Readability of text in multiple typefaces using negative and positive polarization; how different focus of study affects reading on screens

1st André Tørlen Lønvik
Department of Design
Norwegian University of Science and Technology
Gjøvik, Norway
andretl@stud.ntnu.no

2nd Fredrik Veland
Department of Design
Norwegian University of Science and Technology
Gjøvik, Norway
fredrvel@stud.ntnu.no

3rd Petra Vymazalová
Department of media and design
FH Joanneum
Graz, Austria
petra.vymazalova@edu.fh-joanneum.at

***Abstract*—In this paper, we attempted to figure out if there is a link between different fields of study and the students' experience in reading using multiple typefaces. We touch upon the different reading methods, typefaces and reading on screens. We have shown different studies that have performed studies related to our work, and attempted to expand upon those findings. We focused mainly on students within the age-range of 19 to 30, as those were the ones available during our data gathering period. We found that Jetbrains Mono typeface has an increased time-per-word and focus overall, while the other typefaces used had a more flow-like reading within them. We also found that there was not a large difference in the time it takes to read using the different typefaces. In addition, due to the limited number of participants, we did not find any significant statistical difference for reading within the different fields of study.**

***Index Terms*—Eye tracking, color, fonts, legibility, typefaces, students**

I. Introduction

Reading could be considered the main fundamental skill required to participate in higher education, this is due to almost all information available is currently in text form. Usually in the form of long-form articles. In order to participate in higher learning, we expect a general level of literacy, as the students must be able to understand the information we have at our disposal. We know that some individuals are exceedingly quick at reading, and some tend to be on the slower side. While this does not truly affect the information retained, the method of reading does to a certain degree.

In any field related to the praxis of coding, the importance of reading every structural part of the line they are working on cannot be understated. This is due to how code works, and the placement, word and structure is of significance.

Within most other fields of study (FoS), the requirements for reading are more on the importance of remembering the information, while also reading the text without having too much pausing in between each word. Our brain only in most cases requires the first and last letter to be in the correct position to be able to understand the full word [1].

Since the structural integrity of words within the praxis of coding is of utmost importance, the need for individuals to be able to look at, understand and see the entire word is crucial. In this paper, we will attempt to look at how the different typefaces affect the reading of individuals, and if there is a significant difference in between the FoS'.

This is why we ended up with the following research questions;

- Readability of text snippets (2 paragraphs) in multiple typefaces using dark and light coloring, does different focus of study affect screen reading.
- Does the FoS affect the method students read the text and does the FoS affect how quick they read the text?
- Do visual differences in typography and color affect information retention from text, and how differently do similar fields of study perform?
- Is there a significant difference in how students perceive the text's readability with the selected fonts? How long do the students look at single words before reading the next?

II. Literature review

There are multiple studies on typeface design and its visibility, identifiability and legibility over the past years since we read an increasing number of texts on digital screens under various conditions. With the help of eye-tracking, typeface effectivity is measurable and can lead to conclusions about the influence of font on legibility process and reading speed.

An eye-tracking study [2] investigated the legibility of typefaces presented on Liquid Crystal Display (LCD) screens. It was the first study to use LCD screens and eye-tracking technology to analyze the influence of font size and typeface in reading speed. Previous studies have used eye-tracking to measure the legibility of typefaces with and without serifs on LCD screens without making the connection to reading speed. Legibility refers to the ease of transforming typographical symbols in a text into meaning [3]. The authors said that typeface and font size influence reading speed, number and duration of fixations and saccade length. The experiment showed a coherence between the increasing font size and number of fixations. Furthermore, the typeface Verdana demonstrated a higher legibility than Georgia, regardless of the font size. The effectiveness of Verdana in this legibility-experiment can be assigned to a larger x-height than standard typefaces show as well as a minimum of stroke and width variations.

A study into influence on text legibility during reading [4] proved previous studies which stated that the selection of fonts has a significant impact on reading comprehension and speed. In this study subjects read text paragraphs on video monitors as their eye movements were recorded. The passages of text were presented in different formats and five different fonts to analyze how effectively the letters can be encoded. While this experiment was mainly focused on elimination of pixilation by using ClearType which is a format designed to display clear characters on LCD monitors and comparing it to standard fonts, font differences could be noticed. It was observed that the comprehension of Times New Roman (TNR) compared to the fonts Harrington and Script MT was faster as fewer fixations lead to faster reading. The authors stated that the results proved previous experiments and studies pointing out the effectiveness of reading more familiar fonts like TNR over less familiar fonts [5].

A study [6] monitored familiarity of typefaces with specific characteristics. By making use of three fonts, one standard font and two new typefaces, one of common and one of uncommon structure, the authors tested experimental typography. The participants in this experiment read the text which was presented in Arial efficiently. The authors conclude that typeface familiarity has a positive effect on reading speed after testing on common typefaces like Arial and Grid Sans. Furthermore, the universal structure of the typeface decides upon the reading comfort and speed. If it resembles common structures, typeface can be easier read. In addition, the results show that humans can be familiarized and recognize specific structures over a time. This habituation leads to faster reading and legibility. Especially structures which resemble the Humanistic forms, can create an ease of reading. On the other hand, unknown structures can be recognized after an exposing period and become as legible as the common ones.

As automotive-oriented research on impact of typeface design [7] has shown, legibility can be affected by extrinsic and intrinsic factors. The shape of characters, modulation and features of the font are described as intrinsic factors while extrinsic factors relate to size, illumination and external factors which are not shape-based. This legibility study researched the effect of typeface design in an automotive display by comparing two typeface genres. Eurostile as a square grotesque type design was considered as less legible compared to the humanist design of the typeface Frutiger. The factors that were taken into consideration were open space inside letterforms, letterforms, horizontal proportions and letter spacing. As the authors state, humanist typefaces are effectively identifiable as the open space design provides the reader with distinctive and visible features. The similarity or diversity of letterform shapes, decides on weather a typeface is easily distinguishable. Differences in structure and shape of characters help to identify letters without confusing them with one another [8].

A. Reading on screens

Computers have become a large part of our daily lives and have increasingly become the main product for reading content. Many still do not enjoy having to read long form texts on digital screens, as it often feels difficult because of legibility. This might be because long form content (like large blocks of text) is often linked to, and invokes, a feeling of dread when they appear [9], which may lead to the user skipping large parts after reading the first couple of paragraphs.

Users also often start reading full sentences at the start, and slowly reading less and less per paragraph. While this may be related to the reader's personal preference and familiarity with reading on screens, some studies have shown that the choice of typeface may have a high influence on legibility on screens [2, 10, 11].

Since the length of the content on screen matters in terms of feeling dread, Mariano [9] suggested that content might feel less overwhelming when supporting pictures or photographs added to the text. We know that both color and contrast are factors that can facilitate the legibility of the text [11], and these can heavily influence the time the readers spend on each piece of text.

A reading behaviors study found that a lot more people are accepting of reading long form content on a digital platform, but a lot still prefer a physical copy for in-depth reading [12]. This might be because the eye is strained a lot more using bright screens than it is reading it on paper.

In terms of reading, some studies concluded that eye fixations can be used to accurately measure how engaged the readers are with the text and to what extent it holds their attention [10, 13]. We could also be able to use this information to tell us at what point the readers might start losing focus because of the typeface rather than engagement.

Bar-Zvi Shaked et al. [14] noted during a study on children classified as being poor readers, that these people should have supporting imagery to help them understand what they are currently reading. Reading on screens is less effective than reading on paper, especially in higher education, people reading texts on paper were more likely to have a larger comprehension of the reading material provided [15].

Around 15–20% of the population has a language-based learning disability, and approximately 70-80% have deficits in reading [16]. As dyslexia is such a common problem in the world, we expect that one or two of our participants would likely have this disability. We know that there are differences in the eye movements of readers with and without dyslexia, as they tend to make longer and more fixations than normal readers [17].

For large flowing texts, as seen in word processors, the most common black on white is most often used, there are some exceptions, most people tend to have a better comprehension of the visible information when reading dark text on white background [18]. One study also found that dark characters on a light background were significantly more accurate for proofreading than light characters on a dark background [19].

Most participants in a study related to reading and dyslexia said that a grayscale background did not help them read better [20], most appear to like reading longer texts on a black-on-yellow colorization. The question then stands on whether the eye could be lacking a receptor of some kind, or if it is the softer, warmer glow that comes out of the screen. This could require further research into why most prefer this combination and if the blue light from screens is a contributing factor to this.

A substantial number of developers (e.g., those who study computer science) often tend to use a darker color scheme in their integrated development environment (IDE). This might be because of a personal preference, as they spend more time in front of a computer screen, and the dark mode is more likely to cause less visual discomfort by brightness [21]. A study related to using dark mode on a digital keyboard during daytime found that there were no statistically significant differences in light v. dark, but suggested "that users who prefer dark mode were more conscious about the aesthetical appearance and comfort rather than speed and errors." [22].

Josephson [23] found that sans-serif typeface worked better for legibility than serif typeface, as screens have trouble rendering the small details of the serif types. She also found that participants experienced fewer regressions while reading the story set in Verdana type.

B. Reading methods and gaze patterns

Over the years there have been several studies done using eye tracking and other technologies in order to investigate people's Reading methods and strategies and their effect on reading performance and cognitive workload.

In one article [24] the authors aim to understand how developers interact with different software artifacts when they are performing comprehension tasks, and how this has a potential to improve developers' productivity. In this article the authors propose a new method that can be used to analyze eye-tracking data. They used process mining to find distinct reading patterns of how developers interacted with the different artifacts. This was done as a smaller exploratory study using eye-tracking. In the study they used what's known as "behavior driven development".

"Behavior driven development" is a development practice that is increasingly used in Agile software development. The main result from this study suggests that the proposed method can be used to explore developers' behavior at an aggregated level, as well as to identify behavioral patterns.

One paper aimed to propose an eye-controlled interactive reading system that uses human eyes instead of the traditional mouse to control digital text to support screen-based digital reading [25]. The results from the study revealed that the reading comprehension of learners in the experimental group significantly exceeded those in the control group. They also found that the difference was insignificant for the pure text article. An eye-controlled interactive reading system improved the reading comprehension of field-independent learners more than it did that of field-dependent learners. The reading time of the experimental group significantly exceeded that of the control group.

In a recent study [26] they replicate previous studies done by Busjahn and Peachock et al. [27, 28]. The aim being to provide empirical evidence on the influencing effects of linearity of source code, and how programmers' comprehension strategy on linearity of reading order. To do this they conducted a non-exact replication of studies by Busjahn et al. and Peachock et al. In this study they compared the reading order of different levels of programmers, Novice and expert programmers respectively. In order to do this, they used an eye-tracker to record the eye movements of participants while the participants read. The results they found matched with the results that Busjahn had found previously, but differed from what Peachock had found. This suggests that experience changes the reading behavior of participants (programmers) and that as people get more experience reading code their method and strategy when reading code evolved. However, the linearity of source code has an even stronger effect on reading order than experience. Reading comprehension strategy has a minor effect.

In a study about the impact of tilt of head or of a tablet on reading behavior. They investigated this by having subjects read several texts while their eye movements were recorded. The results showed that there were several texts while their eye movements were recorded. The results showed that there was no effect of a 10°tilt. The participants adapted to the postural situation and Cyclotorsional eye movements were measured with and without tilt. [29]

In other studies [30, 31] the authors use eye tracking in order to look at people's reading habits when reading newspapers collecting quantitative and objective information on subject's behavior and combined with other methodologies – usability testing, focus groups, log analysis in order to help improve online functionality as well as paper based news in order to improve reader retention. Factors, such as text-based elements, size and placement, are an important guide to salience in both media. Other factors such as Images, have mostly been found to have little to no effect on participants' visual attention.

A study looking at reading comprehension at 3 different levels of reading, in which they concluded that scanning relevant

hypertext sections is related to lower performance, especially for good comprehension [32]. The results also show that deep processing of relevant hypertext sections is positively related to better performance, independent of reading comprehension skills.

III. Method

In this section we describe our methods of data gathering, how we selected the participants, how the data was gathered and how it was analyzed. A selection of 3 different types of fonts (Table I) and 2 coloring sets were created for this test. We used the Calibri typeface as our baseline, to set the baseline and validation for reading. This was because it is the most commonly used typeface used by students as it is used by Microsoft Word by default (as of December 2nd, 2021).

A. Participants

The participant selection was based on a presupposition that there is a difference in how different fields of study read texts and how well they are adapted to reading different fonts. Our selection of participants included graphic design, interaction design, web development. All of these were selected based on the students available on campus (NTNU Gjøvik), as they were easily accessible for the limited time frame of this study. The participants were selected based on who were easily available at the campus during the data gathering phase and asked if they would like to participate in the study.

As we were recording personal data, all data was stored safely on a purpose made eye-tracking computer and only accessible through physical means in accordance to the General Data Protection Regulation (GDPR) regulations. No data was ever stored online. Everyone was also asked to give consent to the use of the video and data gathered only for academic purposes. All personally identifiable data is destroyed after the research were completed.

B. Apparatus

Detail what equipment is used, and perhaps explain why it is good / bad for our use-case

In our testing we used the SMI RED250Mobile Eye-tracker [33], which provided us with both a sampling rate of 250Hz and high accuracy (0.4°) we needed for this project. It is not the most widely used eye-tracker within research circles. As of 02 of December 2021 we found a total of 135 different articles that mentioned that they used the eye-tracker (Appendix C).

This may be because this version of eye-tracker is not in production any more as the manufacturer has been bought out by Apple as reported by multiple sources [34–37].

C. Procedure

Before the participants performed the reading task, they were asked a couple of questions about their ability to read on screens, this included things like how often they read on computers and what they read most often. While this was not necessary information, this was something we found could influence the overall result. We avoided asking personal questions about disabilities, but if they themselves told us, we noted it down as a comment if we wanted to add it as a separate datapoint.

TABLE I
Fonts selected for this test.

Font type	Typeface
Validation & Serif	Times New Roman
Sans-serif	Calibri
Mono	Jetbrains Mono

After they had answered the simple questions above, we continued with performing a calibration of the equipment, so it would be as accurate as possible on every participant.

D. Data gathering

During this phase, we asked the participants to read a set of texts, each of which had different fonts, color schemes and sizing. Between each image, we asked them to rate the difficulty of the text on a scale of 1 to 10, where 10 was difficult and 1 was easy. In all stimulus we hid simple spelling errors, and mixed-up letters, to see if they either spotted it with the current font, and to see if the error made any difference in fixations.

This data was to get a subjective view of how they perceived reading on these selected fonts, while this was not strictly necessary for our project, we found that having more data points to compare against the eye-tracker data was interesting and could tell us more about perception vs actual data.

To attempt to avoid any bias towards mental exhaustion being the reason for how our participants performed, the tasks were randomized within dark and light groupings. The participants were shown six different texts using three different fonts (table I), with two color schemes. The colors used were total white (#ffffff) and a almost black dark gray (#1A1A1A), this was done to slightly reduce the amount of "halation" on the screen. Halation could be defined as "light leaking" around the text, causing a visual fuzzy looking effect [38, 39].

Between each stimulus, we had a period we define as "break" where the participant could look away from the screen for a moment while answering the questions. How long the break between each stimulus was based on how quick the participant responded to the questions. If the participant was quick in their response, the time between each test was less than if the participants spent a longer time on each question.

The length of all the texts used for creating the stimuli were selected to be two paragraphs long, each from official reading comprehension tests [40, 41].

a) Coloring: Dark text on light background and light on dark background (#1A1A1A)

b) Sentences: Text length was on average 1 paragraph long, where we had hidden 2 to 5 switched letters in the text. All sentences was different to not allow the reader to just skim the text while knowing what it said.

c) Number of tasks per participant: A selection of 4 fonts, 2 different color schemes (See coloring) for a total of 8 different variations. Each of which unique in their own way.

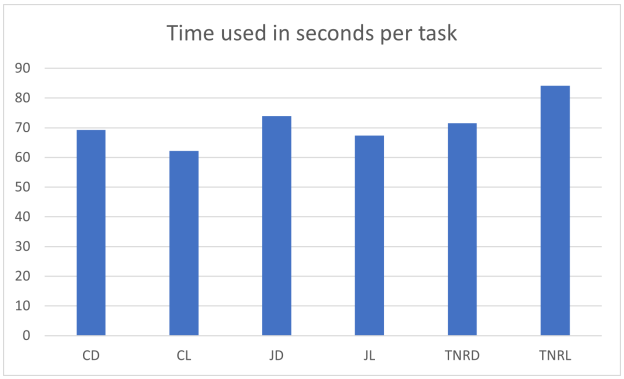


Fig. 1. A summary of the average time spent on reading per typeface on all 8 participants.

d) Randomization of tasks: The order of the fonts was always in the same order, see I for the order, as this would allow us to easily read the text

e) Time limits of the tasks: There was no set time limit for the reading, but we recorded how long they spent looking at the text for statistical analysis.

f) Between tasks: Between each task, we asked the participants what the paragraphs were about, and if they could rate how easy the font was to read on a scale of 1-10, where 1 was difficult and 10 was easy.

Between each variation of fonts we asked how often they felt the needed to backtrack to read the sentence, and how they felt about reading with this font. We also did a validation text to check how often they felt "stuck" on a word before being able to continue reading.

The time between each task is based on the quickness of the response from the participants.

g) Stimuli: All stimuli contained different text such that we could validate that the contents of the text was actually read. These stimuli (Appendix B) had used one of the following sets of typeface and color combination; Calibri Light (CL), Calibri Dark (CD), Times New Roman Light (TNRL), Times New Roman Dark (TNRD), Jetbrains Light (JL), Jetbrains Dark (JD).

IV. Results

The participants on average used 7 minutes to do the reading of all the stimuli, where on average they used about 71 seconds per task. In general, we found that most people, within our participant selection, tended to have a similar reading speed, causing the time-per-task to be less than 2minutes per task. This is within expectation, as we know on average slow readers tend to read about 200 wpm [42], which also correlates and is similar to how our participants focus more than regular to be able to remember the entire text.

In our testing, we also found that most participants preferred reading the Calibri font using a light background and a dark text. This is as expected, as we know that people are students and often use Microsoft Word as their main writing editor.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Scanpath	Sphericity Assumed	115407392.861	2	57703696.431	4.740	0.036
	Greenhouse-Geisser	115407392.861	1.177	98021384.671	4.740	0.070
	Huynh-Feldt	115407392.861	1.325	87113129.734	4.740	0.062
	Lower-bound	115407392.861	1.000	115407392.861	4.740	0.081
Error(Scanpath)	Sphericity Assumed	121738729.806	10	12173872.981		
	Greenhouse-Geisser	121738729.806	5.887	20679782.409		
	Huynh-Feldt	121738729.806	6.624	18378444.397		
	Lower-bound	121738729.806	5.000	24347745.961		

Fig. 2. Figure showing the possibility of significance of the of scanpath length. We found that there could be a significance within a Sphericity Assumed analysis.

Within-Subjects Factors

Measure: MEASURE_1

Scanpath	Dependent Variable
1	Path_Calibri
2	Path_Times
3	Path_Jetbrains

Fig. 3. Figure showing variables and their naming.

In terms of reading speed, we can see that on average CL is fastest, though only by about 10 seconds as seen in figure 1.

Though TNRL was the typeface with the longest Time-Per-Task, it was also the one that was perceived to be easiest to read. While the CL was the one most tended to spend the least time on, it was the second in perceived easiness of reading. The most perceived difficult typeface to read was the JD, while the time used was the second slowest. While this could be because of the low number of participants, it can help set a example of what to expect with a higher number of participants.

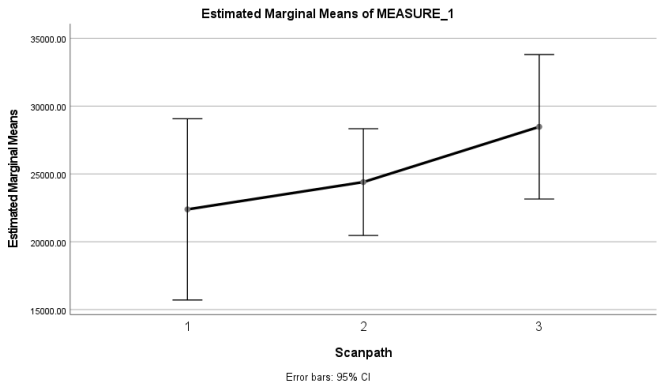
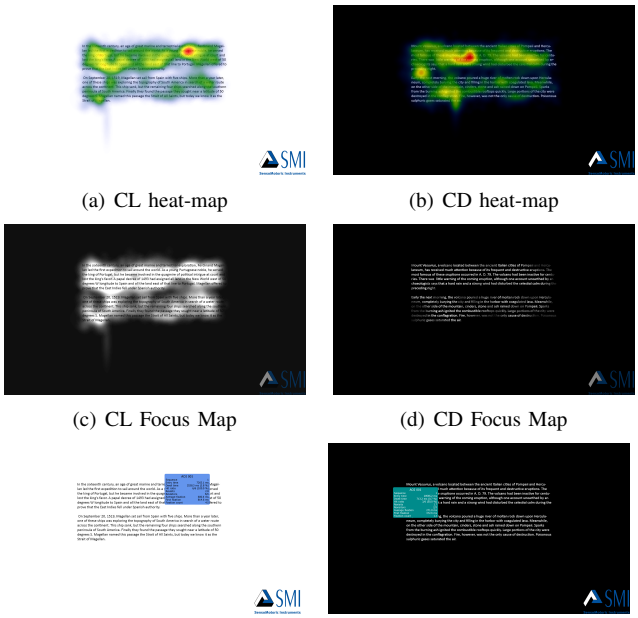


Fig. 4. Difference in scanpath length including error margin. See 3 for explanation as to the factors

Fig. 5. Calibri

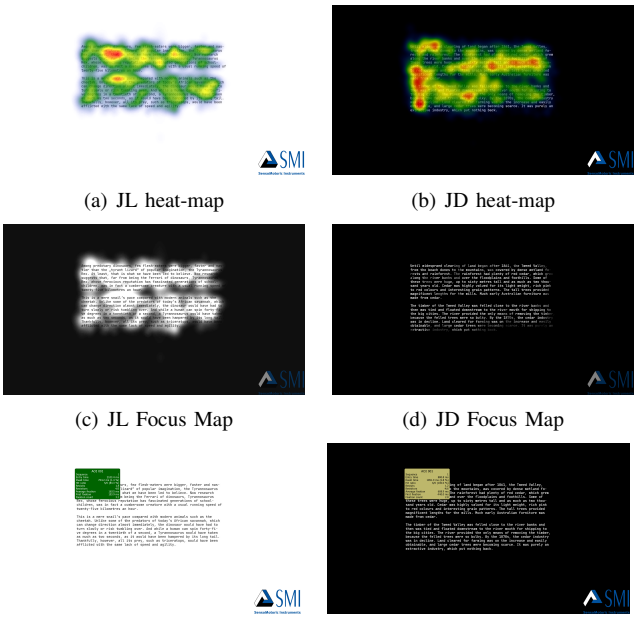


(a) CL heat-map (b) CD heat-map

(c) CL Focus Map (d) CD Focus Map

(e) CL KPI (f) CD KPI

Fig. 6. Jetbrains Mono

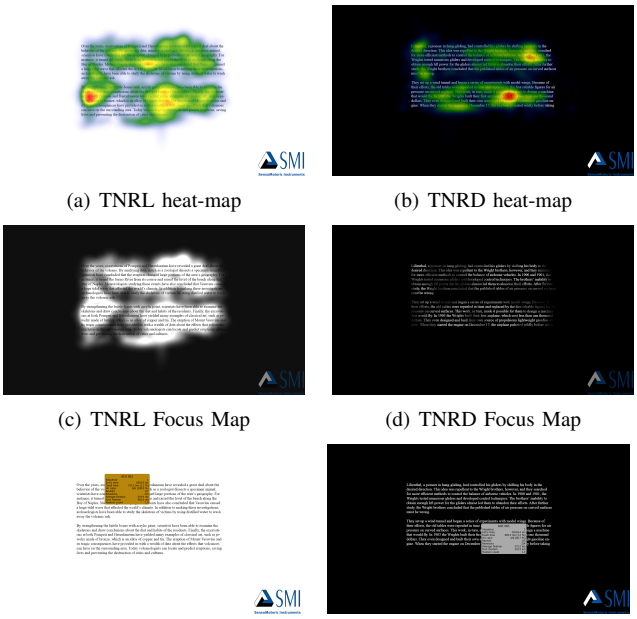


(a) JL heat-map (b) JD heat-map

(c) JL Focus Map (d) JD Focus Map

(e) JL KPI (f) JD KPI

Fig. 7. Times New Roman



(a) TNRL heat-map (b) TNRD heat-map

(c) TNRL Focus Map (d) TNRD Focus Map

(e) TNRL KPI (f) TNRD KPI

Due to a lack of participants we did not manage to find any significant outcomes insight into participants reading behaviours. However, we did find a potential significance in the scanpath length of mono font compared to serif and sans-serif. This could be a further research point, which requires a significantly larger pool of participants.

A. Heat maps, Focus Map and KPI

The graphical illustrations (fig 5, 6 and 7) containing the heat maps, focus maps and Key Performance Indicator (KPI) were generated by aggregating participant data through the SMI software provided to read the data from the Eye-tracker. It shows areas of interest and places where the participants looked the longest.

The heat maps from the stimuli using the Jetbrains Mono typeface (See figure 6), shows that the people have focused only on some parts of the stimuli, mostly related to difficult words. The increased amount of red and yellow seen on fig 7(b), could suggest an increased focus needed to read each word or repeated visitations to understand each word. An increased visit in the top left in both fig. 7(a) and 7(b) which correlates well with the amount of visits seen in 7(e) and 7(f).

Since within the praxis of code and programming, the information contained within each word is often more important than the sentence as a whole. Since each word has such a large impact, the ability to see every word clearly is of utmost importance. We can see this in figure 7(c) and 7(d), which has shown that the focus is clearly more on each word rather than scanning past it. According to what people themselves answered within the subjective parts (Appendix A), the reading using this typeface was perceived as difficult. This might not be because they could not focus on the sentence, but rather that each word had such a significant impact on the sentence, that the people could not attain a sort of reading flow found within the other typefaces.

The light heat maps suggest loss of focus using the calibri typeface in text longer than a paragraph. While the dark heat maps suggest strong focus on each word for both Calibri

and Jetbrains Correlating to our finds in scanpath length. The Heatmaps show that the participants spent more time focusing on specific words to be able to relay that information back to us correctly.

Spelling mistakes were hidden in the text paragraphs. As can be derived from our Area Of Interest (AOI) images in the Appendix, the participants detected the spelling mistakes within the text paragraphs in some of the cases.

For example, we spelt the word Portuguese as *Purtugoese*, was detected by all six participants, or at the very least their eyes looked at the word. The average dwell time came about 2538.3 ms and all six participants revisited the spelling mistake. That they focused on misspelled words can be additionally seen in the heat maps (see fig 5, 6 and 7).

V. Discussion

A. Findings

In the subjective parts of the text, we found that the majority felt the validation text was easier to read than one using the same typeface and size had a higher overall perceived difficulty. Though this is only subjective and does not change our findings in any way.

We have included the subjective results, since they provide information about the perceived readability of the texts in appendix A.

Though we did find a possible significance, we do not know whether it is a statistical possibility of it being accurate or just the typeface itself being the cause of the significance. Monospace typefaces tend to be utilized by a majority of IDEs and software text editors. This is because, according to Spolsky [43], it increases the readability of source code by having all letters be the same width.

B. Limitations

Out of our testing subjects, we found that 4 out of the 11 participants had to be eliminated from our results, as these had a tracking ratio of less than 30%, and another participant had a tracking ratio of 16%. This left us with a total number of 6 valid participant measures. We have also included the subjective results (see Appendix A), since they provide information about the perceived readability of the texts.

The number of participants cannot give a statistically accurate proof that reading is different depending on what field of study the reader is from. Hence our research on how different focus of study affects screen reading did not bring valid statements and further investigation with a higher number of participants from each study field is necessary.

The text paragraphs were presented to the participants in the language English. It must be considered that none of the participants is a native speaker in the selected language. Therefore, reading speed, understanding of content or detection of spelling mistakes are affected substantially. We also did not account for boredom, as that could lead to mind wandering and loss of focus, which could affect our result in terms of reading ability, but would not invalidate our findings in terms of speed.

C. Reliability

We found that in most cases, during the reading of the stimuli, the participants tended to lean forward as they read the text. The SMI has a measurement range of $\approx 55 - 65cm$, in which we can get accurate data. This caused a third of our data to be invalidated as they were outside the range of the SMI sensor.

In addition, according to a study looking into data quality in eye-trackers, the SMI RED250Mobile, exhibits errors exceeding $1°$for movements that are smaller than $10arcmin$ [44]. Which makes reliable eye tracking data, requiring tiny movements not entirely accurate. Though for larger than $10arcmin$, then the data would be accurate. It was also found that the RED250mobile had a mismeasurement that occurred approximately around the same position of the screen [44].

By performing this test multiple times more, we could establish a pattern of knowing whether it was boredom, the font, or the text itself being difficult to read that causes a visible drop-off in reading towards the latter parts of the test.

VI. Conclusion

In this study we found that there were subtle differences in how much of the text was actually looked at, though this could be because of flow-states. Jetbrains did show a larger focus on every letter, but this could be because the participants found the typeface difficult to read rather than the participants remembering and understanding more of the text.

We could also see in fig 6(a) that the most participants tended to have a larger focus on the first part of the stimuli, this suggests that there either were an adjustment period to the font happening at the start (requiring re-reading), or had a larger attention on the first part of the test. We can also speculate that this is due to reaching a “flow” state, allowing one to read without having to read each word seperatly.

Though we did not find a significant reading difference when comparing different fields of study, we expect that there might be a possible difference, however, to account for the difference in people a large sample size would be necessary to find a difference within these groupings. We assume that there possibly are subtle differences in how different fields of study read a text.

REFERENCES

[1] R. L. Johnson and M. E. Eisler, "The importance of the first and last letter in words during sentence reading," *Acta Psychologica*, vol. 141, no. 3, pp. 336–351, Nov. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001691812001564

[2] G. Franken, A. Podlesek, and K. Možina, "Eye-tracking study of reading speed from LCD displays: Influence of type style and type size," *Journal of eye movement research*, vol. 8, no. 1, 2015, place: IFFWIL Publisher: IFFWIL: INT GROUP EYE MOVEMENT RESEARCH.

[3] L. Reynolds and D. Simmonds, "The legibility of type," in *Presentation of Data in Science: Publications, slides, posters, overhead projections, tape-slides, television Principles and practices for authors and teachers*, L. Reynolds and D. Simmonds, Eds. Dordrecht: Springer Netherlands, 1984, pp. 1–19. [Online]. Available: https://doi.org/10.1007/978-94-009-8354-0_1

[4] T. J. Slattery and K. Rayner, "The influence of text legibility on eye movements during reading," *Appl. Cognit. Psychol*, vol. 24, no. 8, pp. 1129–1148, 2010, place: Chichester, UK Publisher: Chichester, UK: John Wiley & Sons, Ltd.

[5] K. Rayner, E. D. Reichle, M. J. Stroud, C. C. Williams, and A. Pollatsek, "The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers." *Psychology and Aging*, vol. 21, no. 3, pp. 448–465, 2006, place: US Publisher: American Psychological Association.

[6] U. Nedeljković, K. Jovančić, and N. Pušnik, "You read best what you read most: An eye tracking study," *J Eye Mov Res*, vol. 13, no. 2, pp. 1–18, 2020, place: Switzerland Publisher: Switzerland: Bern Open Publishing.

[7] B. Reimer, B. Mehler, J. Dobres, J. F. Coughlin, S. Matteson, D. Gould, N. Chahine, and V. Levantovsky, "Assessing the impact of typeface design in a text-rich automotive user interface," *Ergonomics*, vol. 57, no. 11, pp. 1643–1658, Nov. 2014, publisher: Taylor & Francis. [Online]. Available: https://doi.org/10.1080/00140139.2014.940000

[8] F. Attneave and M. D. Arnoult, "The quantitative study of shape and pattern perception," *Psychol Bull*, vol. 53, no. 6, pp. 452–471, 1956, place: United States Publisher: United States: American Psychological Association.

[9] J. Marino, "Reading Screens: What Eye Tracking Tells Us about the Writing in Digital Longform Journalism," *Literary journalism studies*, vol. 8, no. 2, p. 138, 2016.

[10] C. Ralekar, P. Saha, T. K. Gandhi, and S. Chaudhury, "Effect of Devanagari Font Type in Reading Comprehension: An Eye Tracking Study," in *Intelligent Human Computer Interaction*, U. S. Tiwary, Ed. Cham: Springer International Publishing, 2018, pp. 136–147.

[11] G. Franken, M. Pangerc, and K. Možina, "Impact of typeface and colour combinations on lcd display legibility," *Emerging science journal*, vol. 4, no. 6, pp. 436–442, 2020, publisher: Ital Publication.

[12] Z. Liu, "Reading behavior in the digital environment: Changes in reading behavior over the past ten years," *Journal of documentation*, vol. 61, no. 6, pp. 700–712, 2005, place: BRADFORD Publisher: BRADFORD: Emerald Group Publishing Limited.

[13] B. W. Miller, "Using Reading Times and Eye-Movements to Measure Cognitive Engagement," *Educational psychologist*, vol. 50, no. 1, pp. 31–42, 2015, place: ABINGDON Publisher: ABINGDON: Routledge.

[14] K. Bar-Zvi Shaked, A. Shamir, and E. Vakil, "An eye tracking study of digital text reading: a comparison between poor and typical readers," *Reading & writing*, vol. 33, no. 8, pp. 1925–1944, 2020, publisher: Springer.

[15] N. Latini, I. Bråten, and L. Salmerón, "Does reading medium affect processing and integration of textual and pictorial information? A multimedia eye-tracking study," *Contemporary educational psychology*, vol. 62, p. 101870, 2020, publisher: Elsevier Inc.

[16] International Dyslexia Association, "Frequently Asked Questions," 2021. [Online]. Available: https://dyslexiaida.org/frequently-asked-questions-2/

[17] L. Rello and R. Baeza-Yates, "The Effect of Font Type on Screen Readability by People with Dyslexia," *ACM transactions on accessible computing*, vol. 8, no. 4, pp. 1–33, 2016, publisher: ACM.

[18] A.-H. Wang, J.-J. Fang, and C.-H. Chen, "Effects of VDT leading-display design on visual performance of users in handling static and dynamic display information dual-tasks," *International Journal of Industrial Ergonomics*, vol. 32, no. 2, pp. 93–104, Aug. 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169814103000416

[19] C. Piepenbrock, S. Mayr, and A. Buchner, "Smaller pupil size and better proofreading performance with positive than with negative polarity displays," *Ergonomics*, vol. 57, no. 11, pp. 1670–1677, 2014, place: ABINGDON Publisher: ABINGDON: Taylor & Francis.

[20] L. Rello and R. Baeza-Yates, "How to present more readable text for people with dyslexia," *Universal access in the information society*, vol. 16, no. 1, pp. 29–49, 2017, place: Berlin/Heidelberg Publisher: Berlin/Heidelberg: Springer Berlin Heidelberg.

[21] H. Eisfeld and F. Kristallovich, "The Rise of Dark Mode : A qualitative study of an emerging user interface design trend," Student thesis, Jönköping University, 2020, volume: Independent thesis Basic level (degree of Bachelor). [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-50563

[22] L. A. Pedersen, S. S. Einarsson, F. A. Rikheim, and F. E. Sandnes, "User Interfaces in Dark Mode During Daytime – Improved Productivity or Just Cool-Looking?" in *Universal Access in Human-Computer Interaction. Design Approaches and Supporting Technologies*, M. Antona and C. Stephanidis, Eds. Cham: Springer International Publishing, 2020, pp. 178–187.

[23] S. Josephson, "Keeping Your Readers' Eyes on the Screen: An Eye-Tracking Study Comparing Sans Serif and Serif Typefaces," *Visual communication quarterly*, vol. 15, no. 1-2, pp. 67–79, 2008, publisher: Taylor & Francis Group.

[24] C. Ioannou, I. Nurdiani, A. Burattin, and B. Weber, "Mining reading patterns from eye-tracking data: method and demonstration," *Software and systems modeling*, vol. 19, no. 2, pp. 345–369, 2020, place: Heidelberg Publisher: Heidelberg: Springer Nature B.V.

[25] C.-M. Chen, J.-Y. Wang, and Y.-C. Lin, "A visual interactive reading system based on eye tracking technology to improve digital reading performance," *Electronic library*, vol. 37, no. 4, pp. 680–702, 2019, place: Oxford Publisher: Oxford: Emerald Publishing Limited.

[26] N. Peitek, J. Siegmund, and S. Apel, "What Drives the Reading Order of Programmers? An Eye Tracking Study," in *Proceedings of the 28th International Conference on Program Comprehension*. Seoul, Republic of Korea: Association for Computing Machinery, 2020, pp. 342–353, type: 10.1145/3387904.3389279. [Online]. Available: https://doi.org/10.1145/3387904.3389279

[27] T. Busjahn, R. Bednarik, A. Begel, M. Crosby, J. H. Paterson, C. Schulte, B. Sharif, and S. Tamm, "Eye Movements in Code Reading: Relaxing the Linear Order," in *2015 IEEE 23rd International Conference on Program Comprehension*, May 2015, pp. 255–265, journal Abbreviation: 2015 IEEE 23rd International Conference on Program Comprehension.

[28] P. Peachock, N. Iovino, and B. Sharif, "Investigating Eye Movements in Natural Language and C++ Source Code - A Replication Experiment," in *Augmented Cognition. Neurocognition and Machine Learning*, D. D. Schmorrow and C. M. Fidopiastis, Eds. Cham: Springer International Publishing, 2017, pp. 206–218.

[29] J.-L. Perrin, D. Paillé, and T. Baccino, "Reading tilted: Does the use of tablets impact performance? An oculometric study," *Computers in human behavior*, vol. 39, pp. 339–345, 2014, place: OXFORD Publisher: OXFORD: Elsevier Ltd.

[30] D. Zambarbieri, E. Carniglia, and C. Robino, "Eye Tracking Analysis in Reading Online Newspapers," *Journal of eye movement research*, vol. 2, no. 4, 2008, place: IFFWIL Publisher: IFFWIL: INT GROUP EYE MOVEMENT RESEARCH.

[31] S. Leckner, "Presentation factors affecting reading behaviour in readers of newspaper media: an eye-tracking perspective," *Visual Communication*, vol. 11, no. 2, pp. 163–184, 2012, tex.eprint: https://doi.org/10.1177/1470357211434029. [Online]. Available: https://doi.org/10.1177/1470357211434029

[32] H. Jarodzka and S. Brand-Gruwel, "Tracking the reading eye: Towards a model of real-world reading," *Journal of computer assisted learning*, vol. 33, no. 3, pp. 193–201, 2017, place: HOBOKEN Publisher: HOBOKEN: WILEY.

[33] "SMI RED250mobile," Feb. 2021. [Online]. Available: https://imotions.com/hardware/smi-red250mobile/

[34] P. Hartzbech, "SMI and Apple: What It Means for the Future of Eye Tracking," Jul. 2017. [Online]. Available: https://imotions.com/blog/smi-apple-eye-tracking/

[35] L. Matney, "Apple acquires SMI eye-tracking company," Jun. 2017. [Online]. Available: https://social.techcrunch.com/2017/06/26/apple-acquires-smi-eye-tracking-company/

[36] I. Fried, "Archived: Apple confirms it has bought a small German computer vision company," 2017. [Online]. Available: https://web.archive.org/web/20210524180611/https://www.axios.com/apple-confirms-it-has-bought-a-small-german-computer-vision-company-2448697718.html
[37] B. Lang, "Report: Apple Acquires VR & AR Eye-tracking Company SMI," Jun. 2017, section: Apple. [Online]. Available: https://www.roadtovr.com/report-apple-acquires-vr-ar-eye-tracking-company-smi/
[38] Anthony, "Why You Should Never Use Pure Black for Text or Backgrounds," May 2018. [Online]. Available: https://uxmovement.com/content/why-you-should-never-use-pure-black-for-text-or-backgrounds/
[39] Jessica Otis, "Never Use White Text on a Black Background: Astygmatism and Conference Slides," Nov. 2017. [Online]. Available: https://jessicaotis.com/academia/never-use-white-text-on-a-black-background-astygmatism-and-conference-slides/
[40] Victoria State Government, "Reading Comprehension Practice test." [Online]. Available: https://www.education.vic.gov.au/Documents/school/parents/secondary/readingcomprehensionpractice.pdf
[41] Metropolitan Community College, "Reading Comprehension Practice Test." [Online]. Available: https://www.mcckc.edu/tutoring/docs/bt/readwrite/Reading_Comprehension_Practice_Test.pdf
[42] K. Rayner, T. J. Slattery, and N. N. Bélanger, "Eye movements, the perceptual span, and reading speed," *Psychonomic Bulletin & Review*, vol. 17, no. 6, pp. 834–839, Dec. 2010. [Online]. Available: https://doi.org/10.3758/PBR.17.6.834
[43] J. Spolsky, "User Interface Design For Programmers," Oct. 2001. [Online]. Available: https://www.joelonsoftware.com/2001/10/24/user-interface-design-for-programmers/
[44] K. Holmqvist and P. Blignaut, "Small eye movements cannot be reliably measured by video-based P-CR eye-trackers," *Behav Res Methods*, vol. 52, no. 5, pp. 2098–2121, 2020, place: New York Publisher: New York: Springer US.

APPENDICES

APPENDIX A
SUBJECTIVE TEST DATA

This is the anonymous data the participants provided to us voluntarily. Note that anything marked in red is invalidated, and was not counted in the results.

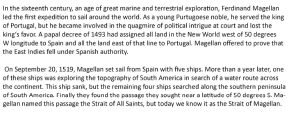
General Information			Percieved Reading Difficulty								
Participant NR	Field of study [1]	Gender [2]	Avg reading difficulty validation [3]		CL [4]	CD [5]	TNRL [6]	TNRD [7]	JL [8]	JD [9]	Notes [10]	T/R [11]	Total time used
-3			5.8		7	4		7	6	5	Invalidated, V	30	
-2			6.6		5	9		7	6	6	Invalidated, V	30	
-1			4.33	1				3	5	5	Invalidated, V	30	
1	2	1	3.83	2	5	5	1	3	5	4		77.3	
2	2	0	2.67	1	2	2	4	3	3	2		75.5	
3	1	1	5.33	1	5	5	5	4	7	6		86	
4	1	0	4.83	2	5	5	2	4	6	7		81.9	
5	2	0	5.83	5	4	6	6	7	5	7	Invalidated	16.9	
6	2	0	5.5	3	3	6	6	4	7	7	Dys, generally	84.6	
7	2	0	4.17	3	3	4	5	4	2	7	generally poste	92.8	
8	3	1	4.67	1	5	3	4	5	4	7	Too low?	62.6	
Summary			Avg Validation	Avg font CL		Avg font CD	Avg font TNRL	Avg font TNRD	Avg font JL	Avg font JD			
			2.375	5.5		6.125	4.125	6.375	7	7.875			

Bracketed numbers explanation
[1]
1. Graphic Design,
2. Interaction Design,
3. Programming
[2]
0 Female
1 Male
2 Other?
[3]
Validation, the first one
Not counted
[4]
Calibri
Light background, dark text
[5]
Calibri
dark background, Light text
[6]
Times New Roman
Light background, dark text
[7]
Times New Roman
dark background, Light text
[8]
Jetbrains Mono
Light background, dark text
[9]
Jetbrains Mono
dark background, Light text
[10]
Notes that could be relevant for how the tester performed.
If there is other, longer notes, put them in a separate location

APPENDIX B
STIMULI

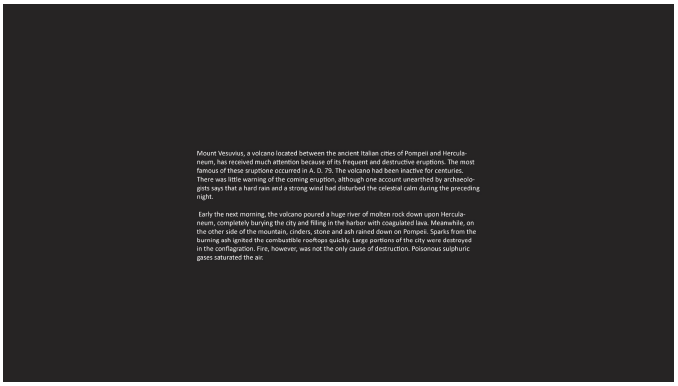
The following are the stimuli used to for the research.

A. Stimuli Calibri Light


In the sixteenth century, an age of great marine and terrestrial exploration, Ferdinand Magellan led the first expedition to sail around the world. As a young Portuguese noble, he served the king of Portugal, but he became involved in the quagmire of political intrigue at court and lost the king's favor. A papal decree of 1493 had assigned all land in the New World west of 50 degrees W longitude to Spain and all the land east of that line to Portugal. Magellan offered to prove that the East Indies fell under Spanish authority.

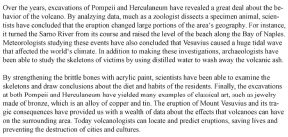
On September 20, 1519, Magellan set sail from Spain with five ships. More than a year later, one of these ships was exploring the topography of South America in search of a water route across the continent. This ship sank, but the remaining four ships searched along the southern peninsula of South America. Finally they found the passage they sought near a latitude of 50 degrees S. Magellan named this passage the Strait of All Saints, but today we know it as the Strait of Magellan.

B. Stimuli Calibri Dark


Mount Vesuvius, a volcano located between the ancient Italian cities of Pompeii and Herculaneum, has received much attention because of its frequent and destructive eruptions. The most famous of these eruptions occurred in A. D. 79. The volcano had been inactive for centuries. There was little warning of the coming eruption, although one account unearthed by archaeologists says that a hard rain and a strong wind had disturbed the celestial calm during the preceding night.

Early the next morning, the volcano poured a huge river of molten rock down upon Herculaneum, completely burying the city and filling in the harbor with coagulated lava. Meanwhile, on the other side of the mountain, cinders, stone and ash rained down on Pompeii. Sparks from the burning ash ignited the combustible rooftops quickly. Large portions of the city were destroyed in the conflagration. Fire, however, was not the only cause of destruction. Poisonous sulphuric gases saturated the air.

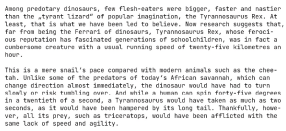
C. Stimuli Times New Roman Light


Over the years, excavations of Pompeii and Herculaneum have revealed a great deal about the behavior of the volcano. By analyzing data, much as a zoologist dissects a specimen animal, scientists have concluded that the eruption changed large portions of the area's geography. For instance, it turned the Sarno River from its course and raised the level of the beach along the Bay of Naples. Meteorologists studying these events have also concluded that Vesuvius caused a huge tidal wave that affected the world's climate. In addition to making these investigations, archaeologists have been able to study the skeletons of victims by using distilled water to wash away the volcanic ash.

By strengthening the brittle bones with acrylic paint, scientists have been able to examine the skeletons and draw conclusions about the diet and habits of the residents. Finally, the excavations at both Pompeii and Herculaneum have yielded many examples of classical art, such as jewelry made of bronze, which is an alloy of copper and tin. The eruption of Mount Vesuvius and its tragic consequences have provided us with a wealth of data about the effects that volcanoes can have on the surrounding area. Today volcanologists can locate and predict eruptions, saving lives and preventing the destruction of cities and cultures.

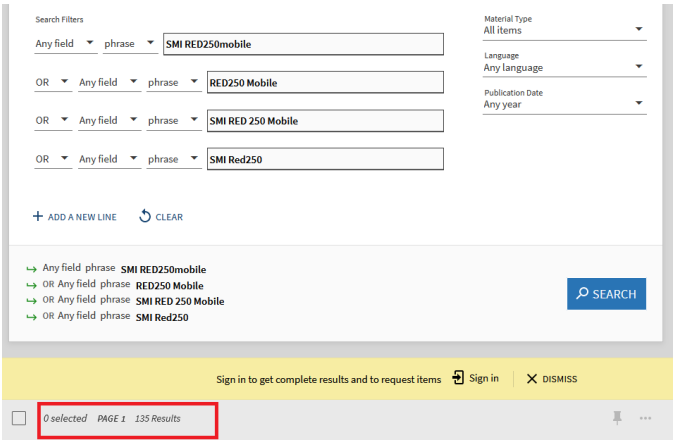
D. Stimuli Times New Roman Dark

E. Stimuli Jetbrains Light

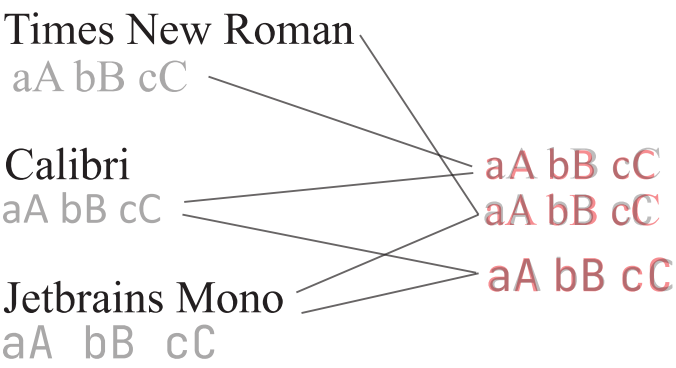


F. Stimuli Jetbrains Dark

Appendix C
Oria Search



Appendix D
Similarities between fonts



Appendix E
Text snippets used

These snippets include the spelling errors highlighted in red. The snippets used are from [41] and [40].

A. Draining/training session/Warm up/training text

Hi! Nice to meet you! My name is John Smith. I am 19 and a student in college. I go to college in New York. My favorite courses are Geometry, French, and History. English is my hardest course. My professors are very friendly and smart. It's my second year in college now. I love it!

I live in a big house on Ivy Street. It's near the college campus. I share the house with three other students. Their names are Bill, Tony, and Paul. We help each other with homework. On the weekend, we play football together.

Source: *https://lingua.com/english/reading/john/*

B. Text 1: Appendix B-A

In the sixteenth century, an age of great marine and terrestrial exploration, Ferdinand Magellan led the first expedition to sail around the world. As a young Portugoese noble, he served the king of Portugal, but he became involved in the quagmire

of political intrigue at court and lost the king's favor. A papal decree of 1493 had assigned all land in the New World west of 50 degrees W longitude to Spain and all the land east of that line to Portugal. Magellan offered to prove that the East Indies fell under Spanish authority.

On September 20, 1519, Magellan set sail from Spain with five ships. More than a year later, one of these ships was exploring the topography of South America in search of a water route across the continent. This ship sank, but the remaining four ships searched along the southern peninsula of South America. Finally they found the passage they sought near a latitude of 50 degrees S. Magellan named this passage the Strait of All Saints, but today we know it as the Strait of Magellan.

C. Text 2: Appendix B-B

Mount Vesuvius, a volcano located between the ancient Italian cities of Pompeii and Herculaneum, has received much attention because of its frequent and destructive eruptions. The most famous of these sruption occurred in A. D. 79. The volcano had been inactive for centuries. There was little warning of the coming eruption, although one account unearthed by archaeologists says that a hard rain and a strong wind had disturbed the celestial calm during the preceding night.

Early the next morning, the volcano poured a huge river of molten rock down upon Herculaneum, completely burying the city and filling in the harbor with coagulated lava. Meanwhile, on the other side of the mountain, cinders, stone and ash rained down on Pompeii. Sparks from the burning ash ignited the combustible rooftops quickly. Large portions of the city were destroyed in the conflagration. Fire, however, was not the only cause of destruction. Poisonous sulphuric gases saturated the air. These heavy gases were not buoyant in the atmosphere and therefore sank toward the earth and suffocated people.

D. Text 3: Appendix B-C

Over the years, excavations of Pompeii and Herculaneum have revealed a great deal about the behavior of the volcano. By analzying data, much as a zoologist dissects a specimen animal, scientists have concluded that the eruption changed large portions of the area's geography. For instance, it turned the Sarno River from its course and raised the level of the beach along the Bay of Naples. Meteorologists studying these events have also concluded that Vesuvius caused a huge tidal wave that affected the world's climate. In addition to making these investigations, archaeologists have been able to study the skeletons of victims by using distilled water to wash away the volcanic ash.

By strengthening the brittle bones with acrylic paint, scientists have been able to examine the skeletons and draw conclusions about the diet and habits of the residents. Finally, the excavations at both Pompeii and Herculaneum have yielded many examples of classical art, such as jewelry made of bronze, which is an alloy of copper and tin. The eruption of Mount Vesuvius and its tragic consequences have provided us with a wealth of data about the effects that volcanoes can have on the surrounding area. Today volcanologists can locate and predict eruptions, saving lives and preventing the destruction of cities and cultures.

E. Text 4: Appendix B-D

Lilienthal, a pioneer in hang-gliding, had controlled his gliders by shifting his body in the desired direction. This idea was repellent to the Wright brothers, however, and they searched for more efficient methods to control the balance of airborne vehicles. In 1900 and 1901, the Wrights tested numerous gliders and developed control techniques. The brothers' inability to obtain enough lift power for the gliders almost led them to abandon their efforts. After further study, the Wright brothers concluded that the published tables of air pressure on curved surfaces must be wrong.

They set up a wind tunnel and began a series of experiments with model wings. Because of their efforts, the old tables were repealed in time and replaced by the first reliable figures for air pressure on curved surfaces. This work, in turn, made it possible for them to design a machine that would fly. In 1903 the Wrights built their first airplane, which cost less than one thousand dollars. They even designed and built their own source of propulsiona lightweight gasoline engine. When they started the engine on December 17, the airplane pulsated wildly before taking off. The plane managed to stay aloft for twelve seconds, however, and it flew one hundred twenty feet.

F. Text 5: Appendix B-E

Among predotary dinosaurs, few flesh-eaters were bigger, faster and nastier than the "tyrant lizard" of popular imagination, the Tyrannosaurus Rex. At least, that is what we have been led to believe. Now research suggests that, far from being the Ferrari of dinosaurs, Tyrannosaurus Rex, whose ferocious reputation has fascinated generations of schoolchildren, was in fact a cumbersome creature with a usual running speed of twenty-five kilometres an hour.

This is a mere snail's pace compared with modern animals such as the cheetah. Unlike some of the predators of today's African savannah, which can change direction almost immediately, the dinosaur would have had to turn slowly or risk tumbling over. And while a human can spin forty-five degrees in a twentieth of a second, a Tyrannosaurus would have taken as much as two seconds, as it would have been hampered by its long tail. Thankfully, however, all its prey, such as triceratops, would have been afflicted with the same lack of speed and agility.

G. Text 6: Appendix B-F

Until widespraed clearing of land began after 1861, the Tweed Valley, from the beach dunes to the mountains, was covered by dense wetland forests and rainforest. The rainforest had plenty of red cedar, which grew along the river banks and over the floodplains and foothills. Some of these trees were huge, up to sixty metres tall and as much as two thousand years old. Cedar was highly valued for its light weight, rich pink to red colours and interesting grain patterns. The tall

trees provided magnificent lengths for the mills. Much early Australian furniture was made from cedar.

The timber of the Tweed Valley was felled close to the river banks and then was tied and floated downstream to the river mouth for shipping to the big cities. The river provided the only means of removing the timber, because the felled trees were so bulky. By the 1870s, the cedar industry was in decline. Land cleared for farming was on the increase and easily obtainable, and large cedar trees were becoming scarce. It was purely an extractive industry, which put nothing back.