Why Do Skimmers Perform Better with Grammar-Preserving Text Saliency Modulation (GP-TSM)? Evidence from an Eye Tracking Study

ANONYMOUS AUTHOR(S)

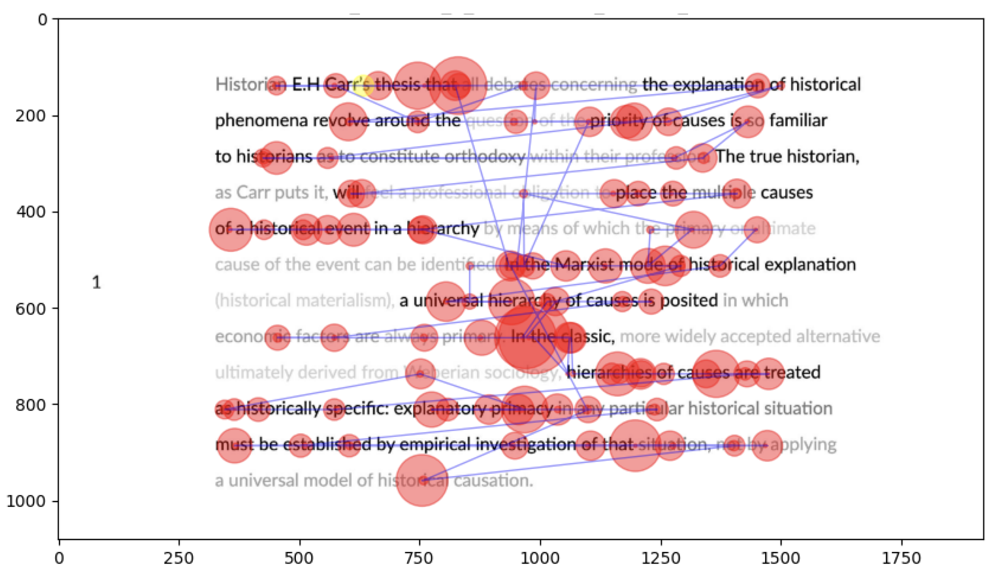


Fig. 1. An example of a study participant's eye movements overlaid on the GP-TSM rendering of a paragraph from the GRE Reading Comprehension test. The red circles represent eye fixations. The bigger the circle, the longer the fixation. The purple lines represent saccades.

Recent work has introduced Grammar-Preserving Text Saliency Modulation (GP-TSM), a novel text rendering technique that has been shown to enhance reading efficiency and experience. However, the mechanism through which GP-TSM augments reading remains unclear. In this work, we conducted a within-subjects eye-tracking user study with 24 participants to understand how GP-TSM influences the reading experience. We found that participants closely adhered to GP-TSM's visual cues, exhibiting gaze behavior that is distinct from that observed with the typical reading interface. From this gaze pattern, we highlight how GP-TSM leads to more efficient and coherent skimming while allowing the revisiting of skipped details.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: eye-tracking, text visualization, human-AI interaction

ACM Reference Format:

Anonymous Author(s). 2024. Why Do Skimmers Perform Better with Grammar-Preserving Text Saliency Modulation (GP-TSM)? Evidence from an Eye Tracking Study. 1, 1 (March 2024), 11 pages. https://doi.org/XXXXXXX.XXXXXXX

Manuscript submitted to ACM

1 INTRODUCTION

In the digital age, people have access to a vast array of digital text, ranging from Kindle books to online news articles. The flexibility of digital text rendering makes it manipulatable programmatically. This opens up an entire design space for reading assistance.

Recent work has introduced a novel text rendering technique called GP-TSM (Grammar-Preserving Text Saliency Modulation) [7] that de-emphasizes recursive levels of detail beyond the core meaning of a passage—while always preserving the grammaticality at every level. De-emphasis is operationalized as text opacity; to preserve *AI-resilience*, GP-TSM's lowest opacity level is constrained to still be legible, in case the AI mis-parses a sentence or de-emphasizes details that are actually important to the reader.

These prior studies have demonstrated that GP-TSM results in better reading experience and higher reading efficiency (compared to a typical reading interface) when it comes to reading comprehension tasks on standardized tests. However, the mechanism by which GP-TSM achieves this effect is yet unknown. We hypothesize that that this technique directs readers' focus to key grammatical subsets of sentences, allowing readers to dynamically adjust their engagement level without losing information or interrupting flow.

In this study, we seek to bridge the gap in understanding how GP-TSM shapes the reading experience through a within-subjects eye-tracking user study (N=24). Eye-tracking offers an unobtrusive and fine-grained method for elucidating readers' visual attention change [21] and cognitive processing of text [12] in real-time. By analyzing where, how long, and in what sequence readers fixate on different text elements, we gain invaluable insights into their reading behaviors, strategies, and the extent to which GP-TSM influences reading patterns, attention distribution, and overall engagement with the text. More specifically, We aim to answer the following research questions:

- RQ1: *How do readers' gaze behavior vary between reading on the GP-TSM interface and a typical reading interface?*
- RQ2: *To what extent is there a correspondence between readers' gaze behavior and GP-TSM's cues (e.g., longer fixation on darker text segments)?*
- RQ3: *In what contexts do readers' gaze behavior deviates from GP-TSM's cues (e.g., longer fixation on lighter text segments)?*

Our findings suggest that readers are greatly influenced by the cues from GP-TSM, leading to a reading behavior that diverges substantially from a typical reading interface and closely aligns with the signals of GP-TSM.

2 RELATED WORK

2.1 Reading

Reading is recognized as a cognitively demanding task, requiring significant mental effort and attention [11, 15]. Studies on speed reading [9, 22] have proposed various techniques and strategies to enhance reading speed, but these skills take a considerable amount of time to practice and employ effectively [13, 17]. Furthermore, prior research suggests that there is an inevitable trade-off between reading speed and comprehension, showing that the comprehension of both essential and non-essential information from a text is equally diminished with an increase in reading speed [5, 17]. When people do skim, the primary predictor of what words are skipped is the length of the word, with processing ease (e.g., the frequency of the word and its predictability in the sentence) having a smaller but still significant effect [1]. Without guidance, skimmers tend to have difficulty staying focused, miss key information, and lack confidence in their comprehension of the text [3, 23].

In response to the challenges of reading and skimming, GP-TSM [7], the subject of our eye-tracking studies, encodes multiple levels of LLM-generated sentence compression within the text rendering by modulating opacity. In this way, it provides readers with immediate, contextually relevant summaries if they choose to skim the text, while making all the details still available, if readers wish to read closely, or dynamically and seamlessly transition between the two modes.

2.2 Eye Movement in Reading and Skimming

There have been numerous studies that focused on eye movement during reading and skimming, and the links between eye movement and cognition have been well established [4, 14]. During reading, eyes make fast jumps between words called *saccades*, and the duration of time the eyes stay focused on words is called a *fixation*. Visual features that aid in word identification are extracted only during fixations, while no meaningful information is extracted during saccades [14].

Although eye movement differs between reading and skimming [4, 20], the ultimate goal of both activities is comprehension. In reading, readers often direct their attention sequentially through the text, from top to bottom and from left to right in a language like English. On the other hand, skimming involves applying a selective reading strategy where eye movement is allocated in a way that skips words, sentences, paragraphs or entire pages of text [4].

Eye movement during reading in general is influenced by many factors including linguistic factors such as the frequency, predictability, and the length of words [14]. Readers often make longer fixations on longer, less-frequent, and unpredictable words, and it is typical for readers to skip short, predictable, or highly-frequent words [14, 18]. In fact it has been reported that a third of words are skipped in reading English [14, 16, 20]. Eye movement studies on skimming show that comprehension is often hindered when skippable words are omitted [6, 17]. At the same time, skilled readers were shown to adopt a “riskier” reading strategy where they skip more words in the text [16].

3 EYE-TRACKING USER STUDY

3.1 Study Design

To answer our research questions, we rely on an IRB-approved within-subject user study to present each participant with passages in GP-TSM interface and in plain text. The presentation is counterbalanced between the two condition over three passages, each consisting of 4-5 paragraphs. The content of the paragraphs is identical across conditions; only the rendering is different. Figure 4 in Appendix A presents screenshots of the GP-TSM and Control conditions in the eye-tracking user study, each displaying the same paragraph from the same passage.

3.2 Participants

The 24 participants in our study were college students with normal or corrected to normal vision. All participants were above the age of 18, and they were native English speakers or proficient second language speakers. Participation was voluntary, and participants were awarded a $10 gift card after the 30 to 40 minute experiment. Eye tracking data from three participants were partially or completely omitted due to problems with recording eye tracking data with glasses or unintentionally skipped paragraphs resulting in incomplete eye-movement recordings.

3.3 Procedure

Before the experiment began, participants were introduced to the GP-TSM interface through an example passage separate from those in the trials. The differences between plain and GP-TSM interfaces were explained, along with a summary of the project and their role. Once the participant had been given the brief, they were given an informed consent form and had the opportunity to address any lingering questions.

An eye tracking chin-rest was used to make sure that eye movement can be captured accurately during the experiment, and a calibration of the eye tracker was conducted once the height of the chair and chin-rest were adjusted to the participant. The trials consisted of three separate passages, all sourced from the Graduate Record Examinations [1], with three randomized conditions, each appearing the same amount of times in the study. The conditions included the type of interface, the order in which the passages were shown, and when comprehension questions were displayed (before reading, after reading, or not at all). The comprehension questions were multiple-choice, meant to measure attentiveness and overall comprehension.

Sequentially, paragraph from the passages were displayed on the monitor screen. The participant had full control of the pace of their reading and could move to the next paragraph once they were satisfied by pressing the space bar. For trials that included multiple-choice questions, readers answered by pressing the letter on the keyboard associated with their choice on the screen. Once participants completed their readings of all three passages, they were guided through a series of NASA TLX survey questions: One set for the GP-TSM interface and one for the plain interface. After that, participants reported their subjective opinions on using the GP-TSM interface through Likert style questions.

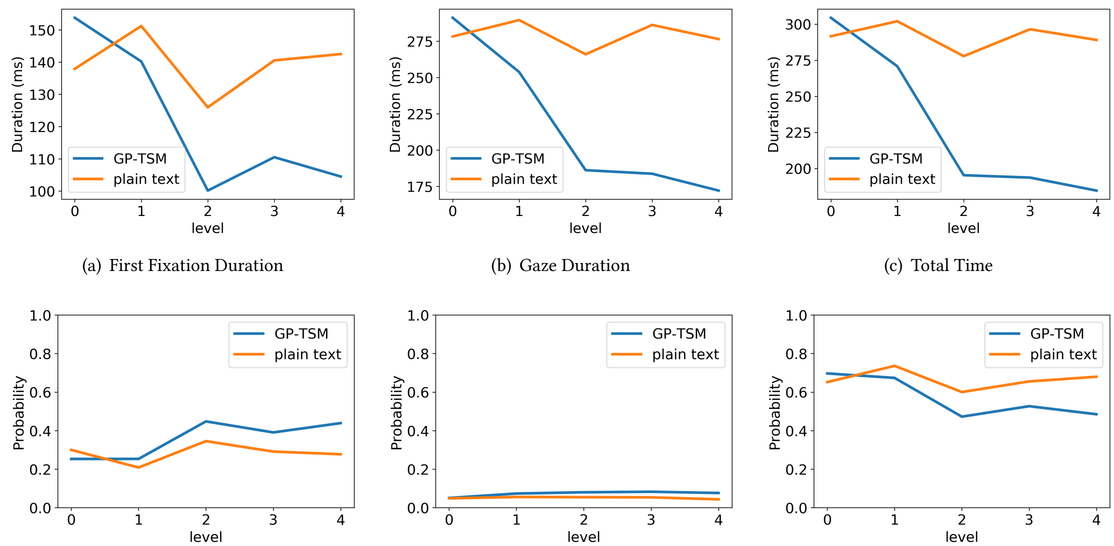
3.4 Measures

The eye movement of the participant was tracked and recorded with an EyeLink 1000 Plus (from SR Research) eye tracker, with a sampling frequency of 1000 samples per second. The screen resolution was set to 1920x1080 px. The eye tracker records fixation durations and their coordinates on the screen, and we match these coordinates to the individual words presented in the stimuli text to get each fixation duration on each word. To correct fixation drift where fixations might move away from corresponding words due to participants moving during the experiment or calibration decay, we apply an automated drift correction algorithm named Warp [2]. From fixation durations we can calculate eye movement metrics that give us insights on reading behaviour and cognition [19]. We use three duration metrics and three probability metrics to compare eye movement when reading text in the GP-TSM interface and the plain-text interface:

- First-Fixation Duration (FFD): The duration of the first fixation on a word (in millisecond).
- Gaze Duration (GD): The duration of all the fixations on a word before moving to the next word (in millisecond).
- Total Time: The sum of all fixation durations on a word, including fixations from regressions (in millisecond).
- probability of skipping (PrS): The probability of a word in getting no fixation (getting skipped during reading).
- probability of making exactly one fixation (Pr1): The probability of a word in getting exactly one fixation not more.
- probability of making two or more fixations (Pr2): The probability of a word in getting two or more fixation.

The connection between eye movement and cognition has been well established in cognitive psychology [14], and in our context First-Fixation Duration is a good representative of word identification or the retrieval of word sound and meaning from memory (lexical access) [14, 18]. Gaze Duration and Total-Time give us an indication of the time taken to complete word identification and comprehension, where longer and repeated fixations are correlated with more processing [10, 20]. In this experiment, the chosen metrics give us the ability to measure the potential differences in

[1]Passages and questions used are from Educational Testing Service (ETS) GRE Practice Tests.



(a) First Fixation Duration (b) Gaze Duration (c) Total Time

(d) Probability of Skipping (e) Probability of exactly one fixation (f) Probability of two or more fixations

Fig. 2. Mean eye movement metrics across Highlighting levels for GP-TSM text and plain text.

reading behaviour as a result of using GP-TSM interface compared to plain text. These metrics could uncover differences during the early word identification stage in addition to latent comprehension processes.

3.5 Quantitative Eye Movement Analysis

Our first and second research questions are concerned with differences in reading behaviour on the GP-TSM interface and the plain text interface. In answering the two question, we attempt to compare the eye movement metrics of GP-TSM readers and plain-text readers across the five highlighting levels described earlier. By making this comparison, we can make an observation on whether there are any differences in reading behaviour between the two interfaces, and also we can verify whether readers were guided by GP-TSM highlighting cues. We hypothesize that this potential correspondence between GP-TSM highlighting and eye movement is reflected in longer and more frequent fixations on darker text (e.g. level 0), while lighter text (e.g. level 4) is more likely to be skipped. Although plain-text trials are not highlighted, we use highlighting level as a way to group the same words when they appeared in the GP-TSM interface and plain-text interface as control.

In comparing GP-TSM reading to plain-text reading we focus on six metrics, Figure 2 shows three duration metrics and three probability metrics across five highlighting levels. First-Fixation duration describes the duration of the first fixation on a word, and in Figure 2 (a) we observe that the mean duration of the first fixation was higher with darker words (level 0) compared to lighter words (level 4). In addition, the slope of the GP-TSM line suggests a duration gradient that correlates to highlighting level. On the other hand, the plain-text line appears invariant. The same pattern is observed in Gaze Duration and Total Time metrics, in fact the mean difference between level 0 and level 4 is greater than 100 ms per word with the GP-TSM interface.

In regards to probability metrics, Figure 2 (d) shows that the mean probability of skipping a word with dark highlighting (level 0) was lower than lighter word (level 4) in the GP-TSM interface, and the trend is gradual across levels. Words in the plain-text interface appear to have the same probability of being skipped across the five levels. In addition, the figure shows that GP-TSM readers were more likely to skip light words in the GP-TSM interface than skipping the same words in the plain interface.

The probability of making exactly one fixation on a word appears very low in both interfaces in Figure 2 (c), yet the GP-TSM line shows slightly higher likelihood of making only one fixation on lighter text. At the same time, the probability of making two or more fixations in Figure 2 (f) shows that darker words (level 0) are more likely to receive two or more fixations in the GP-TSM interface compared to lighter words (level 1). Overall, these results suggest that the eye movement of GP-TSM readers were influenced by the level of highlighting imposed on the text, and we will elaborate more on this in the discussion section.

To verify the previous observations and quantify their size, a linear regression was conducted to examine the relationship between eye movement metrics and word highlighting level. In the GP-TSM interface highlighting level significantly predicted mean Gaze Duration and mean Total Time, Table 1 in Appendix B shows the details of the models and the strong relationship ($R2 > 0.7$). All metrics show a strong relationship, yet not all p-values reached significance. With mean duration metrics, R2 values show a strong negative correlation between mean fixation duration and highlighting level, in other words the darker the highlighting is the longer the the duration. The same relationship is also true for the mean probability of making two or more fixations, where darker words are more likely to receive repeated fixations. On the other hand the mean probability of skipping and the mean probability of making only one fixation on a word are positively correlated with highlighting level, which means that lighter tokens are more likely to be skipped and more likely to receive only one fixation.

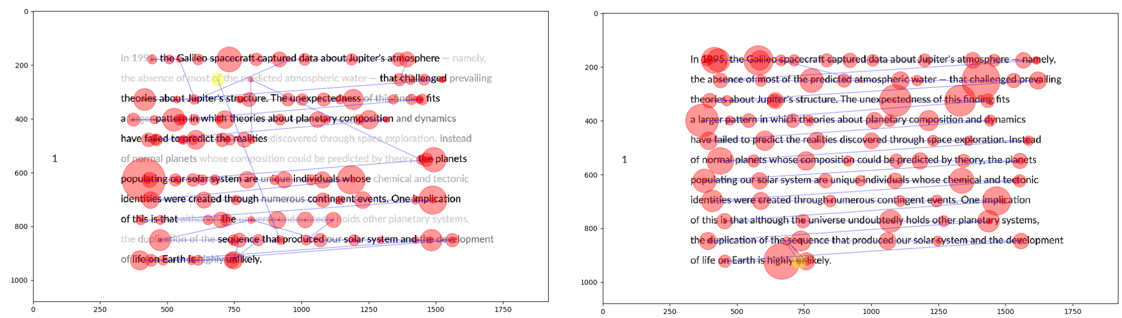
Linear regression was conducted on the plain text metrics to examine the relationship between eye movement metrics and word highlighting level as control. Table 2 shows eye movement metrics means the details of the models, and it was found that the relationships between word levels and mean eye movement metrics were weak ($R2 < 0.4$) and there were no significant differences between levels.

3.6 Quantitative Survey Analysis

Quantitative data on user preference and performance serve as useful references for the eye movement analysis. Table 3 in Appendix C shows participants' answers to survey questions about their reading experience. Overall, participants rated their experience with the GP-TSM interface slightly higher than that with the Control interface. Participants found the reading comprehension task significantly less mentally demanding ($p<0.05$) and reported to have worked significantly less hard ($p<0.05$) when reading using the GP-TSM interface. This is consistent with findings from a previous study about GP-TSM [7]. In addition, participants performed better when using the GP-TSM interface, getting on average 0.71 (out of 2) questions right (SD=0.85), compared to 0.62 (out of 2) questions (SD=0.82) in the Control condition, though this result didn't achieve statistical significance.

3.7 Qualitative Eye Movement Analysis

In this part of analysis, we took a grounded-theory approach, going through each of the eye gaze overlay figures (such as Figure 3) and identifying cases where participants' gaze behavior corresponds to or deviates from GP-TSM's cues, before synthesizing the cases into higher-level themes.



(a) GP-TSM highlighted text reading.

(b) Plain text reading.

Fig. 3. Sample eye tracking data of the same paragraph in the GP-TSM interface and in plain text.

3.7.1 Gaze Behavior Differences. Overall, in the GP-TSM interface, gaze patterns are more concentrated around the darker text segments, indicating that participants focused their attention where the text is rendered with higher opacity. The GP-TSM interface also tends to result in a non-linear reading pattern with more frequent and larger saccades, suggesting that participants jumped over the lighter text to reach the darker segments directly.

The typical reading interface shows a more uniform distribution of fixations across the entire text, reflecting a more traditional reading pattern where participants likely treated all information with similar levels of importance. It also tends to result in shorter saccades, which is typical for line-by-line reading.

3.7.2 Correspondence with GP-TSM's Cues. There appears to be a clear correspondence between participants' gaze behavior and GP-TSM's cues, as evidenced by longer and larger fixations on darker text segments. In many instances, participants' saccades formed connected line segments over darker text only, indicating a reading pattern where participants read text rendered in full opacity only, without referring to the de-emphasized segments.

3.7.3 Deviations from GP-TSM's Cues. While there is a striking correspondence between participants' gaze behavior and GP-TSM's cues, instances of deviations do exist, which are typically in the form of longer fixations on lighter text segments. One type of such deviations is fixating on important logic words such as "therefore", "however", "in addition", etc. If there is a particularly long segment that is de-emphasized, participants tend to fixate on at least some of the words within that segment, possibly due to a sense of insecurity caused by skipping such a long segment. Finally, participants tend to fixate on the beginning and end of a paragraph, even though those parts are de-emphasized by GP-TSM.

4 DISCUSSION

In this section we reflect on the insights gained from our quantitative and qualitative analysis, highlighting the unique mechanism by which GP-TSM shapes reading.

4.1 Mechanism of GP-TSM

4.1.1 Efficient Reading with Reduced Cognitive Load. Previous "fast reading" approaches aimed to train readers to read more words per minute [9, 22]. GP-TSM does not boost the number of words read in a given time directly. Instead, it allows readers to read fewer words without sacrificing too much comprehension by guiding gaze movement towards the most salient and essential parts of sentences, which themselves form grammatical sentences. Our analysis verified that only reading the essential grammatical subsets of sentences is a common pattern among participants. As a result, readers' cognitive load and perceived difficulty of the reading task significantly lowered, as reported by participants.

4.1.2 Smarter choice on what to skip when skimming. Previous research on skimming shows that the selective process employed by skimmers to skip words and sentences seems to be crucial for comprehension [3]. Therefore, it is possible that skimmers are dependent on the way they allocate their attention throughout the text. To augment skimmers' judgement, GP-TSM was designed to provide information scent about what the AI believes to be more or less skippable. Eye movement visualizations demonstrate that, despite occasional exceptions, participants' gaze path largely aligned with the visual suggestions of GP-TSM. This suggests that utilizing a sophisticated interface like GP-TSM may have helped participants make better-informed decisions on what words to skip, resulting in a more coherent reading flow.

4.1.3 Easy recovery from skipping critical details. In section 2, we highlighted previous research that shows that omitting words during skimming often hinders comprehension [6, 17]. This is an important factor that explains why previous "fast reading" approaches had a negative effect on comprehension. By contrast, GP-TSM does not omit any words from the text, allowing readers to go back and read the skipped details and context if they get confused by something or simply want to know more.

The regressions we observe among GP-TSM readers provide evidence on such a sensemaking pattern where readers, after receiving GP-TSM's cues, first follow the cues to either keep reading or skip a subsequent section, before gathering enough additional detail and context that trigger them to stop skimming and (re)read closely. We further hypothesize that readers, when using GP-TSM, engage in an active reading strategy where they constantly reflect on the contextual cues, their working memory of what was just consumed, and their own reading goals. By contrast, when using the typical reading interface, readers receive information linearly, resulting in a sequential sensemaking pattern where readers process words one by one with little saccades among lines.

4.2 Individual differences

Individual differences and reading strategies play a significant role in how participants interact with the text, suggesting a need for adaptive algorithms within GP-TSM to accommodate various reader profiles. For example, the number of deviations from GP-TSM's cues varies significantly among participants, suggesting differences in individual reading strategies that do not align perfectly with the GP-TSM's design. Interestingly, instances of such deviations also co-occur with a high frequency of regressions (jumps back to previous text), suggesting a particular type of meticulous readers who may be less inclined to skip words. Other factors that could lead a user to choose to fixate longer on lighter text might include not trusting the system or not understanding what the different levels of opacity means.

4.3 Limitations and Future Work

Our study is subject to a number of limitations. First, the specific setup of our study, which required participants to wear eye-trackers while reading and displays text one paragraph at a time, may take time to get used to. Furthermore, our conclusions might have been influenced by the limited sample size and the relatively homogeneous sample of participants, which may not fully represent the diverse range of reading behaviors in the broader population. In particular, our characterization of individual differences in reading may have been limited by the small and homogeneous sample. Future evaluation might invite participants from other pools to bring in broader perspectives. In addition, further exploration into adapting GP-TSM and other reading support tools to the needs of different types of readers and reading styles also presents a promising avenue for research.

REFERENCES

[1] M. Brysbaert and F. Vitu. 1998. Word skipping: Implications for theories of eye movement control in reading. (1998), 125–147. https://doi.org/10.1016/B978-008043361-5/50007-9

[2] Jon W Carr, Valentina N Pescuma, Michele Furlan, Maria Ktori, and Davide Crepaldi. 2022. Algorithms for the automated correction of vertical drift in eye-tracking data. *Behavior Research Methods* 54, 1 (2022), 287–310.

[3] Geoffrey B Duggan and Stephen J Payne. 2006. How much do we understand when skim reading?. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*. 730–735.

[4] Geoffrey B Duggan and Stephen J Payne. 2009. Text skimming: The process and effectiveness of foraging through text under time pressure. *Journal of experimental psychology: Applied* 15, 3 (2009), 228.

[5] Mary Dyson and Mark Haselgrove. 2000. The effects of reading speed and reading patterns on the understanding of text read from screen. *Journal of research in reading* 23, 2 (2000), 210–223.

[6] Denis F Fisher and Wayne L Shebilske. 1985. There is more that meets the eye than the eyemind assumption. *Eye movements and human information processing* (1985), 149–158.

[7] Ziwei Gu, Ian Arawjo, Kenneth Li, Jonathan K. Kummerfeld, and Elena L. Glassman. 2024. An AI-Resilient Text Rendering Technique for Reading and Skimming Documents. arXiv:2401.10873 [cs.HC]

[8] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[9] Jon Haupt. 2015. The Use of a Computer-Based Reading Rate Development Program on Pre-University Intermediate Level ESL Learners' Reading Speeds. *The Reading Matrix : an International Online Journal* 15 (2015), 1–14.

[10] Jukka Hyönä. 2010. The use of eye movements in the study of multimedia learning. *Learning and Instruction* 20, 2 (2010), 172–176.

[11] Stefan Jänicke, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. 2015. On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. *EuroVis (STARs)* 2015 (2015), 83–103.

[12] Halszka Jarodzka and Saskia Brand-Gruwel. 2017. Tracking the reading eye: Towards a model of real-world reading. , 193–201 pages.

[13] Marina Klimovich, Simon P Tiffin-Richards, and Tobias Richter. 2023. Does speed-reading training work, and if so, why? Effects of speed-reading training and metacognitive training on reading speed, comprehension and eye movements. *Journal of Research in Reading* (2023).

[14] Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (1998), 372.

[15] Keith Rayner, Alexander Pollatsek, Jane Ashby, and Charles Clifton Jr. 2012. *Psychology of reading*. Psychology Press.

[16] Keith Rayner, Erik D Reichle, Michael J Stroud, Carrick C Williams, and Alexander Pollatsek. 2006. The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and aging* 21, 3 (2006), 448.

[17] Keith Rayner, Elizabeth R Schotter, Michael EJ Masson, Mary C Potter, and Rebecca Treiman. 2016. So much to read, so little time: How do we read, and can speed reading help? *Psychological Science in the Public Interest* 17, 1 (2016), 4–34.

[18] Erik D Reichle, Keith Rayner, and Alexander Pollatsek. 2003. The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences* 26, 4 (2003), 445–476.

[19] Erik D Reichle and Heather Sheridan. 2015. EZ Reader: An overview of the model and two recent applications. *The Oxford handbook of reading* (2015), 277–290.

[20] Alexander Strukelj and Diederick C Niehorster. 2018. One page of text: Eye movements during regular and thorough reading, skimming, and spell checking. *Journal of Eye Movement Research* 11, 1 (2018).

[21] Joan N Vickers. 2009. Advances in coupling perception and action: The quiet eye as a bidirectional link between gaze, attention, and action. *Progress in brain research* 174 (2009), 279–288.

[22] G. Virgili, C. Cordaro, A. Bigoni, S. Crovato, P. Cecchini, and U. Menchini. 2004. Reading acuity in children: evaluation and reliability using MNREAD charts. *Investigative ophthalmology visual science* 45 9 (2004), 3349–54. https://doi.org/10.1167/IOVS.03-1304

[23] Ji Soo Yi. 2014. Qndreview: Read 100 chi papers in 7 hours. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. 805–814.

A STUDY INTERFACES

The recent recognition of a link between increasing rates of deforestation and increasing global climatic warming has focused new attention on the ecological role of forests. Deforestation threatens the continued existence of forests, and their loss would lead to an immediate, irreversible destabilization of the climate because the destruction of forests contributes to increased atmospheric concentrations of such heat-trapping gases as carbon dioxide and therefore to the acceleration of global warming.

The recent recognition of a link between increasing rates of deforestation and increasing global climatic warming has focused new attention on the ecological role of forests. Deforestation threatens the continued existence of forests, and their loss would lead to an immediate, irreversible destabilization of the climate because the destruction of forests contributes to increased atmospheric concentrations of such heat-trapping gases as carbon dioxide and therefore to the acceleration of global warming.

Fig. 4. Example screenshots of the GP-TSM (left) and plain text (right) interfaces in the user study, displaying the same paragraph.

B LINEAR-REGRESSION MODEL OF EYE MOVEMENT METRICS

Table 1. Mean metrics in GP-TSM trials across highlighting levels. R2 is bold when the linear-regression relationship is strong, and *P* is bold when $p < 0.05$.

level	FFD	GD	TT	PrS	Pr1	Pr2
0	153	291	304	0.25	0.05	0.70
1	140	253	270	0.25	0.07	0.67
2	100	186	195	0.45	0.08	0.47
3	110	183	193	0.39	0.08	0.53
4	104	172	184	0.44	0.08	0.48
R2	**-0.85**	**-0.93**	**-0.92**	**0.82**	**0.74**	**-0.84**
p	0.06	**0.02**	**0.02**	0.08	0.14	0.07

Table 2. Mean metrics in plain text trials across highlighting levels. R2 is bold when the linear-regression relationship is strong, and *P* is bold when $p < 0.05$.

level	FFD	GD	TT	PrS	Pr1	Pr2
0	137	278	291	0.30	0.05	0.65
1	151	289	302	0.21	0.06	0.74
2	125	266	277	0.35	0.05	0.60
3	140	286	296	0.29	0.05	0.66
4	142	276	289	0.28	0.04	0.68
R2	-0.02	-0.11	-0.18	0.11	-0.36	-0.08
p	0.96	0.84	0.76	0.84	0.54	0.89

C SURVEY QUESTIONS AND RESPONSES

Question Statements	**GP-TSM**	**Control**
How would you rate your overall experience in this interface?	4.79 (1.25)	4.50 (1.44)
How mentally demanding was the task? ***[Lower is better (LIB)]***	3.29 (1.33)*	4.04 (1.54)
How physically demanding was the task? *(LIB)*	1.91 (1.21)	1.95 (1.26)
How hurried or rushed was the pace of the task? *(LIB)*	2.83 (1.30)	2.41 (1.21)
How successful do you think you were in accomplishing the task?	4.25 (1.53)	4.79 (1.17)
How hard did you have to work to accomplish your level of performance? *(LIB)*	3.00 (1.10)*	3.75 (1.11)
How insecure, discouraged, irritated, stressed, and annoyed were you during the task? *(LIB)*	2.87 (1.54)	2.58 (1.58)
I could recognize the key points in the passage.	4.91 (1.66)	4.75 (1.62)
I could recognize how the key points are supported by additional detail in the passage.	4.75 (1.59)	4.83 (1.37)
The system's choice of what to gray out and what to keep at full font weight made sense to me.	4.62 (1.43)	N/A
I think I know why certain words were lighter than others.	4.95 (1.39)	N/A
I found it helpful that certain words were lighter than others.	4.87 (1.32)	N/A
The different levels of gray helped me see the relationships between different parts of sentences.	3.95 (1.36)	N/A

Table 3. Statistics of scores reported by participants in the survey. Participants were asked to rate their agreement with statements related to their reading experience on a 7-point Likert scale from "Strongly Disagree" (1) to "Strongly Agree" (7). Questions 2 through 7 (and their scales) were adapted from the NASA Task Load Index [8]. The last 4 questions were specific to GP-TSM and were asked only after the GP-TSM condition. *"LIB"* stands for *"Lower is better."* Statistics in column 2 and 3 are presented in the form of mean (standard deviation). Statistically significant ($p < 0.05$) differences compared with Control are marked with *.