university of groningen

faculty of science and engineering

University of Groningen

Colour Perceptual Organisation and its Influence on Reading Speed

Master's Internship report

To fulfill the requirements for the Research Internship course
for the Master's degree in Computing Science at University of Groningen
under the supervision of

S. Bruin, MSc. (Computing Science, University of Groningen)
and
Dr. C. Tursun (Computing Science, University of Groningen)
and
Prof. J. Kosinka (Computing Science, University of Groningen)

Matej Kučera (s4551192)

June 24, 2024

Contents

Page

Abstract **3**

1 Introduction **3**
1.1 Research questions . . . 3

2 State of the Art **4**
2.1 Inspiration . . . 4
2.2 Gestalt principles . . . 4
2.3 Related work . . . 4
2.4 Search terms . . . 5

3 Method **6**
3.1 PsychoPy . . . 7
3.2 Physical setup . . . 7
3.3 Colors . . . 7
3.4 Stimulus . . . 8
3.5 Experiment . . . 8
3.6 NEST . . . 9

4 Results **10**
4.1 Accuracy . . . 10
4.1.1 Edit distance . . . 10
4.1.2 Word comparison . . . 11
4.1.3 Accuracy method choice . . . 11
4.2 Data processing . . . 12
4.3 Results . . . 13
4.4 Expected results . . . 13
4.5 Analysis . . . 14
4.6 Further exploration . . . 14

5 Discussion **15**
5.1 Differences in setup . . . 15
5.2 Future work . . . 16
5.2.1 Replicating the original experiment more closely . . . 16
5.2.2 Diverging from the original experiment . . . 16

6 Conclusion **17**

Bibliography **17**

Abstract

We attempt to replicate results from existing work which finds that reading speed can be increased by introducing various coloring schemes to the text. We use a new experimental setup that allows taking multiple measurements and calculating average reading speed from the resulting data. We replicate the conditions from previous work and perform the experiment on four participants.
We find that the results do not match our expectations. The results have no statistical significance. However, the expected effect is not observed when comparing reading speeds for the same person. We hypothesize that the difference may be caused by the experimental setup or other differences in experimental procedure. The new way of measuring reading speed may have an impact, as well as some minor differences in controlled variables to the experiment we attempted to replicate. We suggest future experiments that can help establish the reason our experiment was unable to achieve the expected results.

1 Introduction

Written text is an extremely important and common way of passing information. Improving reading speed in humans can have great benefits in the speed and efficiency of information transfer. Hence, this research focuses on perceptual organization and its effect on reading speed. It has been proven by Pinna and Deiana [1, 2] that adding color to text can improve or deteriorate the reading experience depending on how it is applied.

Perceptual organization is an important factor while reading text. The organization of text into individual objects to differentiate words from each other is governed by Gestalt principles. These will be described in more detail in Section 2. However, it can be shown that color can be a strong factor in aiding perceptual organization and increasing reading speed [1, 3]. This study aims to replicate these results with a new experimental setup.
We achieve this by performing an experiment to investigate the effect of different coloring schemes on observers' reading speed. In doing so, we aim to validate a new experimental setup and allow further research in this area. Eventually, we aim to provide tools to answer the question of whether color luminance also affects reading speed.

This research will aim to investigate the area described above using an experiment with volunteers in order to gather data and answer the research questions as stated below. The experiment setup will be validated by attempting to replicate the results found by Pinna and Deiana [1]. Volunteers will be recruited in order to ensure data integrity and to provide a meaningful result.

1.1 Research questions

The main research question is:

Can an experimental setup using the `PsychoPy` Python library replicate previous results [1, 2] and provide new insights into the impact of perceptual organization on reading speed?

This research question can be divided into multiple sub-questions:

1. How much does chromatic iso-luminant variation induce a perceptual organisation, and how does this impact reading speed?
2. How do different accuracy metrics influence the experiment results?
3. Do the results from sub-question 1 replicate the expected results?

2 State of the Art

There have been few studies on the topic of perceptual organization and its impact on reading speed. According to our search as detailed in Section 2.4, there are only 2 papers on this topic, both of which are described in this section.

2.1 Inspiration

Pinna and Deiana [1] report that there is a significant positive effect of chromatic coloring on reading speed. They found that alternating per-word isoluminant colors can help with text separation and improve reading speed. They show that the Gestalt principle of similarity helps with perceptual organization and therefore, increases reading speed when used appropriately. They also show that it can be used detrimentally in order to deteriorate the reading experience and slow the reader down. These results are reported in the 2014 paper [1] as well as an earlier paper by Pinna [2] published in 2010.

Since the aforementioned studies used a set of isoluminant colors, they did not have the opportunity to test whether other color attributes, such as luminance, have a comparable effect. The lack of research in this direction inspired the topic of this study. By verifying a different setup, we can set the groundwork to explore this topic further.

2.2 Gestalt principles

The Gestalt principles, sometimes called laws, are rules of organization of perceptual scenes [4]. They were proposed in 1923 by Wertheimer [5]. They govern how humans perceive the world and how they are able to understand and reconstruct complex scenes from an input which is just a set of colored points. They relate to the concept of figure-ground separation, which explains how the presence of objects in an image induces a background and foreground separation, even though no such thing is present.

There are seven Gestalt principles, namely: proximity, common fate, similarity, continuity, closure, good form (Prägnanz) and past experience [4]. Some simple examples showing these effects can be seen in Figure 1. The interactions of these principles are complex and some can be stronger than others. These principles are very important when talking about human perception.

2.3 Related work

There have been studies exploring the effect of color on Gestalt principles in various fields. It is good to have an overview of these even though they are not directly related to reading speed. Ali and Peebles [7] show that a clever use of color leads to improved reader understanding of line charts due to

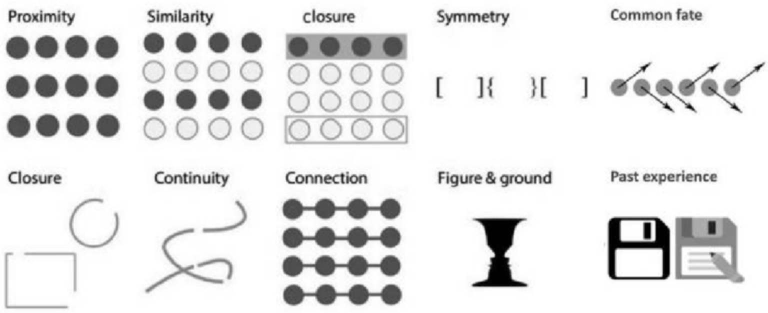

Figure 1: A comparison of accuracy methods. Image by Yalcinkaya and Singh [6]

Gestalt principles. They propose a novel way of coloring graphs to induce better perceptual organization, which in turn improves comprehension. O'Connor [8] finds that the use of color and contrast improves figure-ground separation in visual design and therefore, can help focus attention or distinguish parts of a whole. This shows that color and contrast can be used to enhance Gestalt principles and therefore, empower certain aspects of visual design, including text.

Zhang and Wang [9] explore the application of color in human-computer interfaces and propose rules to guide the use of color. This shows just how important the use of color can be in various visual applications and that attention must be paid to the use of color in order to maximize the efficiency of information transfer. They also show that certain interactions of color can cause interference in the form of visual illusions or other inaccuracies. This can have a negative impact on comprehension.

Abidi *et al.* [10] investigate psychological processing of color in order to improve weapon detection in X-ray scans. They develop several coloring schemes for this purpose and show that human operators make fewer errors when using an appropriate color scheme. They lay out guidelines for choosing color maps as well as finding limitations of the human perception that should be avoided.

Pinna *et al.* [3] investigate the Gestalt principle of similarity. They propose an adjacent principle of dissimilarity and investigate interactions between the two. They investigate the effect of these principles on reading speed using contrast to induce either a positive or a negative effect as shown in Figure 2. This is indeed an example of using luminance to improve readability, while using only the extreme values (black and white). This study does not attempt to quantify the effect as it simply uses its presence as an argument for its overarching point about a principle of dissimilarity. Therefore an investigation into the effect of luminance on reading speed can be viewed as an extension of this research.

Overall, we see that there has been research into the role of color in perceptual organization in various tasks. An effect has been found in all of the studies presented in this section. Therefore, it can be concluded that color plays a strong role in human perception and can be used to enhance reading speed. It has yet to be established whether this only applies to changes in hue or if luminance changes also induce a similar effect. The strength of this effect is also up for investigation.

2.4 Search terms

Three different search engines were used to find relevant literature and compose the current state of the art. Table 1 details the three engines and the search terms used in each. The search was conducted on `20.4.2024`.

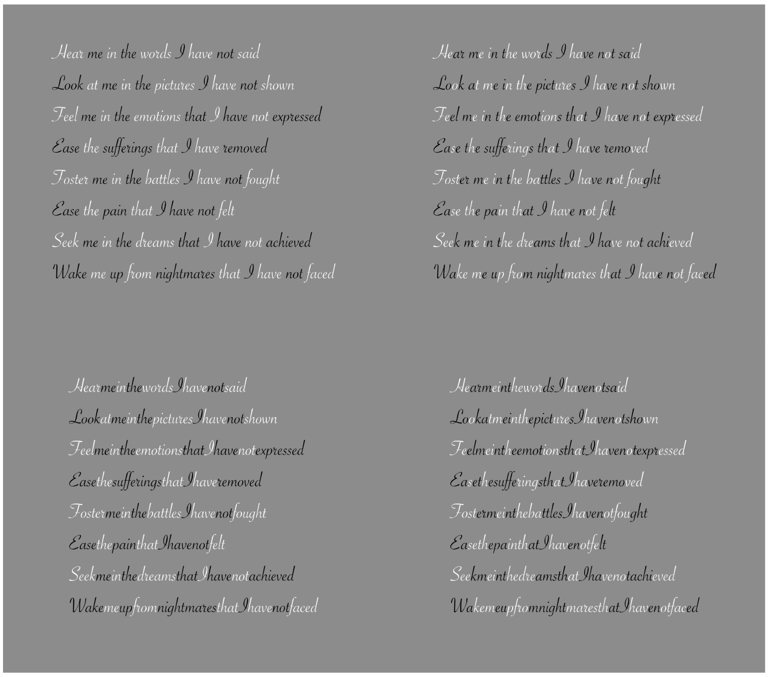

Figure 2: Contrast polarity and its influence on reading. Image by Pinna et al. [3]

Table 1: Search Terms

Search engine	**Search Terms**
WorldCat.org	chromatic color reading speed perceptual organization
WorldCat.org	kw:(gestalt perceptual organization color) AND kw:(reading) AND kw:(text)
WorldCat.org	kw:(gestalt) AND kw:(reading speed) OR kw:(reading rate)
Google Scholar	perception gestalt color reading text
Consensus.app	which factors impact reading speed
Consensus.app	gestalt principles and their impact on reading speed

These searches revealed very little literature on the topic of Gestalt principles in regards to reading speed other than the articles by Pinna and Deiana [1, 2], which inspired this research topic.

3 Method

Our experiment was performed using PsychoPy, a psychophysical experiment library in the Python programming language. The library provides tools for accurate and accessible psychophysical experiments. Our experiments were performed in controlled conditions in the Visualization Laboratory to which the University of Groningen kindly provided access. Details of our entire experiment setup are described below to ensure reproducibility. As a rationale for some choices, it is important to note that our experiment is attempting to find possible improvements to reading speed. Hence, choices were made so that the experimental conditions represent real reading conditions closely. Regular word spacing mirrors the standard reading situation in most cases. Pelli et al.[11] have proven that the crowding effect of surrounding text influences reading speed. We chose our experiment conditions to represent these effects, therefore providing a realistic reading experience.

3.1 PsychoPy

The library used is called `PsychoPy`. More information about this library can be found in the PyPi library listing. This library is implemented by Peirce *et al.* [12]. It allows the configuration and execution of visual experiments and provides tools for accurately setting up stimuli. This library, in conjunction with additional Python code, was used to run our experiment. The source code for our experiment can be viewed upon request.

3.2 Physical setup

Our experiment was conducted in the Visualization Laboratory room in the Bernoulliborg building of the University of Groningen. This room has a light-controlled environment where our experiment was run. The display used was a `LG OLED55C21LA`, a 55-inch television from LG using OLED panel technology. The display's maximum brightness was 170 cd/m^2. To ensure constant distance from the display, a chin and forehead rest was used. This stabilized the participants' head position and allowed for the distance to the display to be kept constant. The distance to the screen was 93 cm. A state-of-the-art computer was used to run the software.

3.3 Colors

The 5 colors from Pinna and Deiana's [1, 2] experiment were used. They are listed below with their respective CIE x, y chromaticity coordinates:

1. Green: (0.34, 0.51)
2. Purple: (0.35, 0.24)
3. Brown: (0.49, 0.40)
4. Blue: (0.17, 0.19)
5. Red: (0.55, 0.37)

These colors were chosen in an attempt to replicate the results obtained by Pinna and Deiana. They are not perceptually equidistant. The colors were used in the order listed above. The color brightness was chosen to be 50% of the maximum brightness. The background was white.

The coloring schemes chosen were the same as in their experiment. Examples with and without spacing (even though our experiment used spacing) can be seen in Figure 3:

1. Per-word. This scheme colored each word a different color. Shown in c) and d).
2. Per-letter. This scheme colored each letter a different color. Shown in e) and f).
3. Per-half word. This scheme colors the first half of each word a different color from the second half. The first half of a word always has the same color as the second half of the previous word. Shown in g) and h).

4. Monochromatic. This condition serves as a control to compare the other results against. It uses black colored text (with the same luminance as the other colors, making it gray). Not shown.

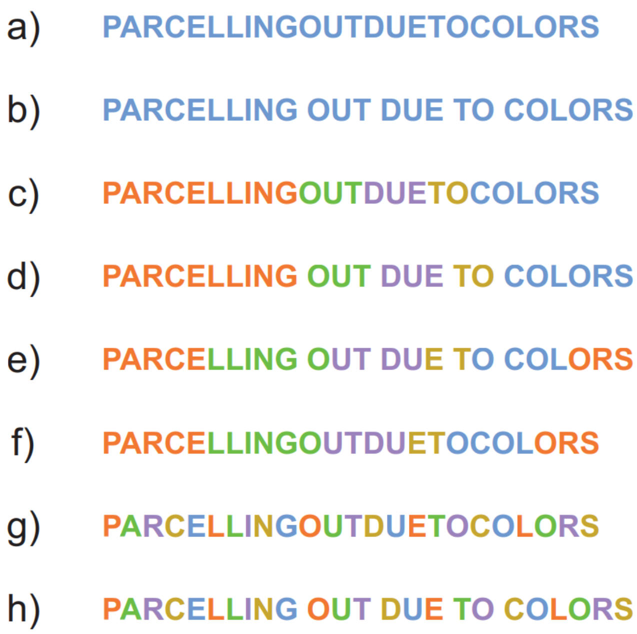

Figure 3: The four coloring schemes used in our experiment. Image taken from Pinna and Deiana [1]

3.4 Stimulus

The stimuli were sentences from the `Corpus of Linguistic Acceptability` [13] dataset. Sentences were selected at random. Sentences were chosen so that they would never wrap to the next line to avoid the delay caused by the observer shifting their gaze. The font height was set to 1 degree of the observer's field of view, calculated using the PsychoPy library given the display size and resolution. Filler text above and below the stimulus was used to simulate crowding. The filler text was the same font size as the stimulus. This text was also colored using the same scheme as the stimulus. The filler text was a random continuous passage from the well-known Lorem ipsum text. The color luminance was set to 50% of the maximum brightness, both for colored text as well as for monochromatic text. For an example stimulus, refer to Figure 4.

3.5 Experiment

Our reading speed experiment consisted of multiple trials. There were six warm-up trials at the beginning of our experiment. The participants were informed about the presence of these warm-up trials. Data recorded during warm-up were excluded from the final analysis. The coloring schemes were alternated in an interleaved fashion.

A trial consisted of the following steps:

1. The desired values of parameters were chosen. The target reading speed was the only input parameter in our experiment.

2. A random sentence to be shown as the stimulus was chosen from the dataset

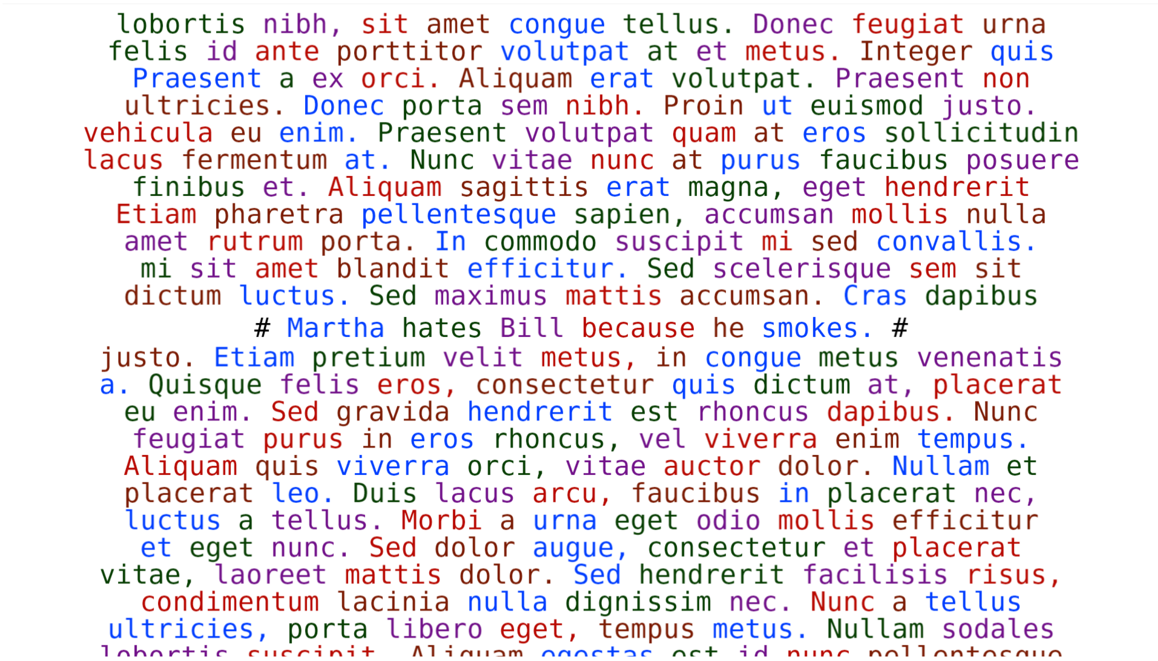

Figure 4: A screenshot of a stimulus in a trial with the per-word coloring scheme. The sentence 'Martha hates Bill because he smokes' is the correct response.

3. A calibration stimulus was shown to the participant for 1 second. This stimulus was a # symbol, and it was positioned exactly where the # symbol denoting the beginning of the sentence would be shown later.

4. The stimulus is presented for a fixed amount of time depending on the target reading speed in words per minute (WPM). Bodies of filler text are shown above and below the stimulus in order to simulate the crowding effect [11] when reading a longer block of text. The stimulus sentence is surrounded by two # symbols on either side. See Figure 4 for an example of a stimulus.

5. The stimulus is removed, and the participant is asked to type in the sentence they were just shown. They get an unlimited amount of time for this.

6. Once the participant submits their response, the accuracy is calculated and recorded.

7. The next trial begins.

For a better understanding of the experimental procedure as well as any details which may be unclear, the Python code used to run the experiment can be viewed upon request.

3.6 NEST

The NEST framework was used to select variable values for trials. NEST [14] is a framework that tries to approximate the psychometric function using a small neural network and then picks the next trial at the decision boundary of the network. This allows for faster data gathering and allows more precise estimation of the psychometric function with fewer trials needed. In this experiment, Fisher convergence was used to estimate when the neural network converged. A running difference average of the last 5 Fisher energy values was used to determine the termination condition. A separate NEST instance was running for each interleaved experiment.

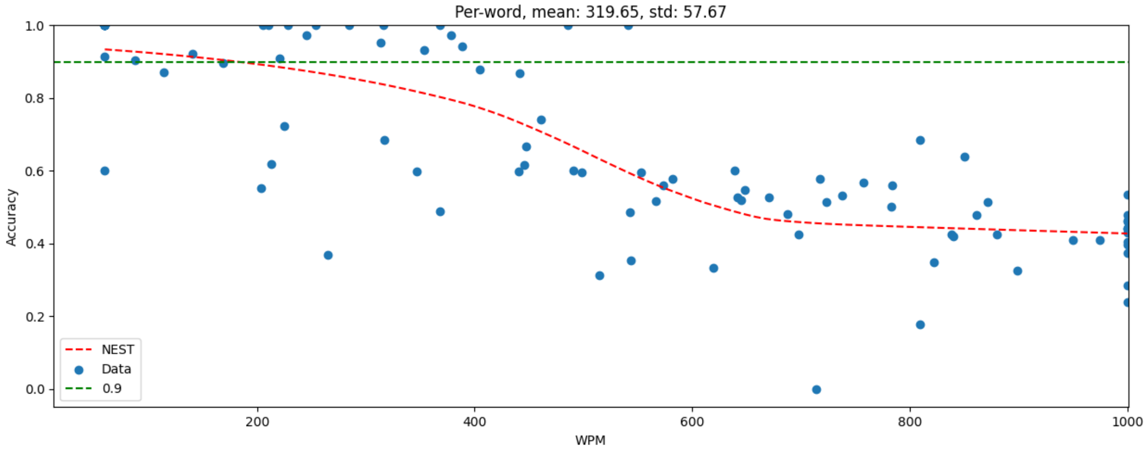

Figure 5: The results from one sub-experiment. The title provides the color scheme name and the computed reading speed in WPM.

4 Results

The experiment was performed on four participants. The participants were 75% male, 25% female, aged 23-38. Each participant filled in a consent form. For each participant, four sets of data points were obtained. Each set represents the relation between the target reading speed in words per minute (WPM) and the accuracy of the observer for a given coloring scheme. A visualization of one set of such points can be seen in Figure 5. This represents the per-word coloring scheme subexperiment. The average reading speed of the observer was calculated to be 320 WPM with a standard deviation of 58 WPM. The calculation is detailed in Section 4.2

4.1 Accuracy

Two different accuracy measures were considered for their different benefits. An approach using edit distance as well as an approach using word comparison were tested.

4.1.1 Edit distance

The accuracy is calculated by the following process:

1. Convert all letters to lowercase
2. Remove all punctuation
3. Remove all accents on letters
4. Compute the Levenshtein edit distance of the given answer and the real answer
5. Divide the edit distance by the length of the real answer to get a percentage accuracy
6. Invert the result (low edit distance means high accuracy and vice versa)
7. Clamp the result to [0, 1].

The Levenshtein edit distance is calculated using the formula introduced by Levenshtein [15]. To calculate it, the Levenshtein library was used. The clamping step is necessary because the result can be lower than 0. This can happen if the observer inputs an incorrect string longer than the correct answer. For example:

$$\begin{aligned} \text{True answer} &= \text{'yes'} \\ \text{Observer answer} &= \text{'wrong'} \\ \text{Levenshtein distance('yes', 'wrong')} &= 5 \\ \text{Normalized distance} &= 5/3 = 1.67 \\ \text{Inverted normalized distance} &= -0.67 \\ \text{Clamped inverted normalized distance} &= 0 \end{aligned}$$

Since the division must always be by the length of the true answer, the clamping step is necessary to avoid this problem and to ensure that the result is between 0 and 1 inclusive. We clamp instead of scaling because any value below 0 indicates that the answer was completely wrong.

4.1.2 Word comparison

An alternative accuracy measure was considered. The alternate accuracy measure used a simpler process:

1. Repeat steps 1-3 from the edit distance measure as described above
2. Split the answer and true answer strings using the space character as a delimiter
3. Iterate over all words in the given answer, count how many of them appear in the true answer
4. Compute the accuracy as the fraction of correct words divided by the total number of words in the true answer

Note that the word-based calculation does not take the order of words into account. This is desirable because a reader may be able to read an entire sentence and understand it even if the words are not in the correct order. This can also help avoid problems with grammatically incorrect sentences.

4.1.3 Accuracy method choice

It was noticed while prototyping our experiment setup that the word comparison method tended to produce more noisy data due to misspelled words. As it cannot accept misspelled words and immediately considers the entire word as wrong, it would often find a very low accuracy due to this problem. For example, the sentence `John reallyh ates Mary` obtains an accuracy of 50% even though it is obviously just one spelling error away from being 100% accurate. With the edit distance method, the same sentence receives a 91% accuracy.
Another problem with the word comparison method is that it does not take word length into account. If the reader is able to read the first few words of a sentence because they are very short, they can achieve a relatively high accuracy score even though they did not read the fraction of the sentence the accuracy result suggests.

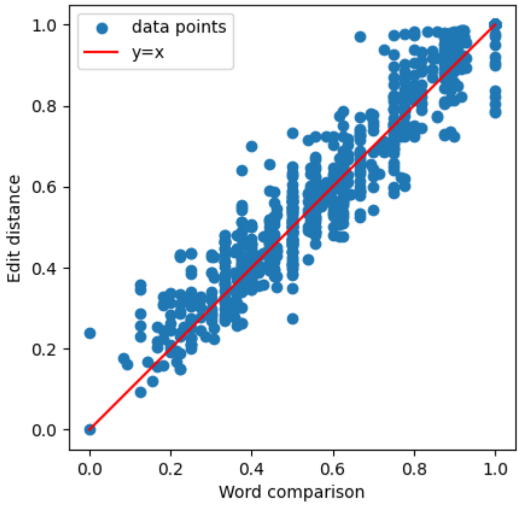

Figure 6: A comparison of accuracy methods

To compare the two methods, the answers from our experiment were gathered and evaluated using both methods. Figure 6 shows the comparison of all answers given by the partcipants. The total number of answers is 1088.

From the plot, we can see that the two methods are clearly linearly correlated. However, if we investigate a specific subexperiment of a specific observer, we notice that the spread in values tends to be higher for the word comparison metric. This is shown in Figure 7. Upon inspection of the box plot, we observe that the edit distance method tends to have a higher average as well as a smaller interquartile range. A similar outcome was observed in our other results as well. This shows that the word comparison accuracy measure causes us to predict lower reading speeds due to flaws in the way it is computed.

In light of these findings, it was chosen to use the edit distance metric due to its robustness against spelling errors. This metric introduces less erroneous behavior to the data and therefore allows a better interpretation of the data.

4.2 Data processing

To extract the average reading speed from a graph such as the one shown in Figure 5, the product of accuracy and expected WPM was used. This measurement method is analogous to the one used by Legge *et al.* [16] in their reading speed experiment. The accuracy is expected to be a measure of how much of a shown sentence the observer was able to read. Therefore the actual reading speed of the observer can be approximated by multiplying the target reading speed by the accuracy achieved at that speed. However, we must keep in mind that accuracy is clamped to the interval [0, 1], so we sample only the part of the psychometric curve which is in the interval [0.1, 0.9] to avoid using regions where the clamping interferes. This is shown in Figure 5 as the green line. If we sample the obtained psychometric function in regular intervals, we can obtain a reasonable average for the reading speed. Since it is not a linear function, there will be differences at each sampling point, and therefore we obtain a mean reading speed and a standard deviation. The sampling interval was chosen to be 10 WPM.

The NEST output function was used as the psychometric function. A logistic function was fitted to

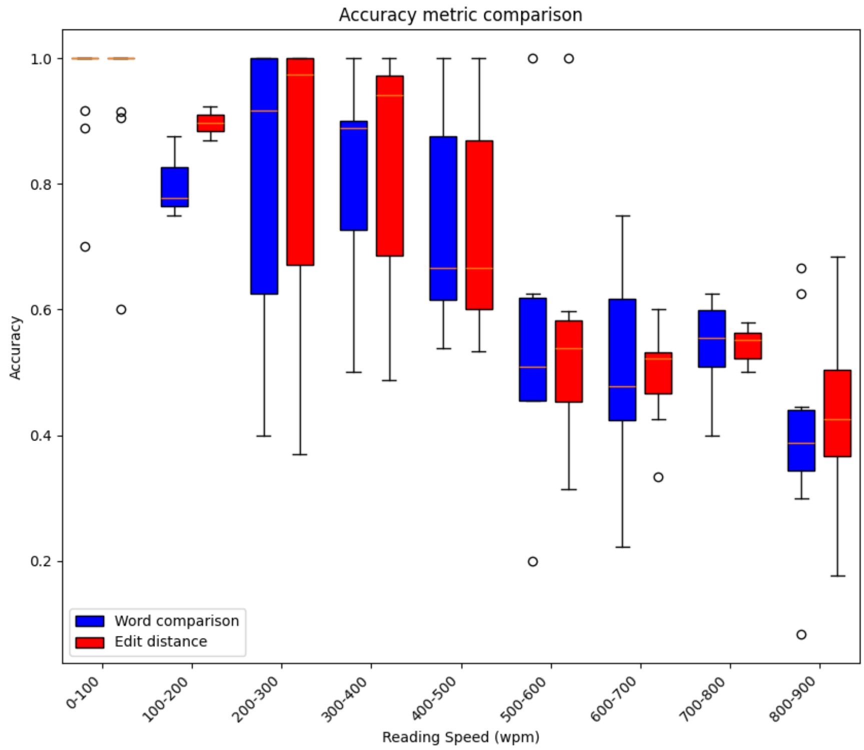

Figure 7: A comparison of accuracy methods for experiment results

the data as well to see whether it would align with the NEST output. In all cases, the fitted logistic function was in almost perfect agreement with the NEST output.

4.3 Results

Our experimental results were gathered as described and plotted to compare against the expected results. The plot can be seen in Figure 8. It is immediately obvious that our results show much higher variance and that our results do not follow the expected results, as shown in Figure 9. Note that the two plots do not have the same units on the y-axis, however reading easiness should correspond to a faster reading speed. This can be seen in Pinna and Deiana's experiment with no spaces, where both reading speed and reading easiness are shown [1].

4.4 Expected results

Figure 9 shows the results obtained by Pinna and Deiana [1]. The plot shows the reading easiness measure for varying sizes of inter-word spacing. The 1 inter-word line (black) corresponds to their experiment which we are attempting to replicate. It is important to note that Pinna and Deiana do not report reading speed measurements for the specific conditions that were tested in our experiment. The plot represents a subjective measure of reading easiness, as measured by allowing observers to pick which color condition they found easiest to read. Some criticisms of their work can be found in Section 5.

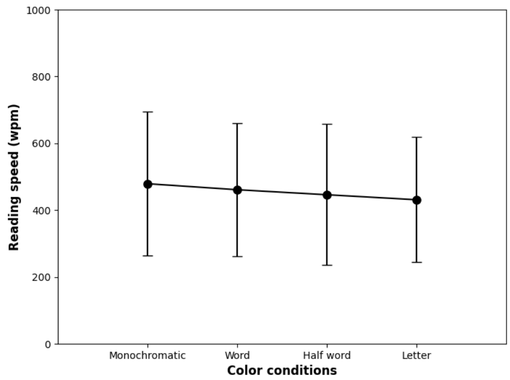

Figure 8: Our results showing the impact of text coloring on reading speed

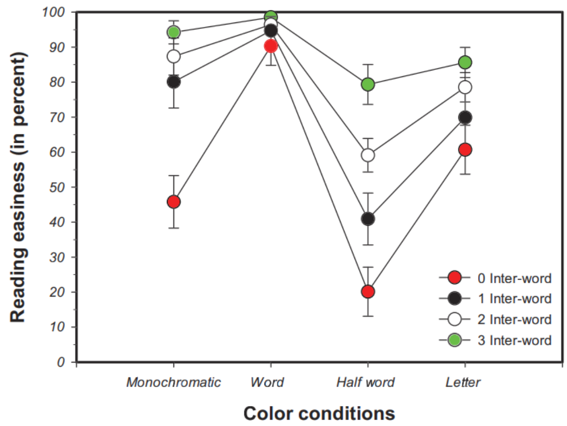

Figure 9: Reading easiness as reported in [1]

4.5 Analysis

Statistical analysis showed that the results are not statistically significant, as is obvious from Figure 8. However, looking at individual experiment results, it was found that for 3 out of 4 observers, the control experiment had the fastest calculated reading speed. In the last case, the control had the slowest recorded reading speed, even slower than the half-word scheme. The variances are too high for any of these differences to be statistically significant. Upon seeing these results, it was decided to abort our experiment and review the possible reasons for the mismatched results. This is the reason why our experiment was performed only on four participants. More on this topic will be elaborated in Section 5.

4.6 Further exploration

Once it was known that our experiment had failed to replicate the expected results, it was found that some experimental variables did not replicate the conditions from Pinna and Deiana's experiment perfectly. These are listed below:

1. The order of colors used was different from the order used in their experiment. This was caused by an oversight. Their paper specifies the colors in a different order than the one shown in the figures. We assume that the order shown in the figures is the one that was actually used.

2. Our luminance did not match their luminance perfectly. Since the display used in our experiment is brighter, the luminance of the background and colors were both slightly higher.

3. Pinna and Deiana never specify the font size used in their experiment. They specify: "The overall sizes of the figures were each ∼5 deg." [1]. It is unclear what this means since the aspect ratio of their text stimulus is not 1:1. Therefore it is impossible to know whether the horizontal or vertical size is 5 degrees. They also never show the entire stimulus, so it is not possible to know how many lines of text it was.

4. The relative distance from the screen was not the same. Taking the size of the display into account, we would need to place observers 140 cm from the display to replicate the horizontal viewing angle of the display width. The setup we used did not allow this distance to be achieved.

5. The duration of our experiment was not the same as theirs. It is not mentioned how long Pinna and Deiana's experiment took. Our experiment took around an hour per participant (with breaks every 30 minutes). This may have strained the participants and skewed the results.

6. Pinna and Deiana do not use black in their monochromatic experiment from 2014, they instead use one of the 5 colors defined in their paper [1]. However, in their experiment from 2010, Pinna *et al.* do use gray color as the monochromatic control [2]. We used black in our experiment due to the similarity to real reading conditions.

Upon discovering these differences, which were originally thought to be too minor to influence the results, a revised experiment was created to correct for them. However the revised version of our experiment ran into an efficiency problem when attempting to decrease the font size. This problem made rendering each experiment trial take 15-30 seconds, making the revised experiment infeasible. Time constraints did not allow the revised experiment to be made functional for smaller fonts, so it is left as a future exploration.

5 Discussion

This section will focus on the reasons behind the obtained results and possible causes for the resulting mismatch. Several differences can be listed, both in our way of measuring reading speed as well as differences in details such as color, distance from the screen and other minor controlled variables.

5.1 Differences in setup

Firstly, there is the actual procedure. Pinna and Deiana measured reading speed by having observers read an entire short story of around 270 words in Italian out loud. This task has its benefits and downsides. It simulates a reading speed average more accurately due to the longer text. It also forces the observer to shift their gaze when encountering line breaks and includes this delay in the total reading time. Reading aloud may also affect reading speed, and possibly have different effects on each individual. The language difference may also be playing a role.

Pinna and Deiana [1] do not actually present reading speed graphs for the exact task that we reproduced in our experiment. They do not give a reason for this. It is unclear whether this is because they found no effect or for other reasons. In the earlier paper [2], Pinna *et al.* do find a statistically significant difference depending on color scheme in this exact task. That study uses the same setup with some minor differences, such as an unexplained change in color luminance. Therefore we assume that in the later study from 2014, the same effect was found. It is, therefore, unclear why the reading speed graph is not included in favor of the reading easiness graph.

On the contrary, our experimental method aims to obtain more precise measurements by eliminating gaze shift and measuring 'straight-line speed'. This method also allows us to take multiple measurements, because we observed large variations in accuracy at the same target reading speed. Overall, this suggests that observers' reading speed is influenced by outside factors such as fatigue, lapses in attention, focus, and other possible distractions. This may cause large variations in the data. Pinna and Deiana do not mention any repeated trials, it seems that they only had each participant read the text once. Our choice of random stimuli instead of repeating the same text allowed us to conduct multiple trials per observer and then attempt to fit a psychometric function to the resulting data.

The method used to measure reading speed in our experiment produces a lot of variation. Due to the fast nature of the task, small errors like lapses in focus or delays caused by reaction time cause relatively large variation in the data. The method could be adapted to present longer pieces of text and possibly measure reading speed multiple times per paragraph to alleviate the errors from this source. An example of such a method could be to present a longer piece of text, having the observer press a button every time they encounter a specific type of word. Stimuli for this task would be harder to generate. A Large Language Model might be useful to generate stimuli for an experiment like this. This would allow for multiple reading speed measurements per trial while extending the total reading time so that small errors do not cloud the data.

Another consideration which should be made when measuring reading speed is the justification of text. Stiff found in 1995 that there was no impact of justified or unjustified line endings on reading speed [17]. However, in our setup the stimulus sentence was not aligned with the filler text, since we focused on keeping the sentence without line breaks. If the sentence was shorter than the width of the filler text, we did not adjust the filler text width, we simply let the sentence be inset on both sides. This minor detail may also play some role in the results we found, and should be taken into consideration in future experiments.

In theory, this setup should be able to produce accurate measurements. The read-and-repeat task focuses on both on reading speed and reading comprehension. The setup also simulates real conditions by presenting sentences from a dataset that contains regular English sentences without contextual meaning as opposed to the short story used by Pinna and Deiana.

There were some minor differences in the setup, as listed in Section 4.6. These may have influenced the results to some extent.

The observed variation in our experiment did not allow us to draw any definitive conclusions. The observed reading speed varied from 250 WPM to over 800 WPM in various observers and color conditions. Naturally, this variation could be reduced by having more participants in our experiment. However, such large disparities show that reading speed varies greatly per individual.

Overall, it is unclear which of these conditions may have caused the disparity in our results. To investigate further, the next section suggests some ways to narrow down the source of error.

5.2 Future work

As mentioned earlier in this paper, multiple proposed future investigations may help us understand the results obtained in this project.

5.2.1 Replicating the original experiment more closely

Several improvements are mentioned in Section 4.6. A new experiment with these revised conditions to follow Pinna and Deiana's [1] experimental procedure as closely as possible would provide more insight into the reasons behind our experiment's failure. The duration should be shortened and the conditions followed as closely as possible.

5.2.2 Diverging from the original experiment

It could also be interesting to find whether altering the conditions can find the desired effect. Possibilities for these include more saturated colors, different colors, different number of colors used, varying font size, varying luminance, etc. It would be interesting to find whether any combination of these conditions can have an effect on reading speed. This would be difficult to achieve as there are many combinations of these variables. The NEST framework [14] could be utilized for this purpose.

Overall, our experimental setup allows testing of multiple variables simultaneously. It can be used for multiple extensions of this work. The source code is available upon request and may be used to execute future work ideas such as the ones mentioned above.

6 Conclusion

It was not possible to replicate the results obtained by Pinna and Deiana with the experimental setup described in this report. The scope of this project did not allow for further investigation into the reasons behind this. We have some ideas why the results are not what we expected, and a detailed explanation of those can be found in Section 5.
The results of our experiment showed too much variance to reliably answer the research questions. It was noted that individual results did not follow the expected outcome either. Overall, it is unclear whether this is caused by some experimental conditions being different or whether our different way of measuring the observers' reading speed simply cannot replicate the results obtained by Pinna and Deiana.
There are multiple improvements that could be made in order to clarify the results of our experiment. As mentioned in the Section 4.6, there are some variables that may have contributed to the results. With more time, it would also be beneficial to invite more participants to our experiment to reduce the observed variation in reading speed.
Overall, even though our experiment did not achieve its intended goal, it provided valuable insight into the experimental procedure and multiple future research topics as extensions to this project.

Bibliography

[1] B. Pinna and K. Deiana, "New conditions on the role of color in perceptual organization and an extension to how color influences reading," *Psihologija*, vol. 47, no. 3, p. 319–351, 2014.

[2] B. Pinna, A. Uccula, and M. Tanca, "How does the color influence figure and shape formation, grouping, numerousness and reading? the role of chromatic wholeness and fragmentation," *Ophthalmic and Physiological Optics*, vol. 30, p. 583–593, Aug. 2010.

[3] B. Pinna, D. Porcheddu, and J. Skilters, "Similarity and dissimilarity in perceptual organization: On the complexity of the gestalt principle of similarity," *Vision*, vol. 6, p. 39, June 2022.

[4] D. Todorovic, "Gestalt principles," *Scholarpedia*, vol. 3, no. 12, p. 5345, 2008. revision #91314.

[5] M. Wertheimer, "Untersuchungen zur Lehre von der Gestalt. II," *Psychologische Forschung*, vol. 4, no. 1, p. 301–350, 1923.

[6] M. Yalcinkaya and V. Singh, "VisualCOBie for Facilities Management: A BIM integrated, Visual Search and Information Management Platform for COBie Extension," *Facilities*, 06 2018.

[7] N. Ali and D. Peebles, "The effect of gestalt laws of perceptual organization on the comprehension of three-variable bar and line graphs," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 55, p. 183–203, July 2012.

[8] Z. O'Connor, "Colour, contrast and gestalt theories of perception: The impact in contemporary visual communications design," *Color Research and Application*, vol. 40, p. 85–92, Dec. 2013.

[9] Y. Zhang and J. Wang, "Human-computer interaction design strategy for color identification on multi-media visual interface," in *2009 IEEE 10th International Conference on Computer-Aided Industrial Design and Conceptual Design*, pp. 1874–1877, 2009.

[10] B. Abidi, Y. Zheng, A. Gribok, and M. Abidi, "Improving weapon detection in single energy x-ray images through pseudocoloring," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 36, p. 784–796, Nov. 2006.

[11] D. G. Pelli, K. A. Tillman, J. Freeman, M. Su, T. D. Berger, and N. J. Majaj, "Crowding and eccentricity determine reading rate," *Journal of Vision*, vol. 7, pp. 20–20, 10 2007.

[12] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv, "Psychopy2: Experiments in behavior made easy," *Behavior Research Methods*, vol. 51, p. 195–203, Feb. 2019.

[13] A. Warstadt, A. Singh, and S. R. Bowman, "Corpus of linguistic acceptability." http://nyu-mll.github.io/cola, 2018.

[14] S. Bruin, J. Kosinka, and C. Tursun, "NEST: Neural Estimation by Sequential Testing," 2024.

[15] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals.," *Soviet Physics Doklady*, vol. 10, pp. 707–710, feb 1966. Doklady Akademii Nauk SSSR, V163 No4 845-848 1965.

[16] G. E. Legge, D. G. Pelli, G. S. Rubin, and M. M. Schleske, "Psychophysics of reading—i. normal vision," *Vision Research*, vol. 25, p. 239–252, Jan. 1985.

[17] P. Stiff, "The end of the line: A survey of unjustified typography," *Information Design Journal*, vol. 8, p. 125–152, Jan. 1995.