the first-pass gaze duration, i.e. the time spent by the reader to fixate a target word for the first time, and its offset EVS, i.e. the time lag between the gaze offset and the articulation onset of the word (Figure 1). By definition, the offset EVS is thus spent to fixate more words *while* processing already fixated words, and is consensually interpreted as a hallmark of parallel processing and automaticity in reading, dependent on decoding speed (e.g. naming velocity: Cohen, Servan-Schreiber, & McClelland, 1992; Moors, 2016), immunity from interference (Cohen et al., 1992), and word familiarity (operationalized as lexical frequency by Silva et al., 2016).

If a reader starts articulating word *N* upon completing its fixation, the corresponding offset EVS is 0. If the reader starts articulating word *N* before completing its fixation, the corresponding offset EVS would be negative. Although

the distribution of offset EVS periods may include o's and negative values, positive offset EVS values are largely prevalent. If the processing of words N and $N + 1$ is more automatic, i.e. if it consumes fewer attentional resources, the processing of more words is facilitated (Protopapas, Altani, & Georgiou, 2013). Thus we expect more skilled readers, e.g. readers with a more rapid articulatory rate (Laubrock & Kliegl, 2015) and a more efficient lexical reading route (Paap & Noel, 1991), to exhibit a longer average offset EVS (Silva et al., 2016). In addition, the effect can also be modulated by other linguistic or text-driven factors such as word frequency, length (Silva et al., 2016) and context-predictability (Kliegl, Nuthmann, & Engbert, 2006), to the effect that high-frequency, shorter and more predictable words typically cause a longer offset EVS.

Why should skilled and efficient readers exhibit a belated articulation onset in the first place? According to Laubrock and Kliegl (2015), the eyes tend to be ahead of the articulatory system because visual processing is faster than articulation. For this reason, a skilled reader buffers the phonological coding of already fixated words until the moment they are articulated. Silva et al. (2016) address the same issue to conclude that delayed articulation is strategic, and is done for the sake of reading efficiency. A list of words can be read more quickly if the eye does not wait for the voice to articulate a word that was just fixated, but rather keep on scanning a list of words ahead of their articulation. The more words are scanned before articulation, the longer their EVS, and the quicker their reading.

These accounts have some merits. Nonetheless, they both appear to neglect the fundamental contribution of word buffering to text reading comprehension and its main acoustic correlate in oral reading: prosodic fluency (Breen, Kaswer, Van Dyke, Krivokapić, & Landi, 2016). Accordingly, text reading comprehension is not just the simple outcome of a process of efficient word recognition. Good readers also “sound good”, as they are able to “chunk” single word recognition into cohesive syntactic units, forming, in turn, coherent prosodic units (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Schilling, Carlisle, Scott, & Zeng, 2007). Text chunking is a challenging task, which requires remarkable look-ahead skills, and a reading strategy aimed at pooling serial information, scattered across multi-word units, into larger meaningful chunks. Letter-by-letter reading defines the entry level of this ability. At the opposite end, reading in chunks is the on-line strategy for achieving a maximum level of reading comprehension via structure integration, with word-by-word reading lying halfway through between the two extremes. The observation is not new. It can be traced back to the 20s, in Buswell's pioneering work (Buswell, 1920, 1921). By switching off the light during the reading of a sentence and counting how many words a reader could articulate after the light was off, Buswell arrived at the conclusion that “[...] an eye-voice span of considerable width is necessary in order that the reader may have an intelligent grasp of

the material read, and that he may read it with good expression" (Buswell, 1920, p. 217). The hypothesis was rekindled and elaborated a few decades later (Lawson, 1961; Levin & Cohn, 1968; Levin & Turner, 1968; Morton, 1964a, 1964b), when some experimental results appeared to support the view that "subjects tend to read in phrase units" (Levin & Turner, 1968, p. 208), and reading rate and EVS were shown to increase with more structured text materials (Morton, 1964a). In particular, Levin and Turner's (1968) suggested "that readers have an elastic span, which stretches or shrinks far enough to read to phrase boundaries" (p. 208). In the remainder of this paper, we provide eye-tracking and finger-tracking evidence supporting such an optimally adaptive reading behaviour.

2. Materials and methods

Participants

Fifty-five young adults (28 female, 27 male, mean age = 27, age range = 18–39) took part in a reading experiment of about 40 minutes, after having signed an informed consent for their involvement. They were all native Italian speakers, with normal or corrected-to-normal vision and no history of neurological disorders. Twenty-two experimental sessions were conducted at the CNR Research Area in Pisa, and thirty-three experimental sessions in SISSA, Trieste. Our sample is larger than any of the samples collected and analysed so far in the literature to detect voice span effects. In addition, we assessed the statistical power of our analyses *post-hoc*, via a simulation-based estimation (Brysbaert & Stevens, 2018) implemented in R with the *mixedpower* package (Kumle, Vö, & Draschkow, 2021). Simulations show that our sample size provides the comprehensive linear mixed-effect model of Section 3.3 with a statistical power of at least 80%. The full results of power analysis are provided in the paper's *Supplementary Materials* (S7).

Procedure

The study is based on a 2 (tracking method: eye vs. finger) by 2 (reading condition: silent vs. aloud) Latin square, fully counterbalanced design. Each participant was involved in four experimental tasks, all taking place in a single day. For each experimental condition, participants were asked to read two different texts while wearing a pair of wireless noise-cancelling headphones with a retractable microphone. Upon reading each text, the participant had to answer a single multiple choice question of reading comprehension. Order of delivery of the tracking

method and reading condition were counterbalanced across participants. Also the presentation of the different reading material alternated among participants, for them to be equally distributed across experimental conditions. Before starting the reading sessions, participants were presented with a short snippet to familiarize with the task.

Text materials

The reading material consisted of eight Italian texts: four excerpts from Roberto Saviano's tabloid news articles, and four passages extracted from Lamberto Maffei's popular neuroscience book *Elogio della parola* ('In praise of words', Maffei, 2018). Each text extended over two tablet pages.¹

All texts were annotated using a battery of state-of-the-art NLP tools including READ-IT (Dell'Orletta, Montemagni, & Venturi, 2011), a readability assessment tool for the Italian language, which combines traditional text features with lexical, morpho-syntactic and syntactic information. Texts were automatically tagged for Part of Speech (PoS), shallow-parsed into word chunks, and annotated for syntactic dependencies.² Chunking (Abney, 1991; Federici, Montemagni, & Pirrelli, 1996; Lenci, Montemagni, & Pirrelli, 2003) defines a non-recursive level of phrasal text segmentation, where lexical heads are always the rightmost word token in the chunk, and provides a linguistically principled way to explore the correlation between acoustic, prosodic and syntactic cues in reading a connected text (Pate & Goldwater, 2011). A summary of the linguistic features annotated in our reading texts is presented in Table 1.

1. Each tablet page fit a 10-inch tablet screen. In eye-tracked reading sessions, words were displayed on a computer screen using a layout comparable to that used in the tablet. The font size was adjusted to the eye distance from the screen so that the angle required to frame a single letter on a computer screen is as close as possible to that required on the tablet screen.

2. For PoS-tagging, we used the coarse-grained level of the ISST-TANL morpho-syntactic tagset (Dell'Orletta, Federico, Lenci, Montemagni, & Pirrelli, 2007), with output in CoNLL format. For annotation of syntactic dependencies, we used the DeSR dependency parser (Attardi, 2006). Frequency distributions were extracted from the Paisà Corpus (Lyding et al., 2014).

Table 1. Annotation statistics for reading texts by text author: IPU (Implicit Prosodic Unit) defines a sequence of words delimited by punctuation marks

	All		Saviano		Maffei	
	Mean	SD	Mean	SD	Mean	SD
word length [letters]	5.17	3.11	4.89	2.95	5.52	3.26
text length [words]	278.75	37.99	308.5	12.79	249	26.49
PoS types	11	0.76	11.5	0.58	10.5	0.58
IPU length [letters]	39	29	35	24	44	34
sentence length [words]	26.99	18.63	20.22	10.98	47	22.20
chunk length > 1 [words]	2.26	0.50	2.24	0.47	2.29	0.54
dependency length [words]	2.44	3.83	2.18	2.54	2.76	4.93
word log frequency	4.32	1.66	4.4	1.77	4.22	1.64

Apparatus

Upon each trial, subjects were asked to read aloud a text displayed on a tablet, while finger-pointing to written words as they read them out. Large streams of time-aligned signals were then automatically captured by the tablet, including time-stamped finger touch events, reading time, accuracy and timing of question answering, voice recording, along with text coordinates on the screen. At the end of each session, anonymised data were encrypted and sent to a centralised server through an Internet connection.

Eye-tracking

Eye movements were recorded via an Eyelink Portable Duo eyetracker (SR Research, Canada), which allows for head-free eye-tracking with a reported accuracy of 0.25° to 0.50° degrees. Only the right eye of each participant was tracked, at a 500 Hz sampling rate. Before the actual experiment started, a 9-point-calibration procedure was conducted until the average error was below 0.5 degrees of visual angle. Next, a 9-point accuracy test was performed to validate eye position. No chin-rest was used during the experiment in either reading mode. Reading materials were shown on a 24 inches Dell S2421H Screen at a resolution of 1920 × 1080. Stimulus presentation and eye movements recording were handled with Matlab Psychtoolbox (Brainard, 1997).

Finger-tracking

Finger-tracking data collection was conducted running the ReadLet application (Ferro et al., 2018) on a 10.1 inches tablet, equipped with a 1.8 GHz Octa-Core processor, 3 GB RAM, 64 GB eMMC and Android 10. The tablet screen was 14.9 cm × 24.5 cm with a resolution of 1920 × 1200 pixels. Finger movements were sampled at a 120Hz rate, corresponding to 24 touch events per syllable when a written word is read at a speed of 5 syllables per second.

Speech recording

In the eye-tracked oral reading condition, participants' speech was recorded with Razer Nari Essential RZ04-02690 earphones equipped with a bidirectional electret microphone (frequency range of 100–6500Hz and a sensitivity of -42dBV/Pa). In the finger-tracked oral reading condition, participants wore BlueParrott S450-XT earphones equipped with a bidirectional electret microphone (frequency range of 150–6800Hz and a sensitivity of -47dBV/Pa). For both eye-tracking and finger-tracking settings, the audio signal was sampled at 48000Hz, 16bit, stereo and compressed to audio-WEBM format at 128kbps.

Data pre-processing and cleaning

Eye-tracking

Out of the 55 participants involved in the experiment, 5 were excluded from the analyses because of technical failure in eye recording (3) or speech recording (2). This reduced the size of the original dataset to 50 readers and 200 individual pages. Of these pages, 16 were further taken out because of noise artifacts in the recorded signal that made fixation sequences uninterpretable (8% data loss). Individual fixations shorter than 50 ms or longer than 800 ms were excluded (2% data loss), together with those falling more than 60 pixels out of each text bounding box (0.65%). Similarly, early fixations falling below the y -coordinate of the first line of the text were dropped (< 1%). After correcting for vertical drifts using the “Warp” algorithm (Carr, Pescuma, Furlan, Ktori, & Crepaldi, 2021), the dataset was further filtered to exclude individual datapoints with first-pass gaze duration longer than 1500 ms (86 individual word tokens, 0.3% exclusion rate). The resulting database comprised 184 pages, 29578 fixations and 19127 fixated word tokens.

Finger-tracking

Text-to-finger alignment was carried out automatically, using a convolutional algorithm finding the closest match between text lines and touch event sequences. For each uninterrupted time series of touch events falling within a letter bounding box, tracking time was computed as the difference between the last time tick and the first time tick in the continuous series of touch events. Finally, the finger-tracking time for all other units in the text was defined as a summation of the tracking times of the letters the unit spans over. One participant was excluded from the analyses because of technical failure in speech recording. 1792 datapoints (6% of the data) were further filtered out because their tracking time was lower than 35 ms or higher than 1500 ms. The final database thus comprised 216 pages and 27935 word tokens.

Speech processing

Speech-to-text conversion was carried out using Vosk (Shmyrev & Vosk Core Team, 2020), an open-source, free speech-recognition toolkit built on Kaldi (Povey et al., 2011). For each word token, Vosk outputs the word's alphabetic transcription and the associated confidence level, as well as the onset and offset time-points of the word's articulation. We collected voice data for a total of 54896 word tokens, out of which 50520 were correctly recognized (94% of the data). All cases of word repetition (326 instances overall, < 1% of the data) were excluded, leaving us with 50194 transcriptions. Of these, 23177 came from eye-tracked sessions, and 27017 from finger-tracked sessions.

Eye-voice span (EVS)

Out of 23177 transcribed word tokens from the eye-tracked data, 5615 were matched with words skipped by the eye (24.2%). The remaining 17562 were correctly associated with the fixated words. The corresponding onset/offset eye-voice span was computed, both spatially and temporally. For our present purposes, we selected word tokens with an eye-voice span between -500 ms and 2500 ms, for a total of 16483 datapoints (6.1% data loss).

Finger-voice span (FVS)

Out of 27017 transcribed word tokens from the finger-tracked data, 25315 (91.2%) were correctly matched with the corresponding finger-tracked words. By analogy to EVS, we also computed the onset/offset finger-voice span, defined as the difference in time (or letter space) between the first/last "touch" event on a given word, and the onset of the word's articulation. We then further excluded any individual datapoint with a finger-voice span below -1000 ms or above 2000 ms (error rate

< 1%), which, by visual inspection, proved to be due to errors. After data trimming, the dataset used for the present analyses comprised a total of 25239 words and 216 pages.

Modelling issues

Time-dependent correlations between signal time-series were computed using *Dynamic Time Warping* (DTW: Sakoe & Chiba, 1978) and *Time-Lagged Cross-Correlation* (TLCC). DTW measures the similarity between two temporal sequences that vary in speed. The method can stretch or squeeze two sequences of time indices non-linearly to make them optimally match, mapping each index in one sequence onto one or more indices in the other sequence, with no crossing of matching lines (i.e. no time shuffling). TLCC identifies *directionality* between two signals (such as a leader-follower relationship in which the leader initiates a response repeated by the follower) and computes their distance in terms of the optimal time-lag required to maximize their cross-correlation. DTW and TLCC were calculated using the `dtw` and `xcorr` functions respectively, both available in the Signal Processing Toolbox version 8.3 of MATLAB 9.7 (R2019b). All other analyses were conducted using the R statistical software, version 4.3.1 (R Core Team, 2023). In particular, we report Generalised Linear Mixed effect Model (GLMM) results specific of eye- and finger-tracking data, as implemented in the package `lme4`, version 1.1–35.1 (Bates et al., 2009). These analyses make it possible to investigate experimental effects with statistical control for possible confounds brought by subject- and word-level variability. In addition, since we wanted to investigate *online* processing data (i.e. the actual processing of a word or sentence from the moment it is visually perceived), which consist of multiple measures per trial that may vary continuously with time, we modelled the non-linear interaction of EVS and FVS using Generalised Additive Mixed Models, or GAMs, using the package `gam4`, version 0.2–6, as they do not assume a linear relation between the fitted variable and its predictors. All plots were created via the `ggplot2` package, version 3.4.3.

All generated measures and statistical models can be found in the paper's *Supplementary Materials*, which include the full set of our data.

3. Results

General descriptive statistics of eye movements, finger movements and articulation duration in oral text reading are reported in Table 2. In the table, “first fixation duration” refers to the duration of the first fixation landing on a word. “First-pass gaze duration” includes possible refixations before the eye moves to

another word. "Total gaze duration" is the sum of the duration of all fixations on a word, including regressive fixations. The table also reports the mean length of a (forward or backward) saccade, the probability for a single word to be skipped, fixated once ("single fixation probability") and fixated more than once ("multiple fixation probability"). Most of these measures are inherent to the specific nature of the eye-tracking signal and have no equivalent in finger-tracking data. Onset and offset temporal voice spans are reported in milliseconds for both the eye and the finger.

Table 2. Descriptive statistics for eye-tracked and finger-tracked oral text reading. Regressive finger movements are virtually absent in adults' finger-point reading

Eye-tracked	Mean	SD	Finger-tracked	Mean	SD
first fixation duration [ms]	280	133			
first-pass gaze duration [ms]	349	198			
total gaze duration [ms]	411	227	word tracking duration [ms]	292	216
forward saccade length [letters]	7.81	4.19			
regression length [letters]	5.99	5.55			
word skipping probability	0.25	0.43			
single fixation probability	0.585	0.49			
multiple fixation probability	0.165	0.37			
regression probability	0.21	0.41			
spoken word duration [ms]	325	212	spoken word duration [ms]	335	220
spoken sentence duration [sec]	76.87	61.66	spoken sentence duration [ms]	92.26	73.22
onset EVS duration [ms]	889	347	onset FVS duration [ms]	278	272
offset EVS duration [ms]	540	382	offset FVS duration [ms]	-014	334

Finger tracking a text requires a reader to be engaged in the additional task of finger pointing to words on a touch screen while reading them out. In principle, this could affect some aspects of a reader's performance. A Welch *t*-test on the distribution of spoken word durations between eye-tracked (ET) and finger-tracked (FT) data shows a small, but statistically significant difference ($t = -7.45$, p -value < 0.001) between the two modalities, with a mean spoken word duration of 0.32 sec for ET sessions, and 0.34 sec for FT sessions. Likewise, the difference in speech velocity between the two modalities is statistically significant ($t = 12.12$, p -value $< 2.2e - 16$), with a mean spoken letter rate of 18.12 letters per second for ET reading, and 17.47 letters per second for FT reading. Additionally, we compared the distribution of voice breaks (i.e. silent pauses between two consecu-

tively spoken words) across ET and FT oral reading, in three different conditions: (i) when there is no intervening punctuation mark between two consecutive written words in the reading text, (ii) when the punctuation mark is weak (i.e. a comma or a colon), and (iii) when the punctuation mark is strong (i.e. a full stop or a question mark). The qualitative pattern is the same in both tracking modalities. Voice breaks in conditions (ii) and (iii) are significantly longer than in condition (i) (Wilcoxon's $p < 0.001$), with breaks in (iii) being systematically longer than breaks in (ii). In ET oral reading, however, pauses tend to be slightly shorter than in FT oral reading, in line with their difference in speed. The evidence shows that the additional task of finger-tracking, however natural, is likely to slightly tax a reader's working memory and consume some attentional resources. Nonetheless, the fact that the difference between ET-only and FT-only oral reading is very small (in terms of both mean spoken word duration/speed and silent pause duration) makes the additional finger-tracking task an affordable price to pay during a reading task.

3.1 The dynamic of EVS and FVS

In the top panels of Figure 2, time series of eye fixations (left) and time series of finger touch events (right) for a single sentence are plotted along with the corresponding time series of spoken data. Eye coordinates of the y-axis vary in "jumps" (corresponding to saccade lengths in letters), and thick horizontal lines correspond to fixation durations. In contrast, finger and voice y-coordinates vary continuously in time and space. In the same figure, the two bottom panels plot the spatial distance between the eye and the voice (or spatial EVS, left panel), and between the finger and the voice (or spatial FVS, right panel). The distance is computed as the difference, in number of letters, between the position of a letter tracked at time t_k and the position of the letter spoken at the same time tick. In the EVS panel, peaks correspond to forward saccades, and valleys mark eye regressions or lingering fixations. In the FVS panel, a smoother jagged line depicts the continuously varying distance between finger and voice during reading, which in turn reflects variation in speed between the two signals.

Figure 3 shows the distribution of onset and offset EVS (left panel), and onset and offset FVS (right panel). Distributions are bell-shaped, with a slightly heavier right-hand tail. The time difference between mean onset EVS (889 ms, $SD=347$) and mean *offset* EVS (540 ms, $SD=382$), shows that EVS has an average headstart reduction of 350 ms at the offset of the word's first-pass gaze duration. In contrast, the average finger headstart of about 278 ms ($SD=272$) at the onset of a word vanishes at the word's offset. This indicates that the voice and the finger are more in synch than the voice and the eye.

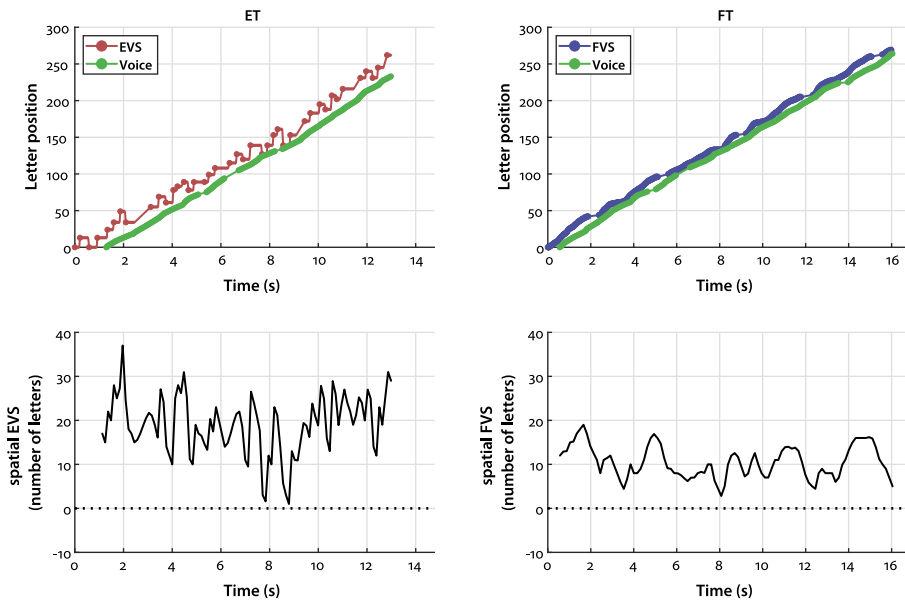


Figure 2. Eye-tracked and finger-tracked reading data for a single sentence, recorded by two subjects reading the same text. Top-left panel: Time series of (i) the position of each fixated letter (red dots) and (ii) of the position of each letter being articulated (green dots). Each dot marks the onset of a fixation, whose duration is shown as a horizontal red segment; a saccade following the fixation is depicted as a thin gray segment. The articulation time of each letter in a word is estimated assuming a constant speed of the voice across the whole word (Laubrock & Kliegl, 2015). Top-right panel: time series of (i) the position of each finger-tracked letter (blue dots) and (ii) the position of each letter being articulated (green dots). Bottom panels: time series of (i) EVS (left panel) and (ii) the FVS (right panel) computed as the difference between the position of each fixated (or finger-tracked) letter at time tick t_k and the position of the letter being articulated at the same time tick. In both panels, a dotted horizontal line marks a null span ($y=0$), i.e. the point in time when the eye and the voice (or the finger and the voice) are, respectively, fixating and articulating (or tracking and articulating) the same letter

A time-dependent correlation analysis of the three time series confirms this observation. Note first that spoken word times, averaged across subjects, correlate with word finger-tracking times more strongly (Spearman $\rho=0.93$) than they do with first-pass gaze durations (Spearman $\rho=0.54$). To factor out the strong correlation of fixation duration, finger-tracking time and articulation duration with word length, we analysed variations in speed (rather than word duration) across the three signals, using Dynamic Time Warping (Sakoe & Chiba, 1978, DTW) and Time-Lagged Cross-Correlation (TLCC). Results of the two analyses (Figure 4)

confirm that the speed of articulation is less correlated to the eye-tracking speed than to the finger-tracking speed. It is worth noticing that the time-shifts required for the the two pairs of time-series to be maximally synchronised using TLCC's optimal cross-correlation lag are in good accord with the average onset voice spans plotted in Figure 3.

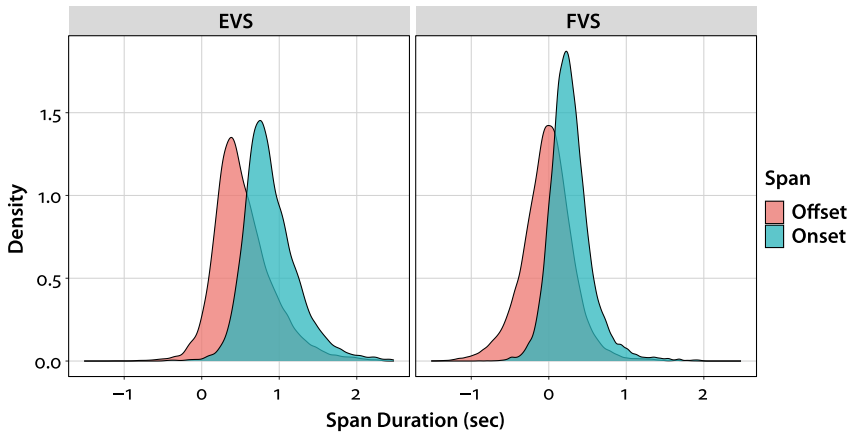


Figure 3. Distribution of time EVS (left panel) and time FVS (right panel) measured in seconds from a word eye-fixation and finger-tracking onset (cyan), and a word eye-fixation and finger-tracking offset (red), respectively

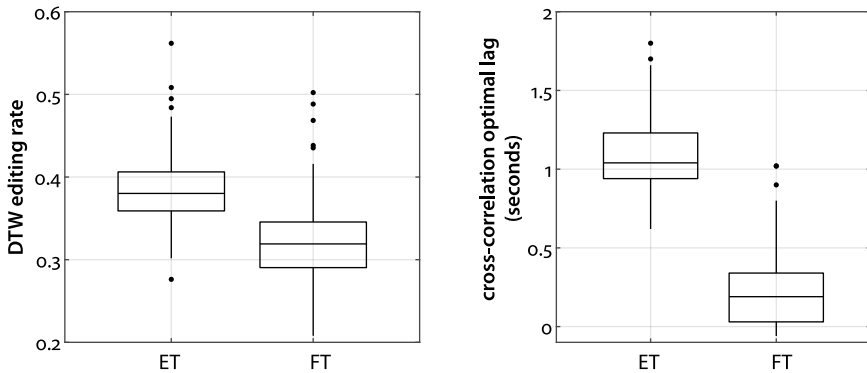


Figure 4. Analysis of the asynchrony between voice speed and eye speed (ET), and between voice speed and finger speed (FV) using DTW (left) and TLCC (right). Left panel: Distribution of the per-page DTW editing cost of aligning the two time series. Right panel: Distribution of the per-page optimal cross-correlation lag. In both the left and right panels, the ET and FT distributions are significantly different (Wilcoxon's $p < 10^{-40}$)

3.2 Onset voice span and eye/finger movements

Table 3. Fixed-effect estimates of GLMM testing first-pass gaze duration as a function of onset EVS and its interaction with word length and frequency

gaze dur ~ poly(Onset EVS, 2) * (length + frequency) + (1+EVS token) + (1+EVS subject)				
fixed effects				
	Estimate	Std. error	t-value	p-value
(Intercept)	-1.32	0.03	-45.83	0.00
poly (Onset EVS, 2) ₁	3.14	2.42	1.29	0.20
poly (Onset EVS, 2) ₂	-23.97	0.93	-25.79	0.00
frequency	-0.04	0.01	-3.75	0.00
length	0.14	0.01	13.30	0.00
poly (Onset EVS, 2) ₁ :frequency	2.58	1.45	1.78	0.08
poly (Onset EVS, 2) ₂ :frequency	-0.50	1.03	-0.48	0.63
poly (Onset EVS, 2) ₁ :length	-1.16	1.48	-0.78	0.43
poly (Onset EVS, 2) ₁ :length	-4.73	1.06	-4.48	0.00

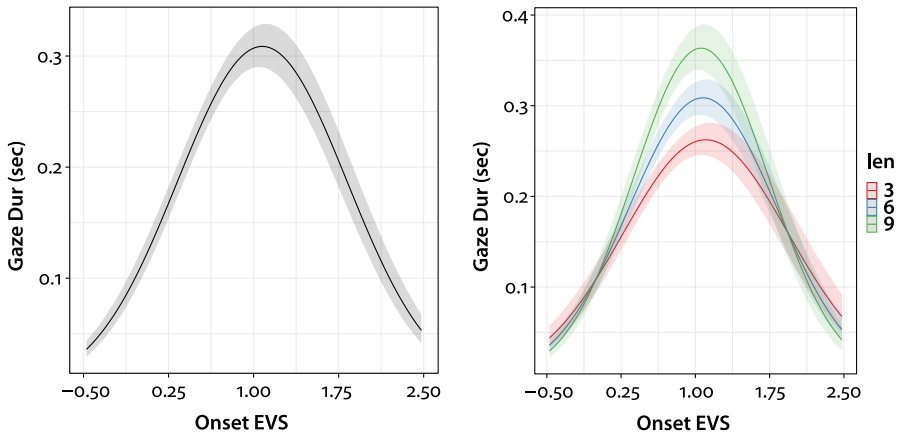


Figure 5. Visualization of GLMM estimates of the main effect of onset EVS on first-pass gaze duration (left) and its interaction with word length (right). Shaded area refer to 95% confidence intervals

A generalized linear mixed effect model (GLMM) fitting fixation duration in adults' reading of connected texts turned out to replicate Inhoff et al.'s (2011) evidence of a significant EVS contribution (entered as a second order polynomial) to fixation duration in sentence reading, with word frequency and length as addi-

tional covariates, and participants and items entered as random effects (Table 3). EVS yields a negative quadratic component with a bell-shaped curvature that goes downwards when EVS increases over 1 second (Figure 5, bottom right corner). A follow-up analysis of the effect of EVS on regression probability and refixation probability (reported in the *Supplementary Materials*) confirms Inhoff and colleagues' hypothesis that the probability of a regression strongly correlates with long onset EVS's.

Similar non-linear effects were replicated with finger-tracking data (Table 4 and Figure 6). A positive first-order component of the non-linear contribution of FVS to tracking duration confirms increasing tracking times for high FVS values. A negative second-order component of a GLMM model fitting finger-tracking times with onset FVS, word length and word frequency as predictors, however statistically significant, has a smaller effect than what we observed for EVS. This is shown by the downward curvature of the partial effect of FVS (> 1.5 seconds), which is less prominent than the bell-shaped effect of EVS.

Table 4. Fixed-effect estimates of GLMM testing finger-tracking duration (*tracking dur*) as a function of onset FVS and its interaction with word length and frequency

<i>tracking dur</i> ~ poly (Onset FVS, 2) * (length + frequency) + (1+FVS token) + (1+FVS subject)				
fixed effects				
	Estimate	Std. Error	<i>t</i> -value	<i>p</i> -value
(Intercept)	-1.69	0.03	-48.44	0.00
poly (Onset FVS, 2) 1	28.76	2.42	11.20	0.00
poly (Onset FVS, 2) 2	-5.09	0.93	-6.51	0.00
frequency	-0.05	0.01	-4.07	0.00
length	0.53	0.01	34.48	0.00
poly (Onset FVS, 2) 1:frequency	1.80	1.38	1.78	0.19
poly (Onset FVS, 2) 2:frequency	0.64	0.81	-0.48	0.43
poly (Onset FVS, 2) 1:length	0.81	1.46	-0.78	0.58
poly (Onset FVS, 2) 1:length	3.08	0.77	-4.48	0.00

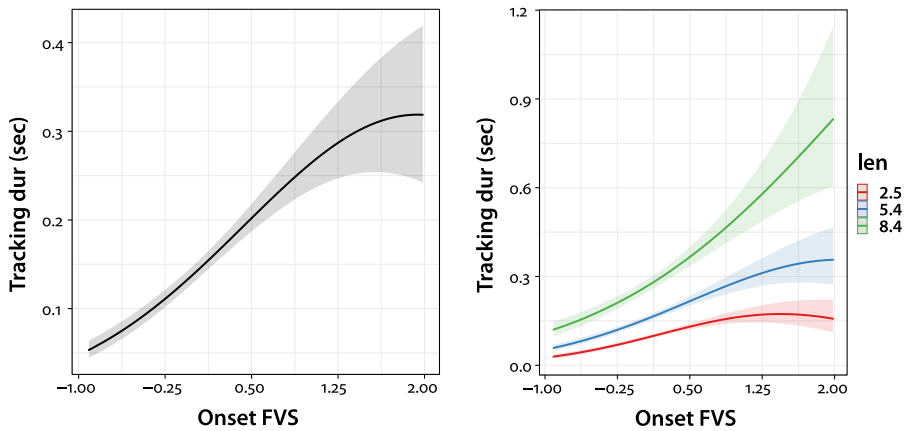


Figure 6. Visualization of GLMM estimates of the main effect of onset FVS on tracking duration (left) and its interaction with word length (right). Shaded areas refer to 95% confidence intervals

3.3 Modelling the offset voice span

To get a more precise understanding of the factors affecting the offset Voice Span, we regressed offset FVS and offset EVS on word length, word frequency, tracking modality and word position to punctuation marks in the text, with subjects and word tokens being entered as random effects (Table 5 and Figure 7). There is a significant interaction between tracking mode and the decreasing effect of word length on the offset span, with the finger being slowed down more by long words than the eye is. Likewise, offset FVS shrinks more prominently after weak prosodic units than offset EVS does. Finally, the increasing effect of word frequency on the voice span is equally marginal in both tracking modalities (no interaction). In contrast, due to the different speed of eye and finger movements, intercepts are always significantly different in the two tracking modalities.

The slow-down influence of punctuation on tracking speed makes an interesting connection with the distribution of voice breaks discussed at the beginning of Section 3. In a follow-up analysis (Figure 8), we then assessed variation in levels of offset FVS (top half) and offset EVS (bottom half) in three conditions: (i) within a syntactic chunk (left panels), (ii) within a weak intonation unit (i.e. a text unit followed by a weak punctuation mark: center panels), and (iii) within a strong intonation unit (i.e. a text unit followed by a strong punctuation mark: right panels). Non-linear regression plots show that the voice span starts high at the left edge of each unit, to get increasingly reduced as more of the unit is read out (the full list of coefficients is provided in the *Supplementary Materials*).

Table 5. Fixed-effect estimates of LMM testing offset voice span (*offset span*) as a function of punctuation type (*punct Type*), word length and frequency for both tracking modes (*trackMode*)

offset span ~ (length + frequency + punctType) * trackMode + (1 token) + (1 subject)				
fixed effects				
	Estimate	Std. Error	t-value	p-value
(Intercept) trackModeET	0.69	0.02	29.09	0.00
length (trackModeET)	-0.03	0.00	-18.13	0.00
frequency (trackModeET)	0.01	0.00	2.54	0.01
punctType1 (trackModeET)	-0.09	0.01	-8.34	0.00
punctType2 (trackModeET)	-0.18	0.01	-12.19	0.00
trackModeFT	-0.44	0.02	-25.16	0.00
length (trackModeFT)	-0.03	0.00	-19.00	0.00
frequency (trackModeFT)	0.00	0.00	1.70	0.09
punctType1 (trackModeFT)	-0.11	0.01	-11.45	0.00
punctType2 (trackModeFT)	-0.04	0.01	-3.20	0.00

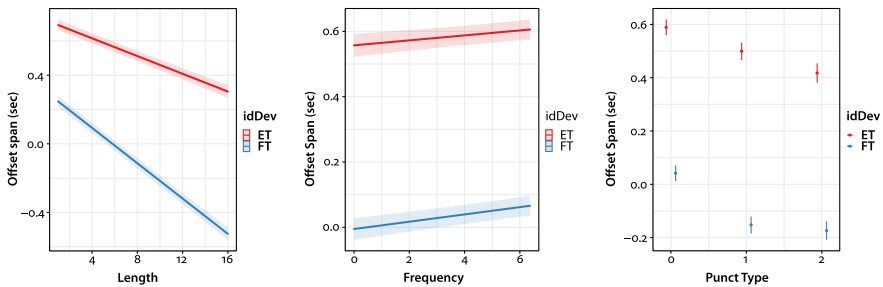


Figure 7. Visualization of LMM estimates of the effect of word length (left), frequency (center) and punctuation type (right) on offset voice span, for eye-tracking (ET, red lines) and finger-tracking (FT, blue lines) data. Shaded areas (left and central panels) and error bars (right panel) refer to 95% confidence intervals

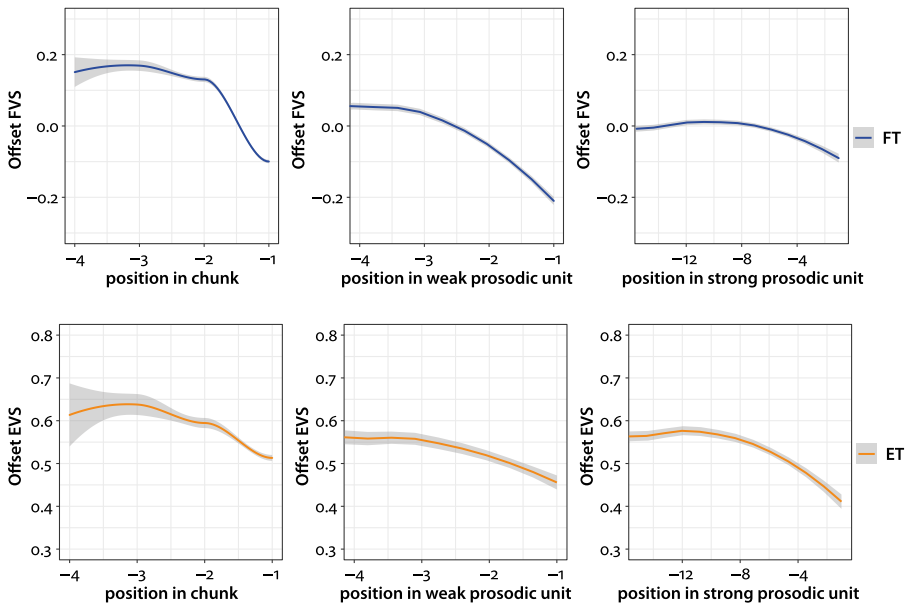


Figure 8. Non-linear regression plots fitting offset FVS (top panels) and offset EVS (bottom panels) as a function of token position from the right edge of a chunk (left plots), a weak (implicit) prosodic unit (center plots), and a strong (implicit) prosodic unit (right plots). In all plots, the position of the right-most word of a chunk/unit is marked as $x = -1$, with $x = 0$ (not shown here) marking the first word token of the immediately ensuing chunk/unit. The plots show the last 4 and 16 word tokens of weak and strong prosodic units respectively. Shaded areas indicate 95% confidence intervals

The plots of Figure 8 average out span variations across units of different length. Possibly, their non-linearity could be an artifact of partially overlaying trends of span variation in chunks of different length. To elucidate the independent contribution of chunk length to variation of the offset voice span across within-chunk positions, two Generalised Additive Models were fitted to offset FVS and EVS, with chunk position and chunk length as fixed effects, and word tokens and subjects entered as random effects (Table 6). Their non-linear plots are shown in Figure 9 for FVS (left panel) and EVS (right panel) respectively, showing that they vary (non-linearly) not only with token position, but also with chunk length. Finally, we modelled how syntactic units and intonation units interact in affecting offset FVS and offset EVS. In the top panels of Figure 10, variations in levels of FVS and EVS are shown for each token position *within* syntactic chunks followed by no punctuation, a weak punctuation, or a strong punctuation. Once more, span levels start high at the beginning of a chunk, to progressively decline towards the end of the chunk. Nonetheless, descents are steeper when the

chunk is followed by a stronger punctuation mark, showing a significant contribution of implicit prosody to span modulation.

Table 6. Fixed-effect estimates of GAMs fitting offset FVS and EVS as a function of token position in a chunk (*chunk pos*) and chunk’s length (*chunk len*), with word token and subjects entered as random effects

offset span ~ (chunk pos * chunk len) + (token=re) + (subject=re)				
	Estimate	Std. Error	t-value	p-value
FVS				
(Intercept) chunk len 2	-0.23	0.02	-9.39	<2e-16
chunk pos (chunk len 2)	-0.10	0.01	-7.38	<0.001
chunk len 3	0.01	0.02	0.71	>0.05
chunk pos (chunk len 3)	0.03	0.01	2.64	<0.01
chunk len 4	0.04	0.04	1.15	>0.05
chunk pos (chunk len 4)	0.06	0.01	4.21	<0.001
tokens, sub jects	<2e-16			R ² 54.5%
EVS				
(Intercept) chunk len 2	0.41	0.03	15.08	<2e-16
chunk pos (chunk len 2)	-0.08	0.02	-5.01	<0.001
chunk len 3	-0.01	0.03	-0.29	>0.05
chunk pos (chunk len 3)	0.02	0.02	1.08	>0.05
chunk len 4	0.03	0.05	0.69	>0.05
chunk pos (chunk len 4)	0.05	0.02	2.39	<0.05
tokens, sub jects	<2e-16			R ² 25.5%

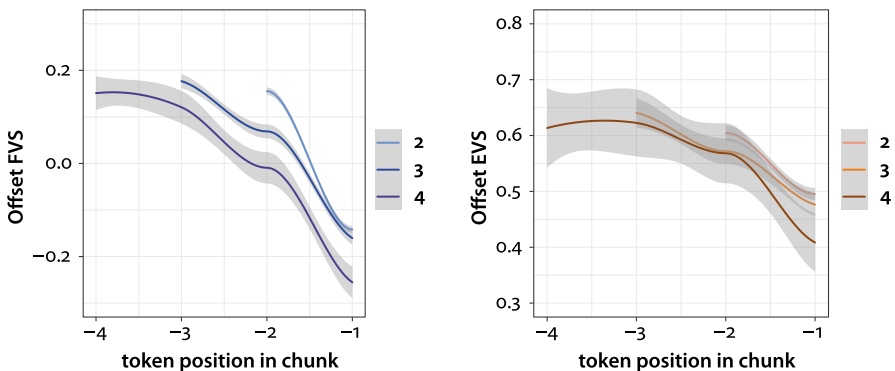


Figure 9. Non-linear regression plots fitting offset FVS (left panel) and offset EVS (right panel) as a function of chunk length (2, 3, and 4 word tokens) and token position from the right edge of a chunk (i.e its lexical head), marked as $x = -1$. Shaded areas indicate 95% confidence intervals

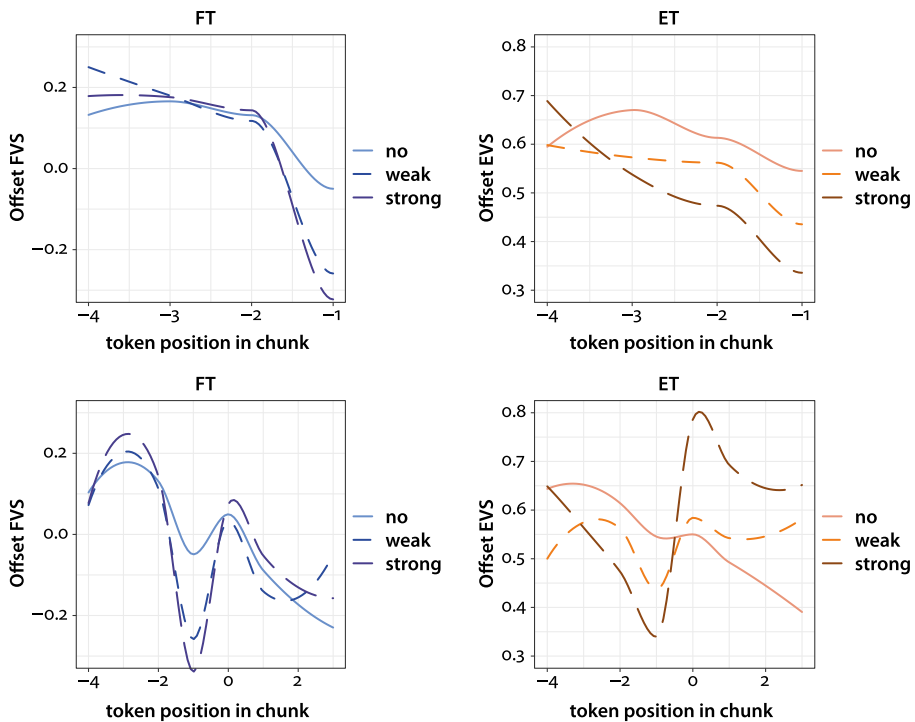


Figure 10. Non-linear regression plots fitting offset FVS (left panels) and offset EVS (right panels) as a function of word position to the lexical head of a chunk ($x = -1$) followed by no punctuation, weak and strong punctuation; top panels plot variation of offset FVS/EVS *within* a single chunk; bottom panels plot the same variation *between* two consecutive chunks

In principle, non-linear effects of chunk length on voice spans are compatible with Laubrock and Kliegl’s (2015) conjecture that span length is a function of the maximum capacity of a reader’s phonological buffer. However, if the span were modulated only by the buffer capacity, we should observe no effect at the transition between consecutive chunks. Since the average chunk length in our texts is 2.26 tokens (Table 1), in most cases readers should simply ignore a chunk boundary, and use up their phonological buffer to stack as many words as possible, no matter whether they belong to the same chunk or two (or more) consecutive chunks. We thus modelled span variation within a larger text window straddling two consecutive chunks. The resulting non-linear plots (Figure 10, bottom panels) show a systematic drop at the transition between consecutive chunks (for $x = -1$, corresponding to the last word in the first chunk, i.e. its lexical head). The drop varies with the type of punctuation marks intervening between the chunks, showing that prosodic units strongly interact with the phonological buffer.

4. Discussion

To understand more of the role finger movements play in oral reading and the ways they are coordinated with ocular and articulatory movements, in this paper we compared classical eye-voice distance measures (EVS) with the functionally equivalent distance between the finger and the voice (FVS). *Oral reading* requires the fine coordination of eye movements and articulatory movements. The eye provides access to the input stimuli needed for voice articulation to unfold at a relatively constant rate. In turn, control on articulation provides internal feedback to oculomotor control for eye movements to be directed when and where difficulties may arise. A factor that makes eye-voice coordination fairly hard to manage is the asynchrony of the two time series. Eye movements are faster than voice articulation, and are much freer to scan a written text forwards and backwards, availing themselves of a wide range of alternative “moves”, including long forward saccades, regressions, refixations and word skipplings. A reader must rely on a tight coordination strategy to ensure that the two processes are optimally integrated. *Oral finger-point reading* introduces a further dimension of complexity by adding an extra time-series of finger movements. Finger movements are slower than eye movements, and need to be controlled and kept in step with the eye and the voice during reading.

A further original contribution of this study is that, to our knowledge, adults' EVS data were collected for the first time during a task of connected text reading (as opposed to single sentence reading and word-list reading). This allowed us to explore a wide range of concomitant processing effects. By comparing effects across different time series, we gained a better understanding of their underlying mechanisms. In particular, we will focus here on assessing the merits of two complementary ways of interpreting our data: (i) the *phonological buffer hypothesis* and (ii) the *adaptive reading hypothesis*.

The phonological buffer hypothesis

Onset *EVS* is classically understood to cover the time taken to process a word for oral reading, resulting from the sum of two components: a word's *first-pass gaze duration* (i.e. the time elapsed between the onset of a reader's first fixation on a word and the onset of the first saccade to a different word), and the *offset span* (i.e. the time elapsed between the moment the eye stops fixating a word and when the voice starts articulating that word). During a word's offset span, a reader typically prepares the articulation of those fixated words that are still waiting to be read out. At the same time, she fixates one or more upcoming words in the text, ultimately engaging in heavy parallel processing.

Laubrock and Kliegl (2015) suggest that Baddeley's (2007) model of a working memory buffer for phonological rehearsal provides a neuro-psychological key to understanding this behaviour. The upper limit of the ability to plan articulation of a fixated word in parallel to viewing other words is set by the capacity of a reader's phonological buffer, where temporarily maintained items are known to decay after a short time. A consequence of this timebound capacity is that the number of units that are stacked in the buffer may vary depending on how long they take to be articulated. A reader with a higher articulation rate can buffer more words than a reader with a lower articulation rate. A further implication is that the number of words waiting to be articulated should not exceed the capacity of the phonological buffer. A too long stack of buffered words may cause some of these words to be uttered in the wrong order, or even forgotten before they are articulated. This explains the need to constantly manage the buffer during oral reading, by controlling eye movements and fixation programming in real time.

Our evidence supports Laubrock and Kliegl's hypothesis. The GLMM models fitting fixation duration in our data successfully replicated Inhoff and colleagues' (2011) models for sentence reading data, showing a non-linear interaction between fixation duration and voice span length (Table 3). Increases in onset EVS are responded to with increases in first fixation duration, for the eyes to be kept relatively close to the word being articulated. We also found a prominent negative quadratic component of onset EVS for span values longer than 1 second (Figure 5), akin to what Inhoff and colleagues found in sentence reading. We concur with them that the negative effect should be interpreted as due to regressive saccades. In reading a connected text, EVS can reasonably be stretched, but an EVS that is longer than 1 second significantly increases the probability of a regression. In our data, this interpretation is indirectly supported by the non-linear interaction between finger-tracking duration and onset FVS (Table 4), which appears to be regulated by the same need to control the distance of the finger from the voice. In this case, however, the negative quadratic component disappears (Figure 6), consistently with evidence that the finger hardly reverses its tracking direction in adults' reading (Table 2).

To investigate interindividual variability, we plotted by-subject random slopes against by-subject random intercepts of the two models (scatter plots can be found in the paper's *Supplementary Materials*). In the first model, random intercepts represent the deviation of each reader's average first-pass gaze duration from the population mean fixation duration, with slower readers being plotted more to the right, and faster readers more to the left in the scatter plot. Random slopes indicate how individual average gaze durations need to be adjusted for increasing EVS values. A strong negative correlation ($r = -0.811$) in the plot suggests that the fixation duration of slower readers is less sensitive to variation of EVS than fast

readers' fixation duration. This is what the phonological buffer hypothesis would predict, as fast readers are more likely to incur longer voice spans. Thus, they are more prone to slowing down their fixations when the span gets too long. The evidence dovetails with Silva and colleagues' (2016) conjecture that faster readers resort to larger EVSs, as larger EVSs increase reading speed. The same negative, albeit weaker correlation is found between individual finger-tracking duration and variation in FVS values, when we plot by-subject random slopes and random intercepts of the model of Table 4. Readers who finger-point more quickly are more prone to slowing down their tracking speed when FVS gets too large. Due to the high correlation between finger-tracking time and articulation time, this evidence accords well with Laubrock and Kliegel's (2015) conjecture that the voice span is modulated by the articulation rate of the reader.

The adaptive reading hypothesis

Average onset EVS (890 ms) in our data is considerably larger than both Inhoff and colleagues' (486 ms, Inhoff et al., 2011) and Laubrock and Kliegel's (561 ms, Laubrock & Kliegel, 2015) mean values. In addition, the interindividual variability of readers' onset spans ($SD=108$ ms for EVS, and $SD=116$ ms for FVS) is smaller than the corresponding intraindividual variability ($SD=330$ ms for EVS, and $SD=246$ ms for FVS). This evidence supports the conjecture, which goes back to Buswell's (1920; 1921) pioneering work, that the voice span is not only an effective indicator of individual reading proficiency and working memory capacity, but it also reflects an *adaptive reading strategy* contingent on the structural features of a reading text. The conjecture resonates well with Levin and colleagues' hypothesis (Lawson, 1961; Levin & Cohn, 1968; Levin & Turner, 1968; Morton, 1964a, 1964b) that "subjects tend to read in phrase units" (Levin & Turner, 1968, p.208), and Morton's view that reading rate and EVS increase with more structured reading texts (Morton, 1964a).

In Section 3.3, offset FVS and offset EVS were regressed on word length, word frequency, and a third, more context-sensitive measure, namely whether a word occurs either in the middle or at the end of a weak/strong prosodic unit (Table 5). Both the eye and the finger appear to be affected by the same factors, with a significant interaction of word length and word position with the two tracking modalities (Figure 7). Notably, FVS is more directly affected by word length than EVS. In addition, FVS shows an overall greater variability than EVS as a function of the presence ($Punct\ Type > 0$) vs. absence ($Punct\ Type = 0$) of a punctuation mark, and less variability between weak vs. strong (implicit) prosodic units. Results are coherent with the high correlation between spoken word duration and word finger-tracking duration, while adding an important piece of information: the

finger tends to make the same pauses the voice makes at the right edge of an (implicit) prosodic unit.

In Figure 8, smaller offset FVSs (top panels) and smaller offset EVSs (bottom panels) are consistently found at the right edge of (i) syntactic chunks (left panels), (ii) weak prosodic units (centre panels) and (iii) strong prosodic units (right panels). The evidence shows that readers tend to slow down their reading pace at the end of chunked multi-word units, no matter how long these units are. Accordingly, a reader's span is not set to the maximum capacity of her phonological buffer. Rather, it appears to stretch or shrink far enough for the reader to be able to process a larger unit, articulate its words with an appropriate intonation contour, and grasp its meaning (Levin & Turner, 1968). We refer to this behaviour as *adaptive reading*.

If the adaptive reading hypothesis is correct, a text passage that is sufficiently long would allow a proficient reader to resort to a larger text-scanning window. Thus, the first word of a longer chunk is likely to be articulated later than the first word of a shorter chunk, because a reader will try to delay word articulation until the whole chunk is fixated. This is confirmed by the plots in Figure 9, where both offset FVSs (left panel) and offset EVSs (right panel) start higher in 3-token chunks than in 2-token chunks, to reach their minimum value when the final word of a chunk is read ($x = -1$). In our data the effect does not extend to 4-token chunks, where offset spans appear to flatten out in both plots, in keeping with what reported by Levin and Turner (1968). In fact, it would be somewhat inefficient for a reader to stretch her span beyond the capacity of her phonological buffer, as this would disrupt oral reading, causing a few buffered words to decay from memory before they can be articulated. Based on our data (Table 2), an average offset EVS may include little more than two fully fixated words. This confirms that fixating a word which is three tokens after the currently articulated word can be suboptimal, and can be done only in particular conditions (e.g. when the chunk is highly predictable, or made up out of short, frequent words).

It is noteworthy that the voice span appears to be affected not only by the position of the word in the chunk, but also by the chunk length. To illustrate, in the left panel of Figure 9, the span for $x = -2$ is higher in 2-token chunks than in 3-token chunks and 4-token chunks. The evidence is compatible with the idea that the finger (or somewhat less prominently the eye) slows down its pace as soon as the reader starts fixating the first token of a chunk, and it does so more or less abruptly depending on the chunk length. Accordingly, the span associated with a word at position $x = -2$ in a 4-token chunk is smaller than the span of a word at the same position in a 3-token or 2-token chunk, simply because the finger has started slowing down its tracking pace a little longer. This evidence lends further support to the idea that structural features of the text, including the

internal structure of a chunk or a prosodic unit, play a role in modulating finger and eye movements during reading.

Finally, Figure 10 sheds light on the relative contribution of syntactic chunks and prosodic units to the timing of finger and eye movements. In the upper half of the figure, plots depict decreasing levels of FVS (left panel) and EVS (right panel) *within* syntactic chunks, by controlling for the presence of a weak vs. strong punctuation mark immediately after the chunk. Both FVS and EVS appear to go down consistently across all unit types. However, offset FVS exhibits a steeper descent than offset EVS at the right edge of chunks that are followed by a punctuation mark (see model S5 in the *Supplementary Materials*). We interpret this evidence as showing that implicit prosodic units (i.e. text chunks followed by a punctuation mark) affect finger movements more than syntactic units do, in keeping with the observation that the finger keeps the pace of articulation. This is confirmed by the plots in the lower half of Figure 10, which show voice span variation *between* consecutive chunks. The finger shows considerably longer tracking duration between two chunks that are separated by a punctuation mark, whether weak or strong. In contrast, we observe (i) longer eye fixations across sentence boundaries, (ii) significantly shorter fixations across weaker intonational boundaries, and (iii) only marginal variations across chunk boundaries when no punctuation marks intervene.

Integrating the two hypotheses

In finger-point reading, eye, finger and articulation movements are highly asynchronous. The length of their time spans is a dynamic function of a variety of interacting factors: (i) duration of a word fixation (based on a word's length, frequency, typicality and predictability), (ii) automaticity of print-to-sound conversion, and (iii) capacity of a reader's working memory buffer (which is, in turn, a function of a reader's articulatory rate). Factor (i) contributes to the first component of the onset span (gaze duration), while factors (ii) and (iii) are determinants of the offset span. Although gaze duration and articulation rate do not necessarily correlate, short gaze durations and high articulation rates independently contribute to longer onset spans. The phonological buffer hypothesis holds the view that such a complex, dynamic interaction is fundamentally due the different functional yield of eye movements and articulation movements in reading. Given a certain time window, the eye can fixate more words than the voice can articulate. The phonological buffer compensates for this functional asynchrony by offering a memory stack of limited temporal capacity where words are temporarily maintained until they are read out loud. The larger the buffer's capacity, the longer the offset span. The hypothesis predicts that readers tend to

keep their span fairly constant during text reading, with relatively small variations in the duration of the offset span, mainly attributable to individual factors (e.g. a reader's working memory capacity). In fact, a careful analysis of the offset span distribution in adults' reading of connected texts shows a large intraindividual variability and a wide modulation of the offset span.

First, we observed that FVS is systematically reset at the right boundary of a syntactic/prosodic text unit, where the finger typically slows down, waiting for the voice to catch up and (often) finish word articulation *before* the word is fully finger-tracked (see negative FVS values at $x = -1$, Figure 10). The same obtains for EVS. The eye jumps from one fixated word to another, to then wait for the voice (and the finger) to catch up at the right boundary of a strong prosodic unit (typically a sentence). Unlike what happens with the finger, a word's articulation rarely starts before the fixation on the word is finished. This dynamic explains why adults' finger-tracking times and fixation durations appear to correlate perfectly at the sentence level ($r = 0.99$), but less strongly at the chunk ($r = 0.79$) and word level ($r = 0.62$) (Crepaldi et al., 2022). Unlike the finger, which keeps the pace of voice articulation and continuously slides across all text words, eye movements appear to be significantly less affected by prosodic units that are weaker than a sentence, and typically skip a few words. Accordingly, while the eye is scanning the text and the voice articulates foveally and parafoveally fixated words, the finger plays the role of a visual marker, pointing to where the voice is on the page at each time tick. Through their fine coordination, these three interlocked time series provide one another with mutual feedback signals. The voice beats the reading *tempo* that the finger must keep to maintain constant its distance from the voice. The eye provides information as to where the voice is more likely to slow down or pause, causing the finger to slow down too. Finally the finger provides a dynamically updated visual marker of the voice position, which is available to the eye as the moving target of a potential, regressive saccade, when something goes wrong through the offset voice span.

If this scenario is on the right track, then the phonological buffer hypothesis puts an upper limit on the eye's ability to look ahead while reading a text, and accounts for the ways this limit may affect eye fixations and eye movements. The adaptive reading hypothesis complements the phonological buffer account to offer an explanation of the flexibility of EVS and FVS, which depends on the linguistic structure of a reading text. Following Buswell (1920) and Levin and Turner (1968), we argue that this strategy is functional to fluent oral reading, as it allows the reader to optimally plan articulation by buffering lexical units into larger meaningful intonation units. Reading fluency has been found to directly relate to text processing and understanding, with appropriate intonational contours and rhythmical patterns providing critical cues to meaningful syntactic

structures (Breen, 2014; Breen et al., 2016), even in silent reading (Kentner & Vasishth, 2016). Our data allowed us to empirically relate reading fluency to delayed articulation: reading more words in parallel is good for online text processing, expressive oral reading and, ultimately, reading comprehension.

5. Concluding remarks

In oral reading, the eye starts scanning the text before voice articulation and finger-tracking set in. How far ahead the eye goes is a function of several factors, including the reader's articulatory rate and phonological buffer, the length and frequency of a fixated word, the larger meaningful text units where words occur. From the perspective we entertained here, the offset voice span is thus understood as the outcome of an optimally adaptive viewing strategy that is interactively modulated by the subject's reading skills, the tracking mode (ocular or tactile) and the linguistic features of a written text: (i) spans may vary from one reader to another, (ii) they tend to get larger when the text contains larger intonational and structural units, and (iii) they are modulated differently depending on the tracking mode, with the eye being mainly sensitive to longer text units, and the finger to shorter ones. Needless to say, this strategy makes reading more effortful, while straining a reader's working memory at the risk of disrupting articulation, and takes a lot of reading skills to be effectively implemented. Nonetheless, according to what we named the adaptive reading hypothesis, this strategy is an essential component of reading proficiency, whereby a reader is able to process, comprehend and express what she is reading while she is reading, by adjusting her span online to the linguistic structure of the text.

This view should not be confused with the large reading literature reporting *inverted* parafoveal-on-foveal effects (e.g., Drieghe, Rayner, & Pollatsek, 2008; Hyönä & Bertram, 2004; Inhoff, Starr, & Shindler, 2000). In this literature, readers are observed to speed up or skip their focal fixations to be able to devote more processing time and resources to what is anticipated as a more demanding input. In contrast, looking ahead for the right edge of a complex (intonational or structural) unit responds to the strategic need to maximise information intake and optimise reading (as opposed to optimise processing resources). Our view is more akin to Lim et al.'s (2019) view that the Eye-Hand Span, i.e. the distance between a pianist's fixation of a note in a score, and the execution of the note in score sight-reading, is a strategy for proficient piano performance. If, as we argued here, the span can be adjusted to structural features of the text, we can expect some specific aspects of textual complexity (e.g. larger phrasal chunks) to actu-

ally lengthen the span, as observed by Huovinen, Ylitalo, and Puurtinen (2018) for music reading.

We believe such a coherent pattern of results to be instrumental for reading research in a number of respects. It is generally assumed (Silva et al., 2016) that easier texts allow a proficient reader to better leverage parallel processing, reading automaticity and, ultimately, voice span duration. However, our evidence suggests that spans are task- and text-dependent: unlike word lists or sentences presented in isolation, a structurally richer text may in fact prompt proficient readers to scan *larger* text chunks, and ultimately *lengthen* their voice spans. Investigating what lexical and structural factors in a written text may dynamically affect the pace of natural reading is, in our view, an important direction for future research, paving the way to a converging perspective between cognitive (Pollatsek & Treiman, 2015), computational (Reichle, 2021) and educational (Grabe & Stoller, 2019) approaches to reading research.

If the finger pace is a rhythmic proxy for articulation rate, we can expect finger-tracking data of silent reading sessions to provide essential information for the Implicit Prosody Hypothesis (Breen, 2014). The hypothesis claims that readers, when reading silently, activate prosodic representations of the text that are similar to those they would produce when reading the text aloud. Hence, we should expect the correlation between spoken word duration and finger-tracking times to remain strong when finger-tracking times are measured in silent text reading, suggesting that the finger keeps the pace of word articulation even when reading is subvocal. In addition, these representations are expected to affect the reader's interpretation of the text. Accordingly, we should be able to observe that silent readers slow down their finger-tracking pace at the right edge of an implicit prosodic unit. Conversely, failure to produce a slow-down effect where the effect is expected could be associated with problems in text interpretation (Kentner & Vasishth, 2016). The capacity of the phonological buffer is known to grow developmentally, together with the articulatory rate and the lexical competence of a speaker (e.g. Gathercole, 2006, and Vihman, 2022 for a recent overview). We thus expect the voice span to get larger as decoding automaticity, orthographic lexical competence and working memory capacity increase developmentally.

On a less positive note, we must be cognizant that finger-tracking *per se* does not provide direct evidence of a reader's processing behaviour. Hence, our data should always be examined and scrutinized meticulously. It is possible, after all, that a subject might just pretend to read a text silently, while at the same time finger-pointing to the text in an apparently natural way. This risk exists and cannot be ruled out completely. Nonetheless it can be minimized by checking if the reader grasped the content of the text with a few comprehension questions. In addition, the more we know about how finger movements relate to lexical and structural features of a reading text, the better we can tell genuine




finger-tracking data from a possible pretence. We should also bear in mind that finger-tracking data are not a substitute for eye-tracking data. Here, we showed that finger movements are in step with voice articulation, guided by the written information captured by the eye. This accounts for a strong voice-finger correlation, and makes finger-tracking data somewhat complementary to eye-tracking evidence. In fact, the former can shed light on aspects of reading (e.g. the influence of subvocal articulation and the role of implicit prosody in silent reading) that elude classical eye-tracking protocols. In the end, we strongly believe that multi-modal reading data, such as those presented here, promise to speed up progress in reading research and teaching by offering richer, more robust and more controlled behavioural evidence.

Finally, we would like to emphasize the practical potential of finger-tracking for reading research. The technology turns out to be considerably less expensive, brittle and noisy than eye-tracking. In prospect, an ecological and robust protocol based on finger-tracking will allow for massive data to be harvested with an ordinary tablet, through an ecological reading task to be carried out in ordinary, familiar contexts (e.g., either in classroom or at home). This promises to provide evidence at scale of more and less typical developmental patterns of reading skills in the first years of primary school, when reading difficulties are more critical but manifest less clearly, while offering a suitable benchmark for continual assessment of reading proficiency.

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), and the *ReadGround* grant, from the Italian National Research Council (CNR), are gratefully acknowledged. Alessandro Lento is a PhD student enrolled in the *National PhD in Artificial Intelligence*, XXXVII cycle, course on Health and Life sciences, organized by Università Campus Bio-Medico of Rome.

References

-  Abney, S. P. (1991). Parsing by chunks. In R. C. Berwick (Ed.), *Principle-based parsing* (pp. 257–278). Springer.
- Attardi, G. (2006). Experiments with a multilanguage non-projective dependency parser. In *The 10th International Conference on Computational Natural Language Learning (CoNLL-X2006)* (pp. 166–170). New York.
-  Baddeley, A. (2007). *Working memory, thought, and action* (Vol. 45). OUP Oxford.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... others (2009). *Package 'lme4'*. r-project.org.
-  Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, 10 (4), 433–436.

- doi Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, 8 (2), 37–50.
- doi Breen, M., Kaswer, L., Van Dyke, J.A., Krivokapić, J., & Landi, N. (2016). Imitated prosodic fluency predicts reading comprehension ability in good and poor high school readers. *Frontiers in psychology*, 7, 1026.
- doi Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, 1 (1).
- Buswell, G. T. (1920). An experimental study of the eye-voice span in reading. *Journal of Educational Psychology*, 4 (12), 217–227.
- doi Buswell, G. T. (1921). The relationship between eye-perception and voice-response in reading. *Journal of Educational Psychology*, 12 (4), 217–227.
- 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.
- doi Cohen, J.D., Servan-Schreiber, D., & McClelland, J.L. (1992). A parallel distributed processing approach to automaticity. *The American journal of psychology*, 239–269.
- doi Crepaldi, D., Ferro, M., Marzi, C., Nadalini, A., Pirrelli, V., & Taxitari, L. (2022). Finger movements and eye movements during adults' silent and oral reading. In R. Levie, A. Bar-On, O. Ashkenazi, E. Dattner, & G. Brandes (Eds.), *Developing Language and Literacy: Studies in Honor of Dorit Diskin Ravid* (pp. 443–471). Springer International Publishing.
- Dell'Orletta, F., Federico, M., Lenci, A., Montemagni, S., & Pirrelli, V. (2007). Maximum Entropy for Italian PoS Tagging. *Intelligenza Artificiale*, 4 (2).
- Dell'Orletta, F., Montemagni, S., & Venturi, G. (2011). READ-IT: Assessing readability of Italian texts with a view to text simplification. In *Proceedings of the second workshop on Speech and Language Processing for Assistive Technologies* (pp. 73–83).
- doi De Luca, M., Pontillo, M., Primativo, S., Spinelli, D., & Zoccolotti, P. (2013). The eye-voice lead during oral reading in developmental dyslexia. *Frontiers in human neuroscience*, 7, 696.
- doi Drieghe, D., Rayner, K., & Pollatsek, A. (2008). Mislocated fixations can account for parafoveal-on-foveal effects in eye movements during reading. *Quarterly Journal of Experimental Psychology*, 61 (8), 1239–1249.
- Federici, S., Montemagni, S., & Pirrelli, V. (1996). Shallow Parsing and Text Chunking: A view on Underspecification in Syntax. *Proceedings of ESSLLI'96 Workshop on Robust Parsing*, 35–44.
- doi Ferro, M., Cappa, C., Giulivi, S., Marzi, C., Nahli, O., Cardillo, F.A., & Pirrelli, V. (2018). Readlet: Reading for understanding. In *Proceedings of 5th IEEE Congress on Information Science & Technology (IEEE CiST'18)*. Marrakech, Morocco.
- doi Fuchs, L.S., Fuchs, D., Hosp, M.K., & Jenkins, J.R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific studies of reading*, 5(3), 239–256.
- doi Gathercole, S.E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied psycholinguistics*, 27 (4), 513–543.
- doi Grabe, W., & Stoller, F.L. (2019). *Teaching and researching reading* (3rd ed.). Routledge.
- doi Hughes, B., McClelland, A., & Henare, D. (2014). On the nonsmooth, nonconstant velocity of braille reading and reversals. *Scientific Studies of Reading*, 18 (2), 94–113.

- doi** Huovinen, E., Ylitalo, A.-K., & Puurtinen, M. (2018). Early attraction in temporally controlled sight reading of music. *Journal of Eye Movement Research*, 11 (2), 1–30.
- doi** Hyönä, J., & Bertram, R. (2004). Do frequency characteristics of nonfixated words influence the processing of fixated words during reading? *European Journal of Cognitive Psychology*, 16 (1–2), 104–127.
- doi** Inhoff, A. W., Solomon, M., Radach, R., & Seymour, B.A. (2011). Temporal dynamics of the eye-voice span and eye movement control during oral reading. *Journal of Cognitive Psychology*, 23 (5), 543–558.
- doi** Inhoff, A. W., Starr, M., & Shindler, K.L. (2000). Is the processing of words during eye fixations in reading strictly serial? *Perception & Psychophysics*, 62 (7), 1474–1484.
- doi** Kentner, G., & Vasishth, S. (2016). Prosodic focus marking in silent reading: effects of discourse context and rhythm. *Frontiers in Psychology*, 7, 319.
- doi** 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.
- doi** Kumle, L., Vö, M.L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in r. *Behavior research methods*, 53 (6), 2528–2543.
- doi** Laubrock, J., & Kliegl, R. (2015). The eye-voice span during reading aloud. *Frontiers in psychology*, 6, 1432.
- doi** Lawson, E.A. (1961). A note on the influence of different orders of approximation to the english language upon eye-voice span. *Quarterly Journal of Experimental Psychology*, 13 (1), 53–55.
- Lenci, A., Montemagni, S., & Pirrelli, V. (2003). “Chunk-it”. An Italian Shallow Parser for Robust Syntactic Annotation. *Linguistica Computazionale, XVIII–XIX*, 353–386.
- Levin, H., & Cohn, J.A. (1968). Effects of instruction on the eye-voice span. In H. Levin, E. J. Gibson, & J. J. Gibson (Eds.), *The analysis of reading skills: A program of basic and applied research. final report* (pp. 254–283). Ithaca, New York: Cornell University.
- Levin, H., & Turner, E.A. (1968). Sentence structure and the eye-voice span. In H. Levin, E. J. Gibson, & J. J. Gibson (Eds.), *The analysis of reading skills: A program of basic and applied research. final report* (pp. 196–220). Ithaca, New York: Cornell University.
- doi** Lim, Y., Park, J.M., Rhyu, S.-Y., Chung, C.K., Kim, Y., & Yi, S.W. (2019). Eye-hand span is not an indicator of but a strategy for proficient sight-reading in piano performance. *Scientific Reports*, 9 (1), 1–11.
- 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.
- Lyding, V., Stemle, E., Borghetti, C., Brunello, M., Castagnoli, S., Dell’Orletta, F., ... Pirrelli, V. (2014). The Paise corpus of Italian web texts. In F. Bildhauer & R. Schäfer (Eds.), (p. 36–43). Gothenburg, Sweden: Association for Computational Linguistics.
- Maffei, L. (2018). *Elogio della parola*. Bologna: il Mulino.
- doi** 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). Bologna.
- doi** 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.

- doi Mesmer, H.A.E., & Williams, T.O. (2015). Examining the role of syllable awareness in a model of concept of word: Findings from preschoolers. *Reading Research Quarterly*, 50 (4), 483–497.
- doi Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual review of psychology*, 67, 263–287.
- doi Morton, J. (1964a). The effects of context upon speed of reading, eye movements and eye-voice span. *Quarterly Journal of Experimental Psychology*, 16 (4), 340–354.
- doi Morton, J. (1964b). A model for continuous language behaviour. *Language and Speech*, 7(1), 40–70.
- doi Nonaka, T., Ito, K., & Stoffregen, T.A. (2021). Structure of variability in scanning movement predicts braille reading performance in children. *Scientific reports*, 11 (1), 1–12.
- doi Paap, K.R., & Noel, R.W. (1991). Dual-route models of print to sound: Still a good horse race. *Psychological research*, 53 (1), 13–24.
- 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).
- doi Pollatsek, A., & Treiman, R. (2015). *The oxford handbook of reading*. Oxford University Press.
- Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., ... Vesely, K. (2011). The Kaldi Speech Recognition Toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society. (IEEE Catalog No.: CFP11SRW-USB)
- doi Protopapas, A., Altani, A., & Georgiou, G.K. (2013). Development of serial processing in reading and rapid naming. *Journal of Experimental Child Psychology*, 116 (4), 914–929.
- R Core Team. (2023). *R: A language and environment for statistical computing*. <https://www.R-project.org/>. Vienna, Austria: R Foundation for Statistical Computing.
- doi Reichle, E.D. (2021). *Computational models of reading: A handbook*. Oxford University Press.
- doi Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26 (1), 43–49.
- doi Schilling, S.G., Carlisle, J.F., Scott, S.E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal*, 107 (5), 429–448.
- Shmyrev, N.V., & Vosk Core Team. (2020). *Vosk Speech Recognition Toolkit: Offline speech recognition API for Android, iOS, Raspberry Pi and servers with Python, Java, C# and Node*. <https://github.com/alphacep/vosk-api>. GitHub repository.
- doi Silva, S., Reis, A., Casaca, L., Petersson, K.M., & Faísca, L. (2016). When the Eyes no longer lead: Familiarity and Length Effects on Eye-Voice Span. *Frontiers in Psychology*, 7.
- 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)*. Agadir, Morocco.
- doi Uhry, J.K. (1999). Invented spelling in kindergarten: The relationship with finger-point reading. *Reading and Writing*, 11, 441–464.
- doi 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.



Vihman, M.M. (2022). The developmental origins of phonological memory. *Psychological review*, 129 (6), 1495–1508.

Address for correspondence

Vito Pirrelli
Italian National Research Council
Institute for Computational Linguistics “A. Zampolli”
via Giuseppe Moruzzi 1
56124 Pisa
Italy
vito.pirrelli@ilc.cnr.it

Co-author information

Andrea Nadalini
Italian National Research Council
Institute for Computational Linguistics “A.
Zampolli”
andrea.nadalini@ilc.cnr.it

Claudia Marzi
Italian National Research Council
Institute for Computational Linguistics “A.
Zampolli”
claudia.marzi@ilc.cnr.it

Marcello Ferro
Italian National Research Council
Institute for Computational Linguistics “A.
Zampolli”
marcello.ferro@ilc.cnr.it

Loukia Taxitari
Department of Psychology
Neapolis University
l.taxitari@nup.ac.cy

Alessandro Lento
Biomedical Campus University
alessandro.lento@unicampus.it

Davide Crepaldi
Scuola Internazionale Superiore di Studi
Avanzati
dcrepaldi@sissa.it

Publication history

Date received: 31 August 2023
Date accepted: 23 January 2024
Published online: 12 March 2024