

Visualisierungsinstitut der Universität Stuttgart

Master Thesis

Evaluation of 2D Combination of Eye-Tracking Metrics for Task Distinction

Caglar Mehmet Metin

Course of Study: Computer Science

Examiner: Prof. Dr. Daniel Weiskopf

Supervisor: Dipl. Inf. Nils Rodrigues

Commenced: October 15, 2018

Completed: May 15, 2019

Abstract

Eye-tracking techniques enable researchers to observe human behaviors by using eye tracking metrics. Machine learning is one of the techniques used in task inference. However, in our research in order to decrease the effort to analyze the task inference, we consider two combinations of different metrics on a two-dimensional scatter plot. Also, we analyze the data with K-Means clustering and correlation analysis to determine the task inference. Two-dimensional scatter plot let the analyst interact with the data in a better manner. In this thesis, we reduced the metrics dimensions, for example, calculating the mean value of the fixation durations that gave us a single value. We examined a few metrics such as crossings of saccades, first fixation duration after the onset of a stimulus, fixation duration mean, and fixation duration median. Furthermore, we created some custom metrics specifically for this research to analyze the tasks for the participants better. Next, we developed a simple game. In the game, there were three game modes for building distinctive gaze behavior. Those game modes include changes in the color tint information, size changes of the stimulus, and as a control mode, a text-only representation which does not contain any color or size differences. Finally, we made a study with six participants. They played our game to give us a dataset which we can work in the analysis with K-means clustering. The results were promising and helpful in distinguishing human behavior on different tasks. However, this research is not enough for task inference, and there are further improvements which could achieve a better result than the current state.

Contents

1	Introduction	11
2	Related Work	13
3	Fundamentals and Eye Movements Metrics	17
4	Study Design	25
5	Implementation	29
5.1	Game Implementation	29
5.2	Analysis Implementation	34
6	Pilot Study	45
6.1	Parameter Selection	45
7	Conducted Study	47
7.1	Participant’s background	47
7.2	Before the Experiment Started	47
7.3	Beginning of the Study	47
7.4	After the Study	49
8	Results	51
8.1	Participant’s comments related to the study	51
8.2	Clustering and Correlation Results Analysis	52
9	Conclusion	57
10	Appendix	59
10.1	Questionnaire Form	59
10.2	Consent Form	65
	Bibliography	69

List of Figures

2.1	Eye movement metrics visualized by a clustered matrix along with a dendrogram that shows the structure of the hierarchical clustering: in part a) metric M1 (path transition), in part b) metric M2 (fixation duration), in part c) metric M3 (saccade length), and in part d) metric M4 (completion time).The figure taken from Kumar et al. [KNB+16]	13
2.2	Kumar et al. present an overview of all the variables in a parallel coordinate diagram. The color information indicates distinct clusters. Rearrangement of metrics reveals a concrete correlation among attributes in part b and contradictory correlations in part c. The figure taken from Kumar et al. [KNB+18]	14
3.1	Saccadic curvature representation. The figure taken from [NH06]	18
3.2	Fixation duration representation. The figure taken from [HNA+11]	19
3.3	Start End Fixation calculation representation. Red lines shows the scanpath, blue circles represents the fixations	21
3.4	Saccadic direction calculation represented. Which illustrates the degree between the two gaze positions which successive one and after. The figure taken from [HNA+11]	22
3.5	Saccadic direction histograms for reading data. First a classical linear histogram, then an angular histogram (rose plot) for the same data. Bin size 11.25 degrees. The figure taken from [HNA+11].	23
4.1	The custom color scale	26
5.1	Game mode 'Number' representation	30
5.2	Game mode 'Size' representation	30
5.3	Game mode 'Color' representation	31
5.4	Game steps from starting until the game ends	31
5.5	Representation of the cluster values. Red values were blocked and the blue values were chosen. Arrows showed the blocked elements after a value was selected.	35
5.6	User interface representation	36
5.7	Clustering Result	37
5.8	Game Json file representation	38
5.9	Two blinks in a velocity plot. The figure taken from [HNA+11].	39
5.10	Idealized gaze position, velocity, and acceleration profile over time showing one saccade in-between the end of a fixation (to the left of the saccade). The figure taken from [HNA+11].	40
6.1	1) Inferno, 2) Magma, 3) Plasma and 4) Viridis respectively. The figure taken from [Jou15].	45

8.1 Two cluster graph side by side, left graph shows the participant's individual results
and the right graph shows all the participants combined results 54

List of Tables

5.1	CSV gaze file representation	32
8.1	Cluster result table	55
8.2	Correlation result table	56

1 Introduction

Eye-tracking techniques have allowed researching on people's behaviors [KNB+16] [KNB+18] [PPKS02]. Direct visualization of the scanpaths can be helpful to investigate the eye movements, but many of the analyzation tools depend on the eye movement metrics [HNA+11]. Previous research already investigated detailed one-dimensional metrics [HNA+11], but sometimes a single dimension is not enough to find groups of similar behavior in the eye movement data. Combination of different metrics can be helpful to distinguish this similar behavior. To be able to create a distinctive gaze behavior, researchers create different tasks for their participants, such as recording the eye movements while the participant was reading a paragraph. Moreover, machine learning is giving promising results on the task inference [PPKS02] [TVPK13]. However, for data analysis, it is hard to understand how machine learning techniques were reaching these results. Because machine learning is working like a black box which blocks investigation on the inner working mechanisms while the algorithm creates a relation between the different information. For decreasing the effort on researching on the task inference, we used more straightforward methods such as K-mean clustering and Pearson correlation.

In this research, we studied eye tracking metrics to distinguish humans behavior on different tasks. The tasks for the person could be comparing sizes, reading numbers or comparing colors. There was a study with the participants, for the study we developed a game for creating these distinctive gaze behaviors. After created an analysis tool to evaluate the collected data from the participants. As the other goal of the study, the results of the metrics should be well visualizable which lets the data analysts work on the metrics more efficiently. Commonly, 2D representation of the results in the scatter plot is easy to work with. This provides a pleasant working environment for data analysts. That is why using more than two dimensions was not considered in this research. To be able to represent two metrics at the same time in the two-dimensional metrics there was a need for reducing the metrics dimension. Because most of the eye tracking metric calculations are actually a time series. Such as the participant can be fixated multiple times in a given reading task and all the fixations have different durations. Then we need at least three dimensions to represent the combination of the two metrics. However, that would not be a good fit for the analysts to investigate the data. By reducing the dimension, it can be fit two metrics at the same time to a two dimensional scatter plot. In this study, we reduced the dimensions such as taking the mean value of the fixation durations which gave us a single value. The metrics for the analysis are chosen according to their ability to cover a wide range of eye movements, which lets the researcher allow to detect the low-level tasks.

This thesis holds several chapters. First, in the related work chapter, we investigate the previous researches to show what has been achieved in using the eye tracking metrics for the task inference. Then in the Fundamentals and Eye Movement Metrics chapter, we investigated the related metrics with our research. Also, the relation of the metric's to the task distinction reviewed in this chapter. In the Study Design chapter, we discussed how the game should be design to collect distinctive gaze behavior from the participants. In the Implementation chapter, we first explained how we developed

the game which we used in the conducted study. Also, the development of the analysis tool detailly explained. Moreover, the collected gaze data's structure described. We collected the gaze data as a raw gaze recording. Which contains nothing but the X and Y coordinates of the eye on the visual area and the timestamps which belong to every recording. All the metric values are calculated from the raw data then the K-means clustering and correlations of the metric combinations explained in this section. After in the Pilot Study chapter, we explained the findings from the pilot study. Which let us decide the hardness level of the game and the choice of the parameters which arranges the color and size information which belongs to the tasks. Then in the Conducted Study chapter, we explained every step in the conducted study, from welcoming the participant until the participant left the lab. In the Results chapter, we explored the combinations of metrics in an effort to find the ones with which we were able to cluster the gaze data according to the task modes. The best combinations and the least combinations of the metrics shown in the cluster result table. Also the correlations between the metrics shown. The results showed promising outcomes on distinguishing human behavior on different tasks even though K-mean clustering was a more straightforward method then machine learning techniques. However, in the current state, our research was not enough for task inference.

2 Related Work

In various researches, many of the eye movement metrics recorded and visualized. Kumar et al. [KNB+16] used the matrix to show the results of the multiple metrics. Kumar et al. [KNB+18] used the parallel coordinates to utilize and present an overview of the metrics. However, it is not easy for beginners to understand the results of these metrics in their visualization environment and as well as seeing the patterns.

For instance, in the research presented by Kumar et al. [KNB+16], they described a visualization technique based on the matrix to compare the data for eye movement visually. This matrix representation is shown in the Figure 2.1. To achieve this visualization, a group of scan path trajectories is first processed and transformed into matrices describing similarities and differences of eye movement paths. In order to keep the rendered plots simple, they visually encoded the generated data set into the cells of the matrix. Apart from just incorporating one individual metric of the dataset into a matrix cell, they extend this standard visualization by a dimensional stacking approach supporting the display of several of those metrics integrated into one matrix cell [KNB+16].

It supports the display of various of those matrices incorporated into matrix cell. To improve the readability further on and finding a pattern among these metric values, Kumar et al. researched metric based clustering and more interaction techniques to manipulate the data to navigate it. To demonstrate the usefulness of the system, they applied it to eye movement dataset about the reading behavior of metro maps. They also discussed the limitations and scalability issues of the methods they used.

It was concluded that the representation would be the best with the data of multiple combinations regarding the values in the used two-dimensional plots. It will make more satisfactory results which will enable the readers to have a better understanding of the results.

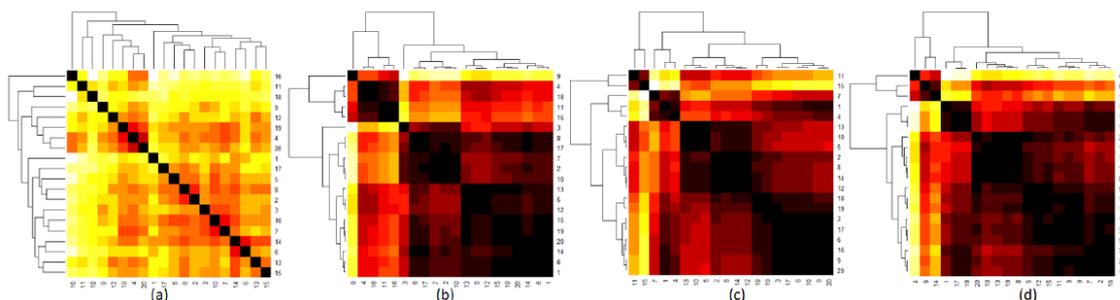


Figure 2.1: Eye movement metrics visualized by a clustered matrix along with a dendrogram that shows the structure of the hierarchical clustering: in part a) metric M1 (path transition), in part b) metric M2 (fixation duration), in part c) metric M3 (saccade length), and in part d) metric M4 (completion time). The figure taken from Kumar et al. [KNB+16]

2 Related Work

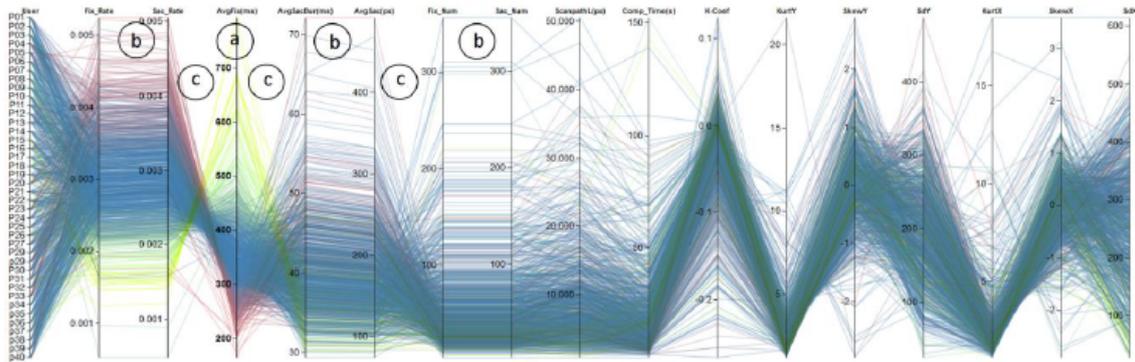


Figure 2.2: Kumar et al. present an overview of all the variables in a parallel coordinate diagram. The color information indicates distinct clusters. Rearrangement of metrics reveals a concrete correlation among attributes in part b and contradictory correlations in part c. The figure taken from Kumar et al. [KNB+18]

In another research presented by Kumar et al. [KNB+18], they derived eye-tracking matrices of the participants from their recorded eye tracking data. Their method utilized two visualization concepts. In the first concept, parallel coordinates were utilized to present an overview of used metrics, also their interactions and similarities. These parallel coordinates help to select the appropriate metrics which demonstrates the characteristics of eye tracking. They have visually encoded the matrix data set that was generated into cells of the similarity matrix of participants, to keep the matrix visualizations comprehensible and straightforward.

In Figure 2.2 the metrics shown using parallel coordinates. The axes show the metrics and giving the data analysis opportunity to shift these axes. By shifting the axes, it is possible to find similar behavior on the gaze metrics which symbolized by parallel line segments. Some of the combinations of the metrics convey similar information and with the disabling of chosen metric possibility lets the analyst discard these unions [KNB+18]. That is why in our research we also calculate the correlation between the metric results. Strong positive or negative correlation means that they are too similar, they would look like parallel or crossing lines [b&c] in the image.

To represent the usefulness of the system, they applied eye tracking technique for reading the metro maps involving approximately forty participants. However, it is hard for a beginner to comprehend those calculations. Therefore, it is vital to come up with a more straight forward and easier approach to visualize.

Some of the researches used machine learning to distinguish low/high level user tasks. At the usability research of Paliouras et al. [PPKS02] they used the machine learning and data mining algorithms to analyze the user behavior on the Internet. They argued about the usefulness of constructing communities of users with collective behavior, using the power of machine learning methods. The paper presented the results of three case studies for three different types of Internet service: a digital library, an information broker and a Web site [PPKS02]. They used unsupervised learning methods which are the conceptual clustering and cluster mining.

However, machine learning is like a “black box”, researchers throw the data in the algorithm, and the algorithm says if the person’s behavior related or not but it is difficult for us as humans to understand why the algorithm detects these results and how the algorithm reaches this conclusion. That is why the use of machine learning is not chosen as a method in this research.

In our research, there was the experiment on the visualization environment which tries to let the results of the algorithm easily understandable by the people. The two-dimensional way of representation regarding the metrics is simply understandable by anyone, even if the person is not experienced in the research topic. This was a difference between our research and all the other previous works [KNB+16] [PPKS02] [KNB+18]. That is why this experiment took place and examined whether the result allows us to detect the low-level tasks reasonably with the combinations of two metrics. Also, reflect the results in an easily understandable visualization environment.

3 Fundamentals and Eye Movements Metrics

There are some metrics chosen to use in this study to represent a wide variety of eye movements. However, there was a problem related to the metrics on the visualization process. As discussed in the introduction chapter we were restricted to use the two dimensional visualization in our research, and we aimed to show two metrics at the same time in the scatter plot. However, when we looked into some of the metrics in the literature such as fixation durations, there exist many fixation durations in the gaze data. Most of the measurements already had a time component, they were in a time series, so that was an additional dimension. We needed a 4D display or at least a 3D environment for visualizing two combinations of the metrics without processing them. We needed to reduce each metric to one dimension to be able to combine and show two metrics at the same time. For example when we were interested in the fixation durations, then we could combine these values into a single dimension by extracting the values from the metric results such as mean, median or total of it. We only worked with the metrics that can be calculated with knowledge of the stimulus and a recording from a remote eye tracker.

Eye Movement's Metrics

Some of the definitions of the eye movement metrics are as follows:

Scanpath Length: The sum of the length of all saccades is giving the value of scan path length [HNA+11] [KNB+18]. This metric can be helpful when the environment is under control to give the person similar or the same tasks. When a simple searching task considered, it can be distinguished by the scanpath length.

For instance consider the task is to find the numbers in the visual area which are higher than the number twenty. As the first environment, consider there are five same size rectangles which contains numbers inside, and their positions are generated randomly on the visual area. Moreover, as the second environment, if we did not change any of the properties but only the increasing the sizes of the rectangles which contain numbers higher then the number twenty. Then in the first task, they would have to look at all the rectangles to be able to compare the numbers whether they are above the twenty or not. However, in the second environment, the participant simply find the rectangles which are bigger than the other rectangles on the screen which are containing a higher number than the twenty. It can be expected that the scanpath length will be longer in the task which contains the same size rectangles due to visiting every rectangle on the screen. This metric could be the indicator of the person's effort to finding a higher number than the 20 tasks. The lower scanpath length values mean, the lower the effort to find the values.

Saccadic Velocity: Saccadic velocity is the first derivative of the position data concerning the time. It is primarily calculated as part of event calculation [HNA+11]. Saccadic velocity is the velocity of the eye movements from a fixation point to the other fixation points. Saccadic velocity has been used

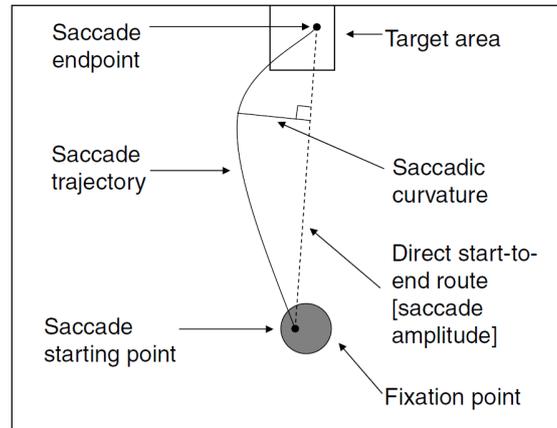


Figure 3.1: Saccadic curvature representation. The figure taken from [NH06]

as a measure of cognitive activation level, or what is often called arousal level [HNA+11] [Gal89]. Saccadic velocity increases as the difficulty of the task increases and likewise decreases with the increasing time over on the task [HNA+11] [MS96] [LCHK91]. With this metric, it can be understood if one of the given tasks requires more effort than the other. For example if there is a complicated math problem on the visual area then the saccadic velocity values decreases. However, if the task is only the looking at an object and naming the color of it, then the saccadic velocity values are expected to increase over the math task due to the ease of the color naming task. In our study, we are interested in two metrics which were extracted from the saccadic velocity: mean and median.

Scanpath Velocity: it is the speed of an overall eye movement. Scanpath velocity is defined as the product of the average saccadic amplitude of saccades [HNA+11]. In some reading tasks, it was shown that the scanpath velocity changes with the font size of the texts. Smaller letters slow down the eye movement speed. Also in the research, the luminance differences in the letters investigated [HNA+11]. We expect similar results from the scanpath velocity metric as the saccadic velocity metric concerning to the difficulty level of the handled task.

Saccadic Curvature: saccadic curvature is the calculation of the overall shape of the saccade. Saccades are thought as ballistic eye movement. Port and Wurtz [PW03] define the saccadic curvature as the angular standard deviation of the tangential direction of the saccade [HNA+11]. The calculation of the saccadic curvature given in the Figure 3.1.

With the saccadic curvature metric, we want to examine if the gaze positions are shifting around too much. It is possible that the person is confused looking at quite different parts of the visual area instead of following a straight path with small jumps. This confusion can occur if all the objects on the visual are had quite a high similarity. For instance, objects with the same color and the same size on the screen can create this confusion if there needs to be an interaction with the objects. However, if some of the hints are added to the objects to increase the person's attention, then we might have some lower curvature results.

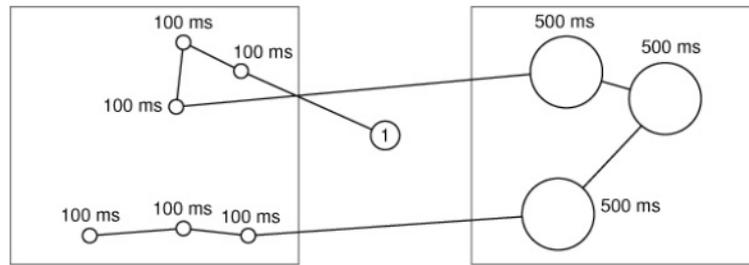


Figure 3.2: Fixation duration representation. The figure taken from [HNA+11]

Fixation Duration: Fixation duration is likely to be the most used measure in eye-tracking. Fixation duration defined as a time when the eye is relatively still. When a participant repeats a task, average fixation durations remain similar across trials; however different people have different average fixation durations [AC99] [JHML12].

Just and Carpenter (1980) [JC80] [HNA+11] formulated the influential strong eye-mind hypothesis, according to which there is no noticeable lag between what is fixated and what is processed. If this hypothesis is correct, and a participant looks at a word or object, he also simultaneously processes it, for exactly as long as the recorded fixation [HNA+11].

Fixation duration is frequently used as an indicator for the cognitive processing intensity. When the average fixation duration increase it often means that a person spent longer time thinking around a visual area. Spending more time on a stimuli means that it has a high complexity or a lack of intuitiveness in it [KNB+18]. Here in Figure 3.2 the several fixation duration is given, as one can see that the fixation durations differs between the visual objects.

The fixation duration time is related to the processing time of the text or object. In our research, we were interested in that if it is possible to distinguish the different tasks by looking at the fixation duration time when we extend the text information with the indicators such as color or size information.

The added color information reflects, and this is an aid when there is a comparison task between the text information. For example, finding the text context which is related to the animal or flower names. If we assign to the animal names as the color of yellow and the flower names as the color of red, then recognition time probably differ than the text without the color information. We are expecting the lower fixation times on the color added text data. If there are mixed texts in the visual area without any indicators, then it needs a reading event to understand the context of the text way more then the color information added data. Reading event might increase fixation durations times. Fixation duration represents the relative engagement with the object. The greater the average fixation duration, the higher the level of engagement [HNA+11]. We used the fixation duration metric as the mean, median and total. [HNA+11].

First Fixation Duration After the Onset of Stimulus: The first duration after the onset of stimulus is the first fixation time after the first stimulus is shown [HNA+11].

The first fixation after onset has a special status, as it corresponds with the very first processing of the audited part of the stimulus, and its duration shows the ongoing information processing [HNA+11].

With this metric, one can expect a longer or shorter duration on the first fixation duration with different gaze tasks. The task of finding the smallest value on the visual area, which is a text comparison, when the participant is searching the context without any additional information, we expect a more extended scan through on this task. Moreover, we expect a longer first fixation duration due to reading the text value and processing it. However, if we enrich the text content with color information, then we expect that the scan through time and the first fixation duration time decrease due to not having any processing time like the text reading.

Saccadic Number: saccadic number counts the number of saccades from the start of the trial until the end of the trail. Different kinds of stimuli evoke varying numbers of saccades. The saccadic number is proportional to the number of fixations and related to the fixation duration [KNB+18]. When the number of saccades is increasing in a fixed amount of time, the fixation duration is decreasing [KNB+18].

If the person is searching the context without any additional information, we expect the number of saccades increase on this task due to jumping all the text and comparing them one by one. Also, if we increase the amount of the different text in the area, then the intensity of the comparison also increases. We expect an increase in the number of the saccades if the comparison intensity increases. However, if we enrich the text context with color mapping then we expect a meaningful decrease on the number of saccades. The person can easily trace the color information on the visuals to find the sequence of the numbers.

Saccadic Rate: the measure of the number of saccades per second is known as saccadic rate. Saccadic rate decreases when task difficulty or mental workload increases [HNA+11] [NTS02]. This interpretation is utilized in several applied studies, for instance, Pan et al. [PHG+04] [HNA+11], who found a difference in saccadic rate between two types of web pages. For a still position stimulus, the saccadic rate should be identical to fixation rate, and this correlates to fixation duration [HNA+11]. In our study environment, we used a stable object structure. Thus, we expect a similar increase and decrease with the metric of saccadic number.

Fixation Numbers: When inside an area of interest, the number of fixations has also been called 'fixation density' [HNA+11] [HH99]. From the onset of the stimulus until the task ends the amount of the fixation number giving the result of this metric [HNA+11].

Search efficiency and difficulty are negatively correlated with the number of fixations. Goldberg and Kotval [GK99] distinguished the number of fixations as a measurement of assessing usability.

Rötting [Röt01] says that a low number of fixations could mean either that the participant is experienced, or that the search task is simpler than the another. If all the participants had the same task for the first time, then the experience difference is not an affecting factor in the number of fixations. The factor of the simplicity between the given tasks can be examined if the persons have the same experience level on the tasks.

A high number of fixations would then be indicative of the difficulty in interpreting the fixated information or layout. This interpretation is supported by Ehmke and Wilson [EW07].

In a finding the smallest value on the text comparison task, if the participant is searching the context without any additional information, we expect the number of fixations increase due to difficulty in interpreting the fixated information. However, if we improve the text context with color or size information, then we expect notable decrease on the number of fixations due to the ease of interpreting the texts. The participant can follow the color information on the visuals.

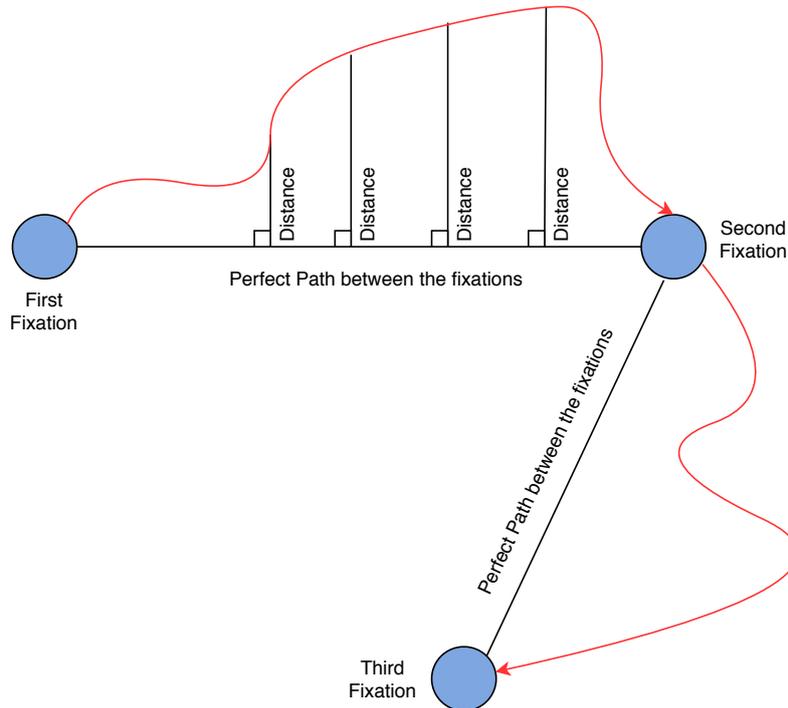


Figure 3.3: Start End Fixation calculation representation. Red lines shows the scanpath, blue circles represents the fixations

Moreover, there is another aspect of the number of fixations. Memory builds up for every new fixation upon an object in a visual area. Memory for the position of that object will improve Tatler et al. [TBG05]. This has the meaning that fewer fixations required to find objects once they have been encoded and a memory representation accumulated through multiple fixations [HNA+11]. For example, in a designed environment if the visuals on the screen have randomly selected positions, then this eliminates the memory built which leads to no increase or decrease in the number of fixations metric.

Fixation Rate: fixation rate is a ratio of the number of fixations and the period such as the duration of a trial in seconds [HNA+11]. There are a few cases where the fixation rate has been used as a measure. Fixation rate was observed to be negatively correlated to task difficulty by Nakayama et al. [NTS02]. Van Orden et al. [VLMJ01] obtained the fixation rate to be correlated with tracking performance over participants. It pointed out that fixation rate is predictive of target complexity in a monitoring task and it could be used as a measure of mental workload.

Start-End Fixation: This metric looks for the distance values within the scanpath and a direct perfect path between the two fixations. The representation of this calculation shown in Figure 3.3. We consider the scanpath started from the beginning of the first fixation until the second fixation and every time stamps the distances between the scanpath and the perfect path calculated. Furthermore, the distance values squared sum and the variance calculated as two distinct metrics.

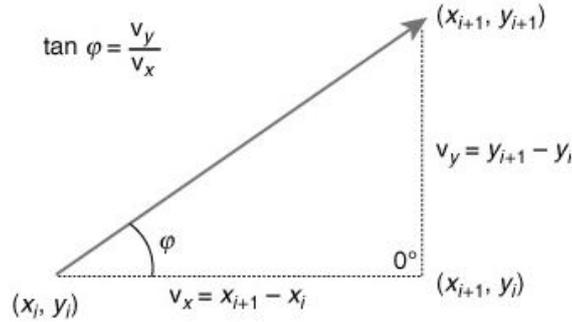


Figure 3.4: Saccadic direction calculation represented. Which illustrates the degree between the two gaze positions which successive one and after. The figure taken from [HNA+11]

Start-End Bubble: This metric is related to the previous start-end-fixation, but at this time we used the positions of the bubbles instead of the fixation positions. Then we consider a perfect path between the bubbles and calculate the distances with the scanpath. We discuss the bubbles later in the study design chapter.

Saccadic Direction: With the results of the saccadic direction, one can analyze the person’s gaze behavior related to the direction of the eye moved. The horizontal angles between the fixations calculated for every exist fixation. The generated values between 0-359 degrees. Two different ways of visualization of the saccadic direction results for the same task are shown in Figure 3.5. As given in the Figure 3.4 the calculation happened between the two fixation points. A saccadic direction between fixations (X_i , Y_i) and (X_{i+1} , Y_{i+1}) can be calculated using basic trigonometry [HNA+11]. Note that many saccades are curved and hence have varying directions along the trajectory [HNA+11]. The saccadic direction only represents an ideal straight line from start to end point [HNA+11].

There were three kinds of calculation concerning the saccadic direction which we were interested which are the mean, median and mode values.

Saccadic Direction Mode: The most repeated value in the dataset of saccadic directions will be the mode value. If the frequency of some of the elements equal to each other then the kernel density approach is used for the decision of the mode selection. Kernel density approach is decides whether the element and the surrounding of the element are denser than the other most frequent value. We couldn’t use the regular value of the mode because there could be more then one mode value appear due to our values are not natural but real numbers. That is why we need to use kernel density estimation approach for calculation of the mode value. We used the Equation (3.1) for the kernel.

$$Kde = e^{\frac{CurrentDegree^2}{(-2*(4.5)^2)}} \tag{3.1}$$

We sample the data domain as 15 degrees while traversing with the density formula. The further information related with this metric given in the Section 5.2.3.

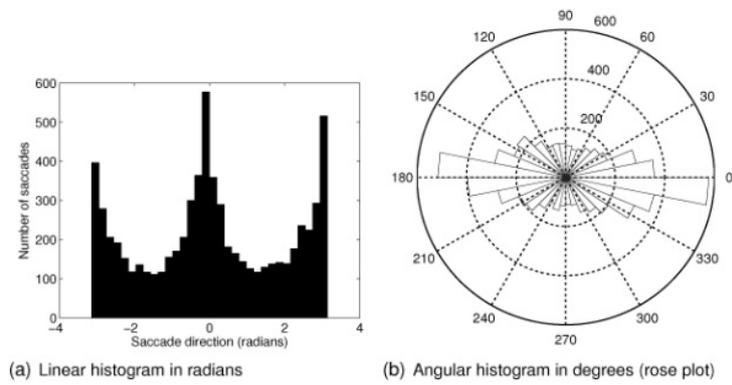


Figure 3.5: Saccadic direction histograms for reading data. First a classical linear histogram, then an angular histogram (rose plot) for the same data. Bin size 11.25 degrees. The figure taken from [HNA+11].

4 Study Design

As was pointed out in the Fundamentals and Eye Movements Metrics chapter, we have discussed some of the metrics which can be useful for us to distinguish differences in human behavior. If we design an environment which the people have to read, then the gaze behavior will be particular for reading. These tasks can also be extended to search somewhere on display for a specific task or comparing the text. In our study, we wanted to create a playground which participants need to read the texts, scan the display with some tasks, make some comparison back and forth in the texts. We could implement that as having different text in the specific areas and force the participant to make a comparison. We make the texts covered with the bubbles, and participants needed to compare them to get an order. Further, we know that people are not restricted with recognizing the black and white colors, as humans we also have the receptor for the different colors. Different colors can create different behaviors of the people. For extending the understanding of the different behaviors of the people we add colors to the numbers instead of just black and white numbers on the screen. Moreover, people are also good at estimating the sizes of the objects. As an addition to the different behaviors, we also introduced the size comparison in the study with assigning different sizes to the objects. With this approach, we have different types of playgrounds which the participants had different types of tasks.

A game designed for the study to cover all the task that we have discussed before. For the purpose of reflecting the low-level eye tasks as explained earlier, three game modes were designed. Moreover, to be able to make the attendee's effort easier, there was only one high-level task which was choosing the smallest number containing bubble and clicking on to it which leads the bubble to disappear.

Clicked bubbles disappeared to make it less crowded the area and to keep the participants focus on the remaining numbers on the screen. Because if the bubbles remained on the screen, then bubbles could confuse people. Moreover, the participants also needed to memorize which bubble they clicked before and compare all the bubbles on the screen with each other. Participants should somehow ignore them, but if it stays on the screen, then it is not possible to ignore. That is why to be able to escape from the confusion, the clicked bubbles were removed from the screen. The name of the 'bubbles' comes from the shape and the pop event on the screen. That was very intuitive because participants had a knowledge of soap bubbles. Also, bubbles have a natural pop up action when somebody touches which reflects a similar behavior as the task. Playing the game was relatively easy for participants, the participant always assigned for the same task, and also the bubble popping was rather fun to play. To take our participants attention in the study and keep them focus until the end, we made the game a bit competitive. They notified as if they click the wrong bubble, there was going to be a time penalty, and because of that reason, they should do it as quickly as possible. Time penalty motivated them when they told as to finish the game quickly but without having errors. When the participants were forced to have to be quick on the task, then they did not have too much time to jump around. Moreover, they needed to follow their first instinct on the tasks with the color or size information.

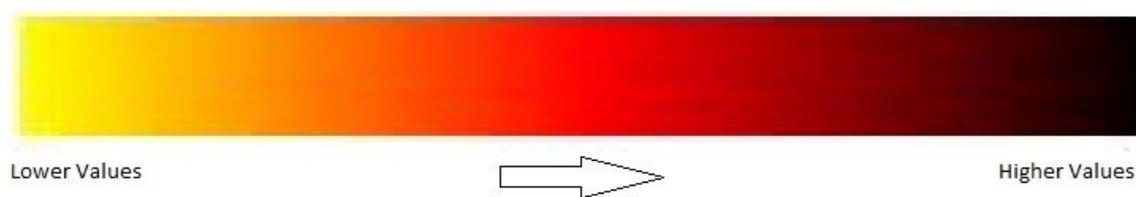


Figure 4.1: The custom color scale

Within this game, there were three different modes which let us assign different tasks for the eyes. The design of the three-game mode was taken place as follows;

The main task performed by the participant was choosing the bubble which is containing the smallest number on the screen at that moment. There were several bubbles, and when the participants continued to click the smallest one, the clicked bubble popped up and disappeared from the screen. In the game mode “numbers”, all bubbles were at the same color and size. The participant needed to read the numbers in order to find the correct sequence and pop all bubbles. In the game mode “size”, the bubbles were the same color again, but their sizes were increased according to their number. In the mode “color”, the bubbles were reverted to the same size again, but their colors were changed according to their numbers.

The amount of dots (bubbles) were decided concerning the duration of the study and also the intensiveness of the game. The amount of dots (bubbles) are just the parameters. Specific values are then adjusted in the Pilot Study section. If the number of bubbles is much more, then the searching time for the smallest number will increase significantly, not just at the beginning of every each game also until the end of the game. Also, the design made the game simple enough to show the differences in size and color changes. If the number of bubble increases then the differences between the colors decreases. This color difference has led the participants not to check the differences in the colors or sizes, just reading the numbers inside enough to decide which one is the smaller one. If the number of bubbles are too less, then participants only read the numbers inside, not affected by the color or size difference. That is why the number of bubbles needed to be arranged.

To be able to reflect the changes in between the bubbles, different colors were assigned to them. Moreover, the selection of this color was decided by looking at many of the color scaling frameworks and their options. The color scales of Inferno, Plasma, Viridis, Magma were tested. The results of these options were not satisfactory. This custom color scale has shown in Figure 4.1.

On the framework chroma.js, the scaling of the 'yellow, red, black' decided in the 'lab' mode. This increases the color differences within the given number of colors by looking at the human perception of the colors. Color generation made it easy for participants to make a decision even at a single glance.

As mentioned before, the game was integrated with the eye tracking device which is recording all the position changes in the eye pupil. Concerning the recording of the eye pupil position, it gives the information about the exact point of the person's focus position on the screen as pixel values. One of the main tasks on the project was to take the recorded raw data from the eye tracking device and process them as meaningful information such as the metrics. There were 22 metrics calculated. See also these metrics in the section Eye Movement's Metrics. Then the research continued with

the analysis part, the 2-dimensional metric combination values were evaluated by the k-means clustering setup. With the clustering results, the question of 'How well the 2-dimensional metric is to distinguish the low-level tasks ?' was studied.

Furthermore, there were two options which can be used for the study. One of them was 'between the subjects' design type. In this design, one could call a participant and ask him to play only one type of a game such as 'color', then ask for the next participant to play another type of a game such as 'size'.

Afterwards the study has a comparison between the modes, but there was a problem. The problem was that there could be a difference in how people play games. As a proper study design, having also differences in the people's gameplay should be avoided to make sure the difference comes from the game mode and not from the people.

Hence, the design of 'within the subjects' type of study allowed the participants to play all three game modes repeatedly and produce all game results for all games.

Last but not least, the numbers which were carried by the bubbles were chosen within a range from the 1-30. The numberings were assigned to the bubbles as randomly within this range. The design of the 'randomly chosen number' was necessary to the elimination of the learning effects on the participants. Otherwise, the participants will always search for the number 'one' as the smallest value. However, in our design, when the color or size information is not provided for the participants, they should always take a look at all the bubbles in order to choose the smallest number. Moreover, the color information and the size information are not depended on this random selection of the numberings. In other words, In 'color' game mode, if the smallest number for one of the participant were selected as number five, and as number three for the other participant, the colors would be the same for both, independent from the numbers. But depending on the ranking order of the numbers such as the smallest one always was yellow. Also for the second smallest and the proceeding bubbles, this harmony continues till the end of the group of bubbles. Moreover, for the game mode 'size' the numbers are not mapped to the sizes directly, same like the 'color'. For example in the different games, all the bubbles which had the same sequence will be the same size not depended on the numbering inside. This dismissal of the mapping was for the elimination of such cases like the two randomly chosen number is too close to each other in the sequence, then there might be bubbles which are highly similar as their size's or their color's.

Specific values are available in the Pilot Study section.

5 Implementation

The implementation of the research environment was separated into two parts. The first part was the Game Implementation part which explains the steps of developing the game. The second part was the Analysis Implementation which explains the steps of developing the Analysis Tool implementation.

5.1 Game Implementation

The first part of the development was the game environment. The game allowed us to investigate the stimulus object changes and at the same time track the eye movements of the participant. Concerning the game development, the main task for the participant was to choosing the bubble which contained the smallest number on the screen at that moment. There were some bubbles on the screen and when the participants searched and continued to click the smallest number, the clicked bubble disappeared from the screen. If the participants clicked the bubble which was not the currently the smallest number, then the bubble was not going to pop up on the screen. All the participants were informed that, there was not a bug in the game, the bubble always popped up when it was clicked if it was the correct one. The only reason the bubble did not disappear from the screen was that the choice of higher value contained bubble then the current smallest bubble in the visual area. After all the bubbles were vanished from the screen when a new game starts. Three game modes built towards adapting the game environment to be able to reflect the stimulus object changes. The same game mode repeated itself as few times to catch multiple data points for the further cluster analysis of the data later on. In the first game mode 'Numbers', all bubbles were the same color and size shown in the Figure 5.1. The participant needed to read the numbers in order to find the correct sequence and pop all the bubbles.

In the game mode Size, the bubbles were the same color again, but their sizes increased according to their number shown in the Figure 5.2.

In the mode Color, the bubbles reverted to the same size again, but their colors changed according to their numbers shown in the Figure 5.3.

The color scale algorithm used the Chroma.js¹ library and with the method selection of 'lab' for scaling the colors. The algorithm can serve as many as the bubbles which needed to be generated.

A sample game steps from starting until the game ends was shown in the Figure 5.4.

¹<https://gka.github.io/chroma.js/>

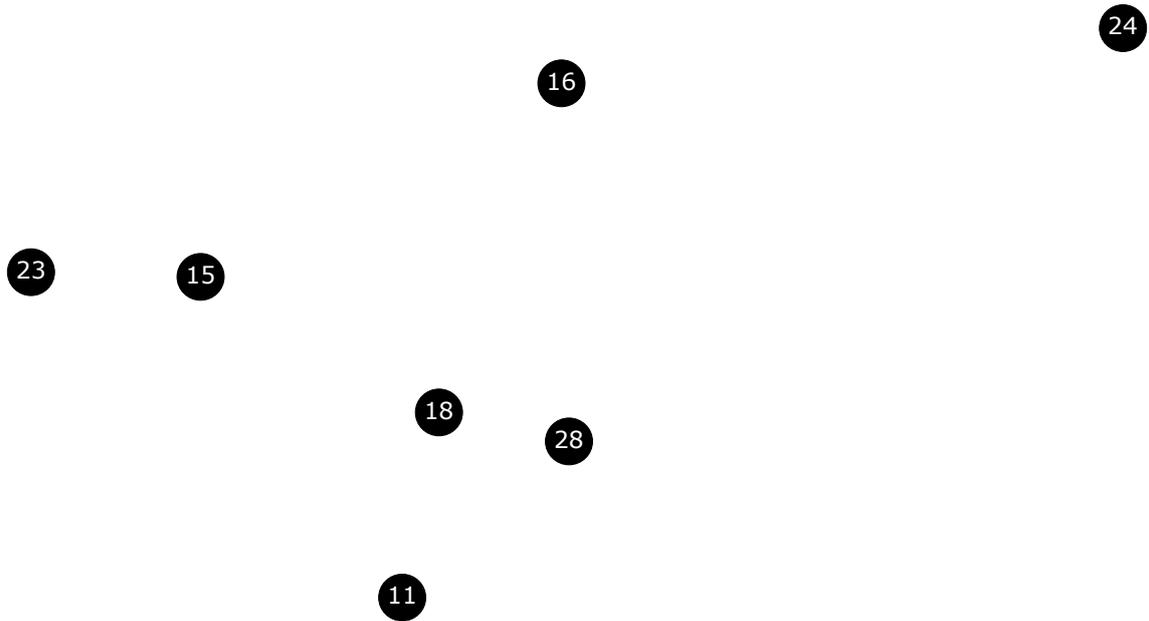


Figure 5.1: Game mode 'Number' representation

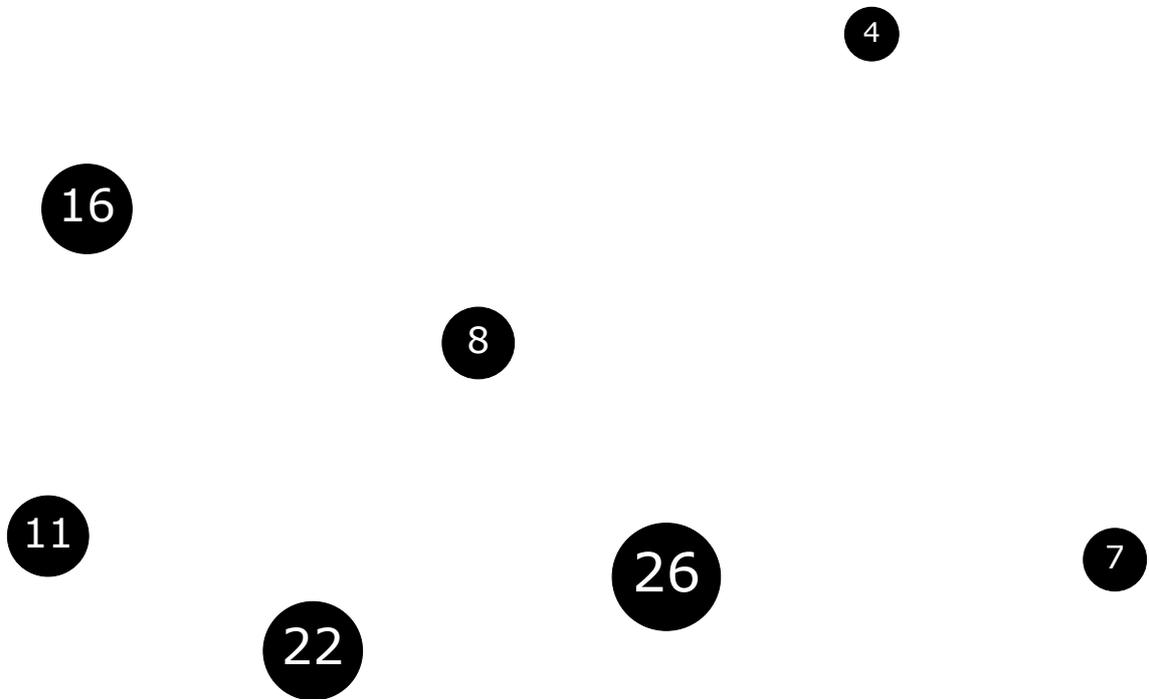


Figure 5.2: Game mode 'Size' representation



Figure 5.3: Game mode 'Color' representation

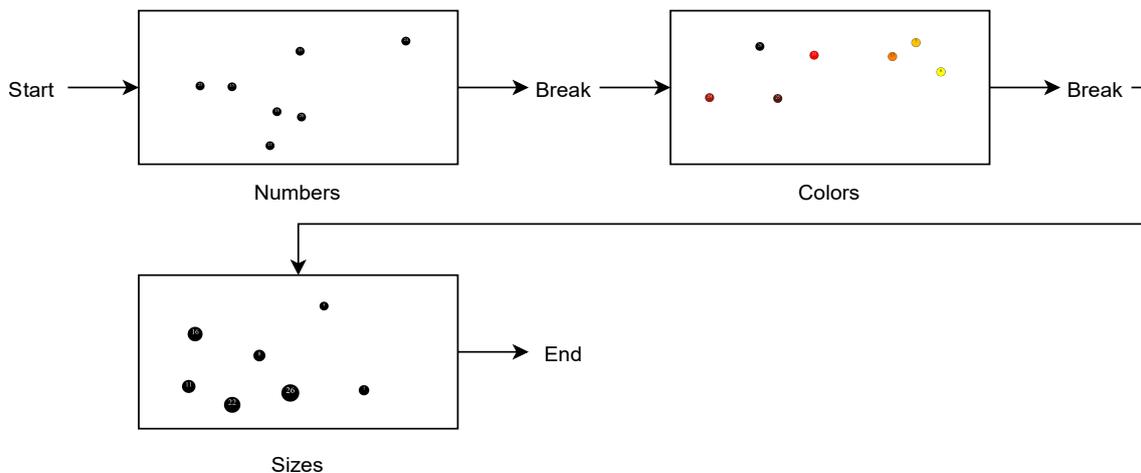


Figure 5.4: Game steps from starting until the game ends

For each game mode, the number information assigned to all bubbles. To be able to increase the participant's effort on the searching of the smallest number at the visual area, the smallest number was not started from the number 'one'. Instead, it started as a randomly chosen value. Additionally, all the values assigned randomly as in the given range parameter and numbers never repeated itself at the same game, participants always saw a unique number set in the screen. While the participant played the current game and all the games came one after another, the participant's eyes could get tired, and maybe the attention could be reduced. For that reason, the 'game breaks' created between the game mode changes. The participants could make game breaks, and the game would continue after the participant regained the focus. Also, every bubble which was generated, has a unique position on the screen which was randomly selected.

The most important challenge in the position assigning was to prevent any collision within the bubbles on the screen. The bubbles should not have collided or touched to each other even if the bubble amounts increased significantly. The problem of random position assignment was complicated when game mode was selected as 'size'. Due to the increased size of the bubbles, a random collision in the remaining area made it difficult to find a free position. This position assigning problem was directly affected by the minimum and maximum bubble sizes. Because, if the claimed smallest bubble diameter was such as 20 pixels and the maximum was 1000 pixels, it was not possible for the bubbles to be placed on the screen without a collision.

5 Implementation

RelativeTimestamp	X	Y	LeftX	LeftY	RightX	RightY	LeftPupil	RightPupil	AbsoluteTimestamp
685909837104	2813.654	419.2262	2870.369	426.5841	2756.939	411.8682	2.781448	2.971222	2019-04-03T12:14:36.0951007Z
685909837939	2813.956	419.8809	2873.225	432.5432	2754.688	407.2185	2.763824	2.971069	2019-04-03T12:14:36.0959357Z
685909838770	2806.492	416.1058	2857.165	430.3507	2755.819	401.8608	2.729935	2.972595	2019-04-03T12:14:36.0967667Z
685909839604	2810.628	418.022	2867.437	428.1589	2753.819	407.8851	2.758865	2.972672	2019-04-03T12:14:36.0976007Z
685909840438	2805.443	425.9318	2859.028	435.7863	2751.858	416.0772	2.770508	2.991089	2019-04-03T12:14:36.0984347Z
685909841270	2815.449	414.6129	2872.7	419.6371	2758.197	409.5888	2.751266	2.968384	2019-04-03T12:14:36.0992667Z
685909842104	2813.371	417.7661	2871.344	430.3978	2755.399	405.1343	2.772995	2.97403	2019-04-03T12:14:36.1001007Z
685909842938	2822.539	410.1747	2890.368	409.2382	2754.71	411.1113	2.724228	2.973389	2019-04-03T12:14:36.1009347Z
685909843770	2810.038	420.2976	2864.37	429.8554	2755.705	410.7398	2.782944	2.980316	2019-04-03T12:14:36.1017667Z

Table 5.1: CSV gaze file representation

The algorithm iterated itself until it could find a valid position series for the bubbles at the beginning of each game. When the environment parameters were selected for the study, the most critical parameters were the minimum bubble size, maximum bubble size and amount of bubbles to be generated. These parameters arrived as the solution of the problem 'If all the bubbles have enough space to be generated or not?'. Moreover, this question will be answered with the pilot study later on in the Pilot Study section.

In all games, participant's actions were recorded such as the randomly generated bubble's positions to be able to integrate the game with the second part of the implementation which was calculating the metric values. Also, the information related to game specific values were stored such as the time of the game's start and end, and also the bubbles' click times. The recordings of all the game values and the recordings of gaze tracking values were needed afterward in the analysis tool implementation.

Last but not least, the game was integrated with the eye tracking device which was recording all the position changes in the eye pupil. Concerning the recording of the eye pupil position, it was given the information as the exact point of the person's focus position on the screen with pixel values.

After the participant successfully finished the game, then the results were saved into the two separate files. One was the CSV file which contained all the eye position information at a frequency of 1200 Hz while the game was played, and all the respected time for each pixel position value which was registered. The CSV file is shown in Table 5.1.

5.1.1 Server Implementation

HTTP-server was used to run the game locally. HTTP-server² is a package that is available for the Node.js³ environment. Also, another server was used to communicate with the eye tracker and with the game application called Gaze server. Gaze server was a custom C# implementation that used the API from the Tobii Spectrum Eye Tracker. These servers were applications that were running on the same local machine. Gaze server triggered the eye tracker to start and also end the game. Further, the Gaze server was used to record the gaze data file and also saved with it. Eye tracker had a specific Windows API, but we were using a browser for our game. With the Windows API the communication could not directly achieve with the windows environment through browser. Also, we could not use the functionalities of the Windows operating system. Thus, there was a need for a

²<https://www.npmjs.com/package/http-server>

³<https://nodejs.org/en/>

server in between, which received the events from the eye tracker, also had access to the file system to log the gaze data. This server was also able to communicate with the website. Moreover, both with HTTP server and the Gaze server also communicated through the website. Then we had two servers which created an inter-server communication in between. We used web application because it could run anywhere, with the current setup we could also use another eye tracker as long as they were providing the same server interface.

5.1.2 Used Frameworks for The Game Development

Papaparse.js

Papaparse.js⁴ used for reading the CSV data from the local file system. This process was needed a special framework like papaparse.js due to the eye tracker's large CSV file sizes when the 1200Hz reading was selected. Papaparse.js was used javascript's 'promise' functionality which worked as a trigger. When all the data was fully uploaded to the web browser's cache, then the Papaparse.js parsed the CVS readings into a format which the javascript could interact. With this format, one can easily could reach the values like a class object such as giving the column names of the CSV file to interact with all the results which were at that CSV file's column.

Hyper Text Markup Language and Snap.svg

All different versions of the browser game were developed with the Hyper Text Markup Language (HTML). HTML is the primary construction segment of the Web. The content structure of the web page can be defined with HTML Also, HTML increase it is power with the collaboration with some of the other partner frameworks. The collaborative partners are CSS for the appearance and JavaScript for the functionality of the web page. HTML can be helpful for the user interaction tools such as buttons, fields or text entries but it has some limitation for the drawing of an element on the visual area. This limitation surpassed by the Snap.svg⁵ which is a JavaScript Scalable Vector Graphics (SVG) library for the web environment. SVG is a tool which can create graphics by using vector imaging in the two-dimensional environment. It has support for interactivity of the elements such as HTML interactivities and animations of the objects. The interaction of the elements such as click events allows us to create a game that the user can interact with. Also, the animations can be useful for different types of studies which are needed to move objects positions.

TypeScript

The functionality of the elements fulfill with the TypeScript⁶. TypeScript is a subset of JavaScript which is typed. Typed means that each type of data (such as integer, string, object) is predefined in the TypeScript. Being typed gives additional power to the TypeScript in which the JavaScript is missing. Certain actions were recognized only with particular data types which minimizes the

⁴<https://www.papaparse.com/>

⁵<http://snapsvg.io/>

⁶<https://www.typescriptlang.org/>

errors on processing the data and also better handling of the generated errors. However, TypeScript needs to be compiled to be able to work on the web browser. After compiled it generates plain JavaScript which is ready to work on the browser. One of the essential abilities of the TypeScript is being object-oriented. This game benefits from the object oriented design due to event handling of objects such as removing events, click events, restarting the game events.

Require.js

The functionalities of the game had to be developed using some of the libraries' collaboration. Before using these libraries, they had to be loaded first. Accordingly, `require.js`⁷ was used in this development for file and module loading. Require.js controls the loading sequence, if the libraries has not been loaded then Require.js let the code wait until all the loading libraries ends.

FileSave.js

Further, one of the other javascript libraries which the game uses is the `FileSave.js`⁸. One can save files on the client side with using `FileSave.js`. The game generated some files and these files saved by the `FileSave.js`. However, before saving the files, it needed to be converted first into a Blob object. Blobs represent data that is not necessarily in a JavaScript-native format. After this conversion, it is possible to save the files into the local file system.

5.2 Analysis Implementation

When we were analyzing every each participant gaze recordings, we were giving all the metric result set which contains ninety repetitions of the games to the clustering algorithm. Ninety game repetitions contained thirty repetitions of game mode 'color', thirty repetitions of game mode 'number', thirty repetitions of game mode 'size'. This ninety game increased to 540 repetitions when we combined all the six participant gaze recordings. The details of the number of participants given in the Pilot Study section. Also, all the game modes increased at the same ratio. After the K-means clustering algorithm gives us the three distinct cluster values. But in the result set, the index information which describes the game mode of the game was missing. After implementing the index information to the clustering algorithm, we were able to analyze the clustering results.

First, we calculated the Jaccard index [Ham+89]. To be able to find the Jaccard index, we calculated the union of each game mode and cluster combinations. Also, we were calculating the intersection of each game mode and cluster combinations. With these results, it was trivial to see which cluster covers how much percentage of each game mode. Then we have 9 cluster-game mode coverage values for every each combination of two metrics.

Then to be able to calculate the overall cluster result for that two combinations of metrics we need to choose three distinct values out of these nine values. Consider these nine values as a matrix which the first row shows the first cluster, the second row shows the second cluster and the third row shows

⁷<https://requirejs.org/>

⁸<https://github.com/eligrey/FileSaver.js/>

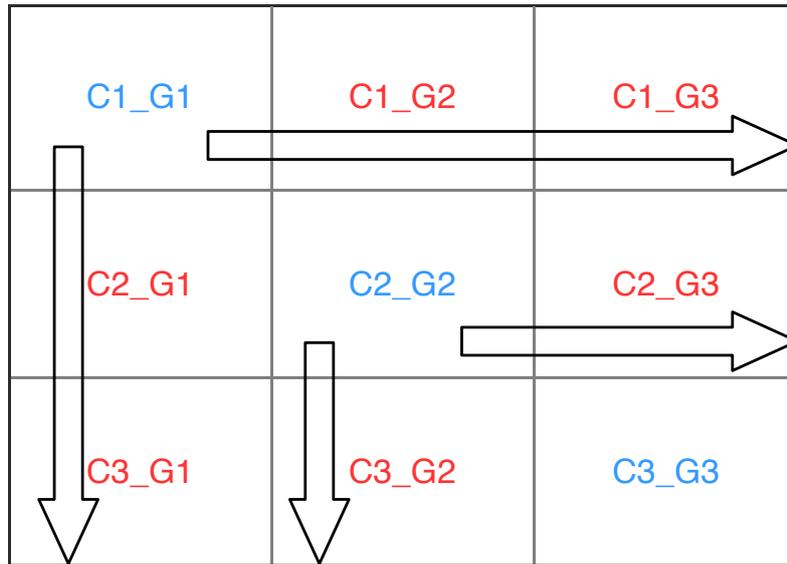


Figure 5.5: Representation of the cluster values. Red values were blocked and the blue values were chosen. Arrows showed the blocked elements after a value was selected.

the third cluster. And the first column shows the clustering result for the first game mode, and the second column shows the clustering result for the second game mode, the third column shows the clustering result for third game mode.

The first step for calculating the best clustering result was to find the best coverage percentage out of these nine values. The representation of these nine value representation is giving in the Figure 5.5. The chosen amount will be stored into the array such as if the first row and the first column was the highest coverage result then it will be saved as the first game mode was covered by the first cluster. Then all the values which belong to the first column and first row was blocked, and as a second best result we were looking at the four left values out of nine. The next highest value from this set such as the second row and the second column gives us the second game mode represented best with the second cluster. And then again all the values from the second row and second column block. Then we were left with a single value such as the third column and third row which gave us the third game mode best representation. This process called 'greedy', no overall optimization, just one best result after the other. Then we left with the three different cluster values for three different game modes. Then we took the mean value of these three value and represented the two metric combinations as this mean value of the three cluster results. K-means Clustering values calculated with the help of the implementation [Eps15] and the algorithm [WH07]. As a further step in the analysis tool, we implemented a function for the Pearson correlation with the help of the implementation [Wes13] and the algorithm [BCHC09]. All the metric result's correlation levels checked and shown in the Table 8.2 which shown later on in the Results section.

Both of the data was stored on the local file system, and for the requirements of the research, all the calculations should be done at one place in the web environment. Thus, it was necessary to select the data recordings from the local file system with a user interface.

Select a file to be loaded as Result

Select a text file: No file chosen

Select Raw File For The Calculation

Select Raw File: No file chosen

Select GameInfo File

Select GameInfo File: No file chosen

Select X-axis:

--Please choose an option--

Select Y-axis:

- Please choose an option--
- Please choose an option--
- Crossings of Saccades
- Saccadic Direction Mean
- Saccadic Direction Median
- Saccadic Direction Mode
- Scanpath Length
- Scanpath Duration
- Saccadic Velocity Mean
- Saccadic Velocity Median
- Scanpath Velocity
- Saccadic Curvature
- Fixation Duration Total
- Fixation Duration Mean
- Fixation Duration Median
- First Duration After Onset of Stimulus
- Saccadic Number
- Saccadic Rate
- Fixations Number
- Fixations Rate
- Start-End Fixation Variance

Figure 5.6: User interface representation

5.2.1 User Interface Implementation

The user interface implemented with the HTML, and for the functionalities of the buttons and the metric calculations, Typescript was used. In the user interface is shown in Figure 5.6, there was two options which the user could select.

First of all, the user had an opportunity to select the raw recordings from the file system and start the calculation process. The algorithm determined the calculations related to the tasks and when it finished, it gave three different information. One of the generated information was giving all the results in a JSON file for reusing the calculation results without needing to calculate it again and again. The following provided information was the best cluster combination metrics results as a CSV file. The third information that it provided was visualizing the three K-mean cluster information with the two dimensions scatter plot shown in Figure 5.7.

5.2.2 Visualization Implementation

All of the game mode information was represented as three different types of shapes and K-mean cluster information was represented with three different colorings of these three game mode shapes shown in Figure 5.7. With the help of two HTML selectors shown in the Figure 5.6, these 22 metric combinations can be arranged and then visualize the result on the screen as the analyst's

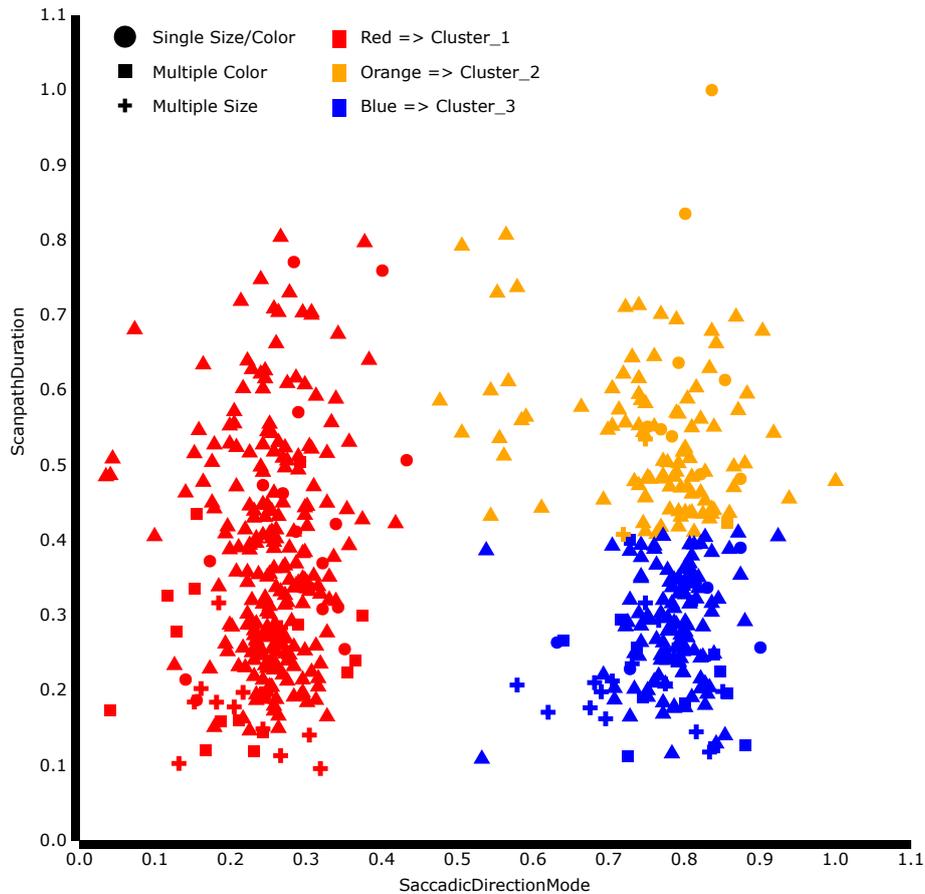


Figure 5.7: Clustering Result

needs. For the visualization of the calculated result of the metrics the D3.js⁹ library used. D3 helps us to visualize data using HTML, SVG, and CSS. Other plotting libraries also tested such as C3.js¹⁰, Plotly.js¹¹ but both of these libraries have limitations. When the clustering and the game modes needed to be shown at the same time on the scatter plot than for a single data point of data it needed to arrange both shape and color information. Also the functionality of showing the data point when the mouse cursor on top of the item was hard to implement for some of the other libraries. As another challenge for the other libraries was to remove the graph and reloaded it without had to reload the page while the selected metric combinations changed.

5.2.3 Metrics Implementation

Tobii Eye tracker was provided the results as raw gaze data, which was contained eye positions as pixel points X and Y coordinates. Also, eye tracker provided the recording of the timestamps as microseconds with the frequency of 1200Hz. Before starting any of the implementations of

⁹<https://d3js.org/>

¹⁰<https://c3js.org/>

¹¹<https://plot.ly/javascript/>

```

1  [
2  {
3      "id": 0,
4      "ValidClickTimes": [
5          "2019-04-03T12:14:38.332Z",
6          "2019-04-03T12:14:39.852Z",
7          "2019-04-03T12:14:41.332Z",
8          "2019-04-03T12:14:42.612Z",
9          "2019-04-03T12:14:43.340Z",
10         "2019-04-03T12:14:44.068Z",
11         "2019-04-03T12:14:44.636Z"
12     ],
13     "ValidClickCircles": [
14         "NONValidClickTimes": [],
15         "NONValidClickCircles": [],
16         "ValidClickedCirclesSVG": [
17             "ValidsnapSVG": [
18                 "NONValidClickedCirclesSVG": [],
19                 "NONValidsnapSVG": [],
20                 "ValidClickedTextSVG": [
21                     "NONValidClickedTextSVG": [],
22                     "AllGameSequences": [
23                         "AllBublePositions": [
24                             [
25                                 1106,
26                                 714
27                             ],
28                             [
29                                 521,
30                                 661
31                             ],
32                             [
33                                 901,
34                                 533
35                             ],
36                             [
37                                 850,
38                                 760
39                             ],
40                             [
41                                 686,
42                                 582
43                             ],
44                             [
45                                 1442,
46                                 638
47                             ],
48                             [
49                                 1659,
50                                 221
51                             ]
52                         ]
53                     ],
54                     "StartTime": "2019-04-03T12:14:36.045Z",
55                     "EndTime": "2019-04-03T12:14:44.636Z"
56                 ],
57             ],
58         ],
59     ],
60 }
61 ]

```

Figure 5.8: Game Json file representation

the metric calculations, the raw eye tracking data needed to be separated and prepared for the metric calculations. The raw data was separated concerning all the start and end played game times. However, we had a problem with the resolution of the time, JavaScript could only go on the milliseconds but while the eye tracker needed 1200hz which was something less than milliseconds. This problem was solved by putting a small margin between the start and end times of the games. The start and end times were taken from the JSON game file which shown in the Figure 5.8.

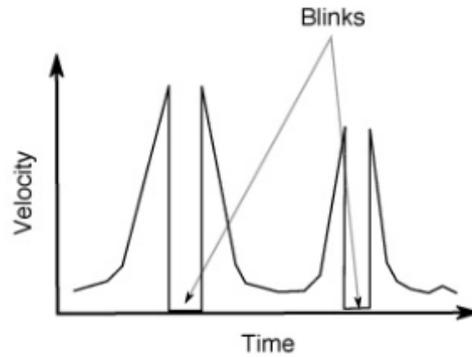


Figure 5.9: Two blinks in a velocity plot. The figure taken from [HNA+11].

All the information from the gaze recordings divided into the containers which reflecting each game with the help of the start and end game times. Hence the algorithm let us give a game-breaking ability while the recordings of the participant's eye movements continued. We did not use the unnecessary gaze tracing values in the game breaks. These parts of the gaze tracking data were taken out. After the division and separation of the eye tracking data, it was necessary to understand the natural failures of the eye tracking data. One of the failures happens when the user blinks their eyes [Ols12]. When the blinks accrue, the gaze information changes accordingly shown in Figure 5.9.

The received data of the gaze positions stopped until the eyes opened back and then the accurate gaze results continued to be generated. This 'no value' readings and 'high-velocity' readings handled in the further steps in the calculations.

Gap filling

Gap filling: as we mentioned when the participant blinks or somehow looks away from the screen, the data loss was almost inevitable [Ols12]. Due to this problem, one has to fill the gaps in the data. Otherwise, the distance values were unusual in the calculation of the eye tracking velocity. Moreover, the algorithm came across with high-velocity spikes in the calculations [Ols12]. For avoiding the high-velocity spikes issue, a certain threshold was calculated. If the loss of the data was in the threshold, then the last valid recording of the gaze position and the first new valid record after the failure was calculated [Ols12]. Then the gap between these two values was filled with the equal distribution of the differences between the last valid result and first new valid result. This functionality was hardcoded by 1200hz if there is any recording frequency change in the future then this gap filling algorithm needs to be converted into the new recording frequency. When the gap was higher then the defined threshold then the fill gap algorithm would not occur [Ols12].

The first and most fundamental calculation is to calculate the point to point velocity. With the help of the high-frequency recordings from the Tobii eye tracking device, the fixations, and saccades found with respect to the velocity of the eye movements [Ols12]. The point to point velocity calculated as Equation (5.1).

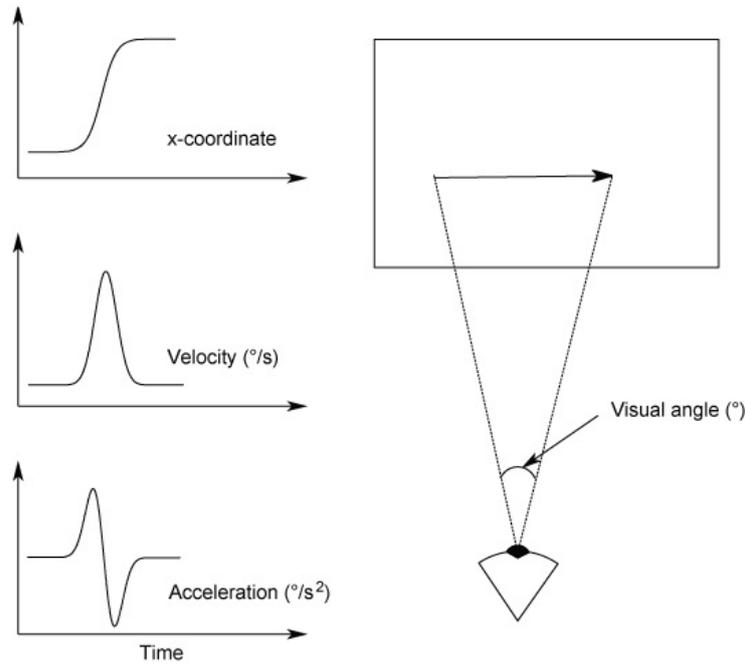


Figure 5.10: Idealized gaze position, velocity, and acceleration profile over time showing one saccade in-between the end of a fixation (to the left of the saccade). The figure taken from [HNA+11].

$$Velocity = \frac{\sqrt{(X1 - X2)^2 + (Y1 - Y2)^2}}{time} \quad (5.1)$$

Even the point to point velocity was enough for calculation of the fixations and the saccades, the research standard was using the angular velocity, and our study also followed the same path with this standard. By using the angular velocity, the effect of the participant's distance to the eye tracker eliminated to influence the results. The angular velocity was calculated by the screen pixel density and the participant's distances through the screen's upper side and the lower side. The distances are approximately the same for all participants, regarding to the Tobii device's calibration process at the beginning of all the recordings. In our study, we tried to avoid the use of the head mounts to control the distance of the participant's head to the eye tracker. With these values, the angular velocity was calculated with the Law of Cosines (Three sides known triangle's degrees) as Equation (5.2).

$$\cos(A) = \frac{b^2 + c^2 - a^2}{2bc} \quad (5.2)$$

Then the calculated angle was used to divide the amount of the pixels of the screen and the pixel to angle conversion calculated by using this translation shown in Figure 5.10.

After the implementation of the velocity algorithm, the noise cancelation algorithm was implemented for using this algorithm on the previously calculated velocity results. The smoothing performed as follows. Smoothing algorithm started from the second value of the velocity calculation array,

with the help from the three values which are the preceding, proceeding and the current values taken into account. After the mean of all three values were calculated, the current element set to this determined value. This process started from the second reading of the data set until the second last value in the end. After these preparations, the angular velocity values were ready for usage in the calculations of the following metrics. In the array of calculated velocities, the algorithm looped through and found the values higher than the values above the threshold, and also for the high speed spikes it uses an above threshold for not considering these values as saccades instead recognizing as blinks. It was defined as above the 100degree/second counted as saccade movements [Ols12]. All the values out of the upper and lower threshold eliminated by giving the '-1' value for these unnecessary values. Then all the values were duplicated into another array due to not damaging the initially calculated velocities. Then also the fixations were calculated under the same lower threshold. Fixations from 0 to 100, saccades above 100 to 500, blink/error more than 500 degree/second. If the values under the lower threshold, these values were counted as fixations [Ols12]. After the saccade's timestamps collected we could start to calculate our eye tracking metrics as fallows;

Saccadic Velocity Mean: The algorithm was looped through the saccade array and it added all the valid velocities together. Then it was divided by the number the elements. Next, all the saccade velocities were sorted starting from the smallest one to the highest one.

Saccadic Velocity Median: Later the middle point was taken as the median of all the saccade values.

Number of Saccades: The metric which was the number of saccades calculated as following. The array which contains saccade velocities was looped trough then all the continues saccades velocities' (which is not -1) beginning and ending times were collected and putted into a unique array. Then this new array's half size was used to calculate the number of saccades in that particular game.

Saccadic Rates: The metric which was the rate of the saccades calculated as following. The total time was calculated as the time differences between the beginning of the game and the ending of the game. Then the current game's number of saccades was divided by the total time of the same game.

Crossing of the Saccades: The metric which was the crossing of the saccades calculated as follows. Using the array of starting and ending times of the saccades, every saccade were compared with only the remaining saccade with higher index compared such as: intersect 0 with 1-n, 1 with 2-n, 2 with 3-n, and so on with using the math.intersect() function from the math.js library. The function of math.intersect took four values which were the current saccades start point of X, start point of Y, and the endpoint of X and end point of Y values. The algorithm calculated the intersection point of two lines at the two dimensions. If there was an intersection between these line equations, the function returned the exact point of the intersection. However, it did not mean that those given points intersect at the interval of our recorded X and Y values. The investigation of the exact intersects was fallowed as if the intersection point was between the both of the X and Y coverage areas of the 4 points then the exact intersection happened.

Number of Fixations: The metric which was the number of fixations calculated as similar to the number of saccade calculation. Under the velocity threshold, all the fixations start and end timestamps were calculated. Metric was derived with taking the half-length of the array which contains the start and end times.

Fixation Rate: The metric which was the rate of fixations calculated as similar to the rate of saccade calculation. The time differences between the beginning of the game and the ending of the game gave us the total time. Then the current game's number of fixations were divided by the total time of the current game.

Scanpath Length: The metric which was the scanpath length value calculated with calculating the distances between the following fixation positions. For calculating the exact point of the fixation, we looked into that current fixation movement than we took all the fixation points with related the movement and after an average of all the X and Y positions was calculated. This average gave us a single X and Y position for that fixation movement. Later, all the distances between the fixation points were calculated.

Scanpath Duration: The metric which was the duration of the scanpath calculated by the start and end times of the current game. The times between all the fixations were giving us the duration of the scanpath.

Scanpath Velocity: The metric which was the velocity of the scanpath calculated as taking the scanpath length and divided into the duration of the scanpath.

Saccadic Direction: The metric which was the mean of the saccadic direction calculated as taking all the horizontal angle values of the saccade's till the end of the game and taking the mean value of these values.

Saccadic Direction Median: The metric which was the median of the saccadic direction calculated as similarly for the previous mean of the saccadic direction. The only difference was the middle point of the values taken for this metric.

Saccadic Direction Mode: The metric which was the mode of the saccadic direction calculated as finding the most repeated value in the saccadic direction dataset for the particular game. If the frequency of some of the elements equals to each other, then the kernel density approach was used for the decision of the mode selection. Kernel density approach was deciding whether the selected element and the surrounding of that element are denser than the other most frequent values. For the saccadic direction mode, we had 360 degrees such as 0 to 359 degree. Also, using this, the 15 degrees of kernel sizes were chosen for the density calculation. All the degree values were sorted from the smallest to highest. Later with the help of the predefined Gaussian kernel function, the 15-degree kernels were processed. Then the one with the maximum density kernel was chosen.

Saccadic Curvature: The metric which was the saccadic curvature calculated as looking at the saccadic movement in the eye. From starting a fixation to another fixation point, the eye movement was creating horizontal angles. All the angles which were created by the eye movement were collected and added together.

First Fixation Duration After the Onset of a Stimulus: The metric which was the 'first fixation duration after the onset of a stimulus' calculated by searching the first fixation's starting timestamp. After finding the first fixation, the duration of that fixation was calculated as a result.

Start End Fixation Squared Sum: The metric which was the start end fixation squared sum calculated with the distance values of the gaze position concerning the current the scan path. All the distances in the per timestamps derived and the squared sum of the values was saved. Then all the values were added and set as the result of this metric.

Start End Fixation Squared Variance: The metric which was the start end fixation variance calculated similarly as the previous metric the start end fixation squared sum. Only the differences were instead of taking the squared sum, the variance of it the values are were taken.

Start End Bubble Squared Sum: The metric which was the start end bubble squared sum calculated with respect to the given game's bubble positions. Between the bubbles, we imagined there was a perfect scan path which starts from the smallest bubbles and goes one by one till the highest bubble. This path was taken as the base of the calculation. All the gaze point distances through this perfect path were calculated, and the squared sum of them were made. After, all the values were added up together to got a single result value.

Start End Bubble Squared Variance: The metric which was the start end bubble variance calculated similarly to the previous start end bubble squared sum. The only difference was instead of taking the squared sum, the variance of it the values were saved as a result.

Fixation Duration Total: The metric which was fixation duration total calculated as looking at the start and end times of the fixations. All the respected fixation duration's were calculated and then the sum of all of them was set as the single value result.

Fixation Duration Mean: The metric which was fixation duration mean calculated similarly as the fixation duration total. The only difference was instead of taking the total, the mean of the values was calculated.

Fixation Duration Median: The metric which was fixation median calculated similarly as the fixation duration total. The only difference was instead of taking the total, the median of the values was calculated.

6 Pilot Study

It was necessary to had a pilot study before the actual study took place. The pilot study was held in the actual study environment to be able to analyze the intensity level of the game and as well as selecting the correct parameters for generating the actual study game.

6.1 Parameter Selection

When the parameters were changed at the game settings, the game reflected those changes by generating the variations of the games. Some examples would be how many games was desired, how many bubbles decided, which colors to use with the bubbles, what was the minimum bubble size, what was the maximum bubble size, where should the bubbles appear on the screen as randomly, which numbering range was chosen to be appearing on the bubbles. These parameters were decided after some repetitions of the pilot studies. Also, the necessity of creating the game breaks between the game changes decided after the game repetition amount was chosen. It was difficult for a participant to repeat ninety games right after each other without having a break. The attention level of the participant was decreasing, and the exhaustion of the participant's eye was progressing while the number of games increases.

6.1.1 Choice of Color Scales

Several color scales tested in the pilot studies such as Viridis, Inferno, Magma, Plasma shown in Figure 6.1.



Figure 6.1: 1) Inferno, 2) Magma, 3) Plasma and 4) Viridis respectively. The figure taken from [Jou15].

In the end, a custom design was chosen as creating the color scale which is consist of red, yellow and black. The lower values start from the yellow, and it goes through red and then the black which is shown in Figure 4.1.

6.1.2 Choice of Bubble's Size and Bubble's Amount

Several minimum and maximum bubbles sizes were tested. If the bubbles are too small then reading the contained number was significantly hard and the participant spent a lot of time on fixating and trying to read the number. Besides, if the sizes are too big, then one cannot put too many bubbles on the screen. In the end, the ratio between the minimum and maximum of 1:2 was chosen, after the min and max bubble sizes were decided, the number of bubbles examined. First, the number of twenty bubbles were tested. However, in the game mode 'size', the differences between bubble sizes were too small, and this made it hard for the participants to understand the number sequence just by looking at a single glance to the bubbles. Also, there were too many colors which are similar together in the game mode 'color'. Then the number of seven bubbles were tested and after seeing the result amount was set to seven.

6.1.3 Choice of Game Repetition Amount

After setting the amount of the bubbles the number of repetitions were studied. The number of repetitions was essential for the generated result set which we used in the analysis. If the repetitions are too less then it was hard to get a valuable clustering result in the end. Therefore thirty repetition of each game were chosen. This limit was not also pushing the participant's eye too extensively.

6.1.4 Choice of the Number of Participants

The chosen number of participants was the result of the number of repetitions. If each participant only played one game, we would have needed much more participants. The number of participants chosen as six. Per participant, we had 90 games and for six participants we gathered 540 data result.

7 Conducted Study

7.1 Participant's background

All six participants are selected from the university students. The participant's age scale was between 24 and 31. The experiment time was arranged by eye tracker's availability. Students were informed about the study before they arrived, for example, they are told that they need to play a game, and the eye tracker tracked their eye movements.

7.2 Before the Experiment Started

First, the participant was welcomed at the University of Stuttgart VISUS building. Then they were directed to the eye tracking room which was specialized for eye studies. All the room lighting could be arranged to be able to get the best results from the eye tracker. Since the eye tracker uses infrared lights to track the eye and if there is natural light coming through the windows or from the door the eye tracking could be affected by them. For each participant, the same procedure was applied.

7.3 Beginning of the Study

The participant was directed to in front of the eye tracker and the computer screen which the game was taking place. Later on, participant and the examiner sit, and the examiner started to explain the consent form. The consent form was started with the information related to the game which was needed to play by the participant shown in the Chapter 10. Then participants were told that the game has three different game modes. The necessary information related to the collection of the data and the analysis of the data was explained as followed:

The participants were told that they see bubbles on the screen with numbers inside. When the participant click the bubbles, bubbles will going to pop and disappear. The participant was told they should pop the bubbles in numerical order. It was explained that they need to start with the smallest and finishing with the most significant number. The participant was informed that if they click the bubbles in the wrong order, they will not going to pop. The participant was informed the numbers will not always start with the number '1'; they can also start with any value. The participant was informed that there are three game modes. They have to play multiple games in a single mode before advancing to the next. It was explained that in the game mode "numbers", all bubbles were the same color and size. It was explained that they need to read the numbers in order to find the correct sequence and pop all bubbles. The participant was informed that in the game mode "size", the bubbles are the same color again, but their sizes are not increase according to their number. In

the mode “color”, the bubbles were changing back to the same size again, but their colors are varies according to their numbers. Further, the color scale picture was shown to the participant which represented the colors presented in the bubbles at game mode 'color'.

Also, all three game mode examples were given on paper drawings to the participants, and each game was explained with those pictures regarding how to play them. Then processing of the measured data which was pointed to the data anonymity shown. In the section of data privacy, the participants were told that the data never leave the VISUS emphasized. After, in the section of risks and rights, the participant was informed that there were not any health threats from the experiment described. The participant was asked by the examiner to read all the form and participants sign the form before the original study starts. After participants signed the consent form, they paid by the examiner's supervisor, and he was asked to sign the form of 'money receiving'. If the participant is under the employment of the university, the participant was not paid.

After the participant signed this form, he received another document which was a questioner shown in the Chapter 10. The participant was asked by the examiner to fill the necessary information related to the current level of math skills. Also in the questionnaire, he was asked about how many hours he spends on computer games. Moreover, the participant's education levels were claimed. After the participant finished the questionnaire, the eye test called 'Snellen Chart' [Sne68] took place. In the range of 3 meters, all the eye test table was asked to read by the participants. When the first error happened by the participant, examiner took notes about that. For the eyeglass users also asked to read the eye test table without the eyeglasses. This step was taken for, if later the eye tracker could not calibrate with the eyeglasses, then the participant asked by to take out their glasses before calibrating the device. After the eye test was finished, the participant was asked to name the given colors in the form which consist of yellow red and black colors. This step took place to understand whether the participant could recognize the differences between the colors or not. There were six game series for each participant which defines how the game modes start and proceed. Participants were assigned random IDs. These IDs encoded the order of the game modes. First, the participant interacted with the game mode 'color' then the 'size' then 'number'. Then the participant was shown by six closed papers which contained numbers inside. The participants were asked to select one of them which was deciding the game mode. As a next step, the participant was asked by the examiner to sit comfortably to be able to keep up 15 to 20 minutes of play time. Then the eye tracker calibration process started. The tracker was arranged by the optimal distance which the calibration software device needed to work correctly. There were not any head mounts which the participant needed to wear. As a next step in the calibration process, the participant was asked to look at the circles on the screen until they pop up. However, these circles belong to the calibration process not related to the game which was going to start. Most of the time the calibration process led participants to a confusion. Many of the participants thought the game already started. After the calibration process was ended, the game start screen placed to the eye trackers screen by the examiner. The participant was told that when he clicks to the start button, the game is going to start recording the eye tracking data and also the game information. The color scale was shown to the participant for the last time to let the participant's memory to refresh. The participant asked by the examiner to play the game as fast as possible and try to avoid to wrong selection of the bubbles. The participant was informed if he makes any mistake, a one second of penalty was going to be added to his score which it was not going to take place truly. This step was only to motivate the participants to play as fast as they could without having any mistakes. Then the participant was also informed that there was going to be some game brakes between the game modes which the participant can rest their eyes if it was necessary. When the continue button was clicked, then the

game was resumed. Also one last time the participants were informed as if they felt any health issue they could leave without finishing the study. Lights around the eye tracker were turned off to be able to create a better eye tracing reading data. The participant played all the assigned 90 games. Moreover, the games were continued nearly 20 minutes up to the participant's skills.

7.4 After the Study

After the game, the participant received questions about general thoughts regarding the game. Moreover, the participant asked which was the hardest and the easiest game modes in the played game. The process took one hour and a half hours nearly, dependent on the eye tracker calibration process and the participant's skills.

8 Results

8.1 Participant's comments related to the study

We ask several questions shown in the Chapter 10 to the participants at the end of the study. First, we asked about general comments related to the game which the participants played. We also asked the participants which game mode was the hardest to play and also the easiest to play.

One participant commented as; at the game mode 'number' the participant had to look every each bubble to be able to distinguish which number was the smallest one at the current state of the game. The participant told this comparison process was hard to accomplish. At the game mode 'color' the participant already knew the smallest bubble colored with the yellow, and he directly searches the yellow bubble. He told us it was even unnecessary to read the numbers inside the bubbles in this game mode. At the game mode 'size' the participant was able to find the correct sequence by looking at the sizes only, but in the end, the bubble sizes getting smaller, and it was getting harder to compare by the sizes of the bubbles. At the end of the games, he started to read the numbers which the bubbles were contained. As the easiest game, the participant's choice was the game mode 'color' and the hardest was the game mode 'number'.

One participant commented as; at the game mode 'number' the participant said the numbers generated between the 1 and 30, but due to this small scale, it was hard to compare the numbers. He commented that if we increase the scale such as 1-99 it would make decrease the comparison effort between the numbers. He had no other comments on the game modes. As the easiest game, his choice was the game mode 'size' and the hardest was the game mode 'number'.

One participant commented as; at the game mode 'color' the participant said the game mode was easy to perform. Without focusing on the bubbles numbers, he was able to see the overview of the sequence of the correct order of the bubbles. Most of the time he was directly selecting the right bubble without reading the numbers inside.

As the easiest game, the participant's choice was the game mode 'color' and the hardest was the game mode 'number'.

One participant commented as; at all game modes, the size of the bubbles was too small. The participant said it would be better if he were playing the game with bigger bubbles. He also told that he was clicking on the bubbles the mouse pointer can be changed to represent the clicks in a better manner.

As the easiest game, the participant's choice was the game mode 'color' and the hardest was the game mode 'number'.

One participant commented as; At the game mode 'number' the participant had to look each bubble to be able to distinguish which number was the smallest one at the current state of the game. At the game mode 'color' the participant compared only the bubbles which close as the same color. At the

game mode 'size' the participant compared only the bubbles which close as the same size but the participant was having a hard time only using the size information he also read the numbers inside to give the last decision on which bubble to click.

As the easiest game, the participant's choice was the game mode 'color' and the hardest was the game mode 'number'.

One participant commented as; At the game mode 'number' the participant had to look each bubble, and it was effort consuming process for him. At the game mode 'color' the participant also had a hard time to decide with the colors. The participant needed to read every bubbles' number to be able to decide the correct bubble. He has only used colors with choosing the smallest one which was the yellow bubble in the beginning. At the game mode 'size' the participant said the sizes reflect the sequences better than any other game modes.

As the easiest game, the participant's choice was the game mode 'size' and the hardest was the game mode 'number'.

With respect to the given comments, we understood that we were able to create a different behavior on different tasks which something that we want to achieve. Some of the participants had better performance at the game mode 'color' and others had at the game mode 'size'. But as a common conclusion, the game mode 'number' was the hardest to decide the correct sequence.

8.2 Clustering and Correlation Results Analysis

For the analysis of the data, all the gaze recordings and the game information was collected from the participants. The analysis tool was fed with the data which was gathered from the participants. As a result, each participant's clustering values were generated with respect to the game modes. This 'best result clustering process' also discussed in the Chapter 5. The clustering tables describe how well was the game mode represent by each combination of the metrics. Then all the clustering result's averages taken which is shown in Table 8.1. As the second step, all the gaze recordings combined, also each game information merged. The analysis tool was fed with this data which generated a single best clustering table. Additionally, each metric's correlation calculated. With the correlation results, we were able to see if there was a negative or a positive correlation between the metric results.

8.2.1 Correlations

As a result, we saw that some of the metric results were highly correlated such as Start End Fixation Squared Sum and Start End Fixation Squared Variance with a 0.95/1.00. This kind of high correlation was expected if the calculation of the metric was highly similar. However, we also saw that even the calculation of the metrics was very different, some of the metrics were highly correlated such as Saccadic Curvature and Saccadic Number with 0.99/1.00. Moreover, the Saccadic Number was highly correlated to the Fixations Number 0.94/1.00. Then when we investigate Saccadic Curvature was highly correlated to the Fixations Number 0.89/1.00. Use of only one of the three metrics was expected to give us similar results due to high correlation. Furthermore, we were able to saw this result in the best clustering table. Such as fixation number

and fixation rate combination gave us 0.08/1.00 which is the lowest result for the fixation number with another metric combination. However, if the calculation of the metrics were not similar such as fixation number and Saccadic Curvature combination gave us 0.11/1.00 which is the highest result for the fixation number with another metric combination.

8.2.2 Average Clustering Results and Single Merged Gaze Data Differences

When we calculate each participant's best clustering values separately, we get very promising results as the average of all the participant's results. However, when we look at the combined gaze recording results, it was much worse than the single calculation results averages. When we investigated this change in the results, we saw that in the case of single participant calculation's cluster results, they were all well separated with respect to the game modes. The well separation shown in the left side of the Figure 8.1. However, when we examined the case of combine gaze record calculation, we saw the clusters meshed with all type of game modes which is shown in the right side of the Figure 8.1. By just looking at the graphs it was hard to detect the cause of this disorganized cluster values. We can only conjecture the cause for further research on the topic.

One of the reasons could be the lack of head mount used on the experiment with the participants. The experiment took a long time and after some time participants were moving a lot unconsciously. This could affect the recording results and create differences between the participants.

As a second reason, the participants give us comments related to the gameplay and nearly half of them choose the game mode 'color' as the easiest one to find the correct sequence on the bubble popping task, and for the other half the easiest one was the game mode 'size'. This could also affect the combined result.

8.2.3 Valuation

We achieved promising outcomes to distinguish the human behavior on different tasks in this study even if the K-mean clustering was a simpler method than machine learning techniques. However, in the current state, it was not better than machine learning techniques on task inference, but we still achieved even 0.58/1.00 results which are highly successful. Moreover, to be able to represent two metrics at the same time in the two-dimensional metrics there was a need for reducing the metrics dimension. In this study, we reduced the dimensions such as taking the mean value of the fixation durations. This dimension reducing also decreases the precision of the metrics which also decreases our detection on distinguishing the human behavior on different tasks. Even with this highly compressed metrics, 0.58/1.00 was a success.

8 Results

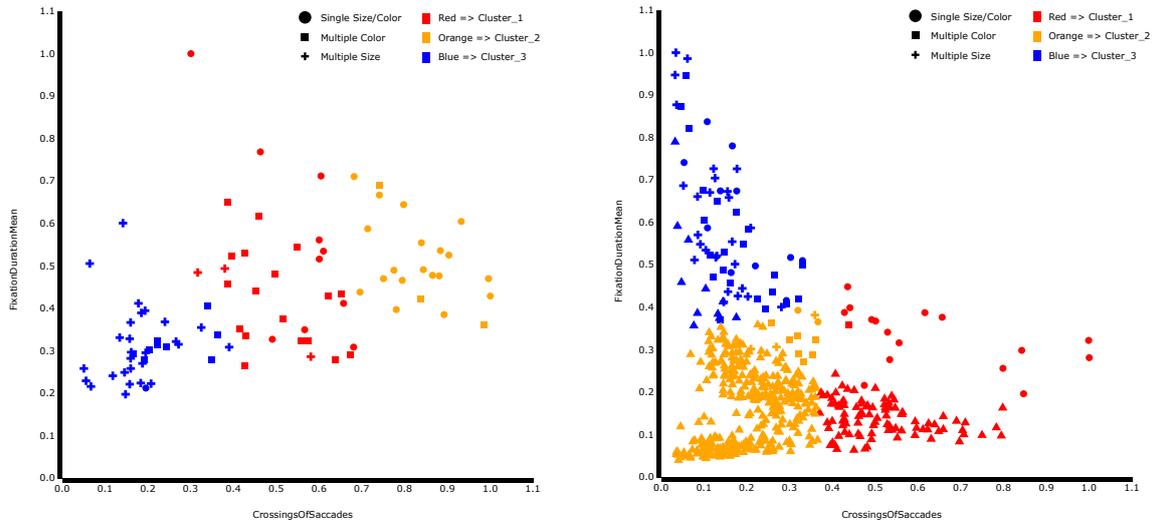


Figure 8.1: Two cluster graph side by side, left graph shows the participant's individual results and the right graph shows all the participants combined results

8.2 Clustering and Correlation Results Analysis

Metrics	1) Crossings Of Saccades	2) First Fixation Duration	3) Fixation Duration Mean	4) Fixation Duration Median	5) Fixation Duration Total	6) Fixations Number	7) Fixations Rate	8) Saccadic Curvature	9) Saccadic Direction Mean	10) Saccadic Direction Median	11) Saccadic Direction Mode	12) Saccadic Number	13) Saccadic Rate	14) Saccadic Velocity Mean	15) Saccadic Velocity Median	16) Scanpath Duration	17) Scanpath Length	18) Scanpath Velocity	19) Start End Bubble Squared Sum	20) Start End Bubble Variance	21) Start End Fixation Squared Sum	22) Start End Fixation Variance
1	-	0.38	0.40	0.38	0.51	0.50	0.40	0.46	0.41	0.41	0.32	0.47	0.38	0.40	0.39	0.41	0.48	0.40	0.39	0.39	0.40	0.38
2	0.1	-	0.33	0.18	0.34	0.54	0.35	0.47	0.17	0.17	0.20	0.51	0.35	0.21	0.32	0.30	0.37	0.24	0.19	0.23	0.22	0.22
3	0.21	0.18	-	0.31	0.33	0.52	0.33	0.48	0.31	0.31	0.26	0.45	0.29	0.27	0.32	0.30	0.37	0.22	0.24	0.28	0.26	0.30
4	0.18	0.14	0.15	-	0.33	0.51	0.40	0.44	0.26	0.34	0.20	0.44	0.33	0.29	0.29	0.30	0.35	0.26	0.22	0.23	0.31	0.31
5	0.08	0.12	0.12	0.13	-	0.54	0.43	0.55	0.36	0.36	0.30	0.56	0.47	0.32	0.38	0.44	0.41	0.37	0.37	0.43	0.35	0.35
6	0.12	0.09	0.17	0.09	0.11	-	0.52	0.54	0.58	0.58	0.31	0.55	0.48	0.54	0.54	0.54	0.53	0.45	0.54	0.50	0.55	0.56
7	0.11	0.08	0.1	0.09	0.09	0.08	-	0.47	0.45	0.45	0.31	0.52	0.44	0.34	0.43	0.48	0.56	0.44	0.40	0.36	0.43	0.43
8	0.1	0.09	0.19	0.09	0.09	0.11	0.09	-	0.47	0.47	0.35	0.49	0.39	0.49	0.50	0.49	0.57	0.38	0.37	0.41	0.49	0.48
9	0.1	0.04	0.19	0.15	0.12	0.09	0.11	0.09	-	0.28	0.22	0.49	0.41	0.25	0.32	0.33	0.38	0.31	0.21	0.24	0.27	0.22
10	0.1	0.07	0.19	0.15	0.12	0.09	0.08	0.09	0.07	-	0.23	0.49	0.41	0.25	0.32	0.33	0.38	0.32	0.21	0.24	0.25	0.24
11	0.1	0.04	0.15	0.11	0.05	0.08	0.06	0.08	0.06	0.06	-	0.35	0.33	0.25	0.30	0.26	0.29	0.26	0.24	0.24	0.19	0.21
12	0.1	0.11	0.19	0.09	0.12	0.1	0.08	0.09	0.09	0.09	0.08	-	0.40	0.54	0.53	0.51	0.55	0.55	0.38	0.40	0.52	0.51
13	0.09	0.09	0.17	0.14	0.06	0.08	0.12	0.08	0.07	0.07	0.06	0.08	-	0.33	0.37	0.44	0.56	0.38	0.31	0.33	0.41	0.39
14	0.09	0.07	0.16	0.14	0.09	0.1	0.1	0.08	0.09	0.09	0.06	0.1	0.07	-	0.29	0.35	0.38	0.27	0.24	0.25	0.22	0.24
15	0.08	0.07	0.13	0.06	0.14	0.08	0.08	0.08	0.06	0.06	0.06	0.08	0.07	0.07	-	0.39	0.39	0.34	0.29	0.29	0.33	0.33
16	0.08	0.08	0.16	0.16	0.14	0.1	0.1	0.07	0.09	0.08	0.07	0.07	0.12	0.09	0.07	-	0.35	0.32	0.32	0.34	0.33	0.32
17	0.1	0.09	0.16	0.13	0.13	0.09	0.1	0.1	0.09	0.09	0.08	0.09	0.09	0.09	0.09	0.09	-	0.40	0.35	0.36	0.37	0.38
18	0.11	0.07	0.16	0.15	0.1	0.09	0.08	0.09	0.08	0.08	0.06	0.11	0.1	0.09	0.07	0.08	0.11	-	0.31	0.25	0.31	0.28
19	0.08	0.03	0.14	0.12	0.05	0.09	0.06	0.07	0.05	0.05	0.06	0.07	0.06	0.07	0.06	0.08	0.09	0.07	-	0.24	0.22	0.22
20	0.08	0.07	0.15	0.13	0.05	0.09	0.06	0.08	0.05	0.05	0.05	0.08	0.06	0.07	0.07	0.07	0.09	0.07	0.05	-	0.27	0.25
21	0.1	0.05	0.18	0.15	0.11	0.09	0.08	0.09	0.02	0.02	0.04	0.11	0.09	0.09	0.06	0.08	0.11	0.1	0.04	0.05	-	0.23
22	0.1	0.04	0.18	0.15	0.12	0.09	0.08	0.09	0.02	0.03	0.04	0.11	0.1	0.09	0.06	0.09	0.11	0.08	0.05	0.05	0.02	-

Table 8.1: Individual metric calculations of all the participants are calculated then the cluster values was found. Moreover, the average of these cluster values measured and shown in the upper half of the graph. Secondly, gaze information of all the participants are combined and the metric calculation processed. The cluster values of these metrics shown in the lower half of the graph. The dark blue values show the highest values within all metric combinations. The light blue values show the best result for particular metric combinations. The dark red values show the lowest values within all metric combinations. The light red values show the worst result for particular metric combinations.

8 Results

Metrics	1) Crossings Of Saccades	2) First Fixation Duration	3) Fixation Duration Mean	4) Fixation Duration Median	5) Fixation Duration Total	6) Fixations Number	7) Fixations Rate	8) Saccadic Curvature	9) Saccadic Direction Mean	10) Saccadic Direction Median	11) Saccadic Direction Mode	12) Saccadic Number	13) Saccadic Rate	14) Saccadic Velocity Mean	15) Saccadic Velocity Median	16) Scampath Duration	17) Scampath Length	18) Scampath Velocity	19) Start End Bubble Squared Sum	20) Start End Bubble Variance	21) Start End Fixation Squared Sum	22) Start End Fixation Variance
1	-																					
2	-0.17	-																				
3	-0.23	0.05	-																			
4	-0.20	-0.00	0.92	-																		
5	0.28	-0.09	0.72	0.72	-																	
6	0.81	-0.25	0.05	0.06	0.56	-																
7	0.59	-0.30	-0.14	0.01	0.30	0.82	-															
8	0.93	-0.20	-0.21	-0.17	0.35	0.89	0.63	-														
9	0.16	-0.02	-0.02	-0.00	0.09	0.15	0.13	0.13	-													
10	-0.07	0.07	-0.03	-0.03	-0.04	-0.05	-0.05	-0.08	0.45	-												
11	0.03	-0.04	-0.02	-0.04	-0.01	0.01	-0.02	0.04	0.19	0.17	-											
12	0.92	-0.22	-0.20	-0.14	0.39	0.94	0.70	0.99	0.14	-0.07	0.03	-										
13	0.72	-0.28	-0.37	-0.26	0.05	0.76	0.91	0.74	0.12	-0.07	0.01	0.7	-									
14	-0.04	-0.14	0.57	0.65	0.65	0.33	0.38	-0.00	0.01	0.02	-0.03	0.08	0.07	-								
15	-0.03	-0.21	0.40	0.55	0.55	0.39	0.52	0.01	0.05	0.03	-0.08	0.11	0.19	0.86	-							
16	0.33	0.06	-0.62	-0.65	-0.43	0.01	-0.32	0.35	0.02	0.01	0.05	0.28	-0.03	-0.66	-0.68	-						
17	0.52	-0.15	0.09	0.11	0.55	0.67	0.27	0.67	0.06	-0.07	0.03	0.67	0.22	0.23	0.17	0.30	-					
18	-0.19	-0.09	0.67	0.70	0.64	0.20	0.18	-0.08	-0.01	-0.01	-0.03	-0.02	-0.11	0.81	0.70	-0.58	0.33	-				
19	0.01	0.03	-0.04	-0.04	-0.02	0.00	-0.02	0.02	-0.02	0.02	0.03	0.01	-0.00	-0.03	-0.06	0.05	0.17	-0.00	-			
20	-0.00	0.04	-0.05	-0.04	-0.04	-0.01	-0.01	-0.01	-0.01	0.02	0.03	-0.01	0.00	-0.02	-0.06	0.04	0.01	-0.00	0.89	-		
21	-0.28	0.26	-0.28	-0.31	-0.37	-0.41	-0.47	-0.36	-0.01	0.06	-0.03	-0.37	-0.43	-0.23	-0.32	0.25	-0.28	-0.27	0.01	0.03	-	
22	-0.31	0.24	-0.16	-0.18	-0.24	-0.34	-0.39	-0.37	0.03	0.05	-0.04	-0.36	-0.41	0.01	-0.16	0.07	-0.26	-0.08	0.02	0.04	0.95	-

Table 8.2: All the participant's gaze information combined and with this data the metric calculation processed. With the metric results, the correlation between these metrics were found and were shown in the graph. Positive values show a positive correlation and the negative values show the negative correlation between the metrics.

9 Conclusion

The question of 'Can we distinguished those different tasks using only two metrics will be answered in this chapter. Concerning the clustering tables, the results show that these metrics do not represent correctly to cluster the gaze data according to the task mode. Some combinations work better than the other such as Saccadic Direction Mean and Number of Fixations. While others do not work that well such as Saccadic Direction Mean and First Fixation Duration.

By looking at the correlation data, the metrics which Fixation Number and Fixation Rate are redundant because they correlate very highly. That means just one of them can be used to analyze the low-level tasks. Like for the instance metric A and B worked well to distinguish the clusters A and C have a high correlation and B and C also work well that means A and C are very similar. It is enough to use one of them. We saw this case in the Saccadic Curvature, Saccadic Number and Fixation Number metrics.

Moreover, if all combinations which used one of this redundant combination correlate well then combinations of these another metric also work wells too distinguish for the low-level tasks.

There are some estimations to determine 'if there is a notable difference between the means of two collections'. It is called inferential statistic which shows the data may be related in specific characteristics. For the recorded data of the three different game modes, it was assumed that those three different game modes would give enough data, but as statistically it has not been calculated. We assumed there are enough data to distribute the game values into three separable clusters.

For the future extension of this study, more metrics can be checked to find a better fit for the low-level task detection. Furthermore, two dimensional two metric combinations can be increased to the three-dimensional metric combinations to find better clustering results. As a visualization respective three-dimensional approach might not be a better choice to reflect on the three dimensional scatter plot. However, it could be beneficial as the detection wise.

As another approach, it could work well to use dimensionality reduction techniques. When the data is in high dimension, it can be reduce by mapping them to 2D. Maybe it could be tried with different dimensionality reduction techniques and analyze after the mapping results to 2D whether that it gives 3 accurate clusters which also correspond to the game modes. This approach goes in the same direction as machine learning because internally machine learning uses dimensionality reduction techniques and creates clusters as using the results. Moreover, for the future extensions of this research, this dataset could be beneficial as it has been tested by some of the inferential statistic tests which prove there are significantly different data sets used in the analysis.

10 Appendix

10.1 Questionnaire Form

Introduction

Started at ____:____

- Info about proceeding of study
 - Welcome to our study
 - You can leave your **belongings** here
 - Let me introduce you to our study **with the Consent** form
 - Please fill in the information form and sign the form
 - Try to select bubbles as fast as possible but be careful not to select wrong
 - Are you an **employee** at the university or does it **pay** you in any other way?
 - Please sign the form and get your money
 - Choose random number for Identification number – Anonymous
 - Let us perform an Eye Test (with and without glasses)
 - Let us calibrate the Eye tracker
 - You can give breaks between the changes of the game modes
 - You may **always ask questions**, if you don't understand something
 - Please **turn off** your **phone**
- Please fill in questionnaire
- Please give feedback

Personal questions (to be filled out by participant)

Sex: Male / Female

Age: _____

Is English your first language? Yes / No

What was your last math grade in school? ____ (1 = very good/best, 6 = failed/worst)

How would you gage your math skills now? ____ (1 = very good/best, 6 = failed/worst)

How many hours do you work with a computer per week? _____

How many hours do you play computer games per week? _____

Degree: High School

Bachelor

Master

State examination

Diploma

Other _____

Primary field of study: _____

Secondary fields: _____

Degree you are aiming for: Bachelor

Master

State examination

Diploma

Other _____

Do you wear optical aids? No / Glasses / Contact lenses

Vision test (read by researcher)

We need to perform a vision test, for evaluation of your study results, later on. The test is specifically tailored to this study and does **not** provide a base for relevant medical evaluation.

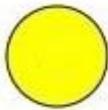
- Look at the **wall**. The distance is 3 meters. Read out the letters in each line, starting at the top, please. (**With and without glasses**)

Snellen test

Lines, which couldn't be read:

- 1 20/200
- 2 20/100
- 3 20/70
- 4 20/50
- 5 20/40
- 6 20/30
- 7 20/25
- 8 20/20
- 9
- 10
- 11

- Can you name these three colors?



Feedback

1. What are your thoughts or general feedback about the game ?

○ Number

○ Color

○ Size

2. Which game mode was the easiest one? (Number, Color, Size)

3. Which game mode was the hardest one? (Number, Color, Size)

10.2 Consent Form



Directors

Prof. Dr. Thomas Ertl
Prof. Dr. Daniel Weiskopf

Contact person

Nils Rodrigues

Contact

Allmandring 19
70569 Stuttgart
T 0711 685-88634
nils.rodrigues@visus.uni-stuttgart.de

www.visus.uni-stuttgart.de
www.twitter.com/vis_visus

Information and declaration of consent for the participation in the research study „Evaluation of Combined Eye-tracking Metrics “

Dear prospective participant,

04.05.2019

We would like to invite you to participate in the following study within the scope of the research carried out at the VISUS. In the current study, „**Evaluation of Combined Eye-tracking Metrics**“, you will be required to play the same game in three different modes. The collected data will be analyzed to find relationships between measurements from eye movements and game modes.

Study procedure

You will see circles (bubbles) on the screen with numbers inside. When you click the bubbles, they will pop and disappear. You should pop the bubbles in numerical order. Start with the smallest and finish with the largest number. The bubbles starting from the smallest number to a higher number until all the bubbles in the screen end. If you click the bubbles in the wrong order, they will not pop. Numbers will not always start with '1'; they could start with any value. There are three game modes. You will play multiple games in a single mode before advancing the next.

In the game mode "numbers", all bubbles will be the same color and size. You will need to read the numbers in order to find the right sequence and pop all bubbles.

In the game mode "size", the bubbles will be the same color again, but their sizes will increase according to their number.

In the mode "color", the bubbles will be reverted to the same size again, but their colors will change according to their numbers. Please memorize the scale before starting the study.

Mode: Number



Mode: Size



Mode: Color



Bank

Baden-Württembergische Bank
Stuttgart – BW-Bank

IBAN

DE51 6005 0101 7871 5216 87

SWIFT/BIC

SOLADEST600



Processing of the measured data

The datasets collected during the study are available only in anonymous form that cannot be linked to a specific person. They are processed with the results of other participants to obtain mean values, which are evaluated statistically and published in scientific journals or conference proceedings. You indicate your consent to the use and publication of your data for scientific purposes. This consent may be withdrawn at any time by a simple written statement to that effect. No reasons for that decision need to be given.

Data privacy

Any personal data, which can be tracked back to you, will be treated confidentially and not handed to third parties outside the research project or published in any form. If you wish an information transfer to third parties, you must explicitly express this via written consent. Only authorized research staff of the research project have access to your data. The computers on which such data is stored are connected to the Internet and are specially secured against unauthorized access. Legal access from outside is not possible.

Expense allowance

There are no costs associated with your participation. You will be paid an allowance of 10 € (EUR) in cash from the funds of the Collaborative Research Center/Transregio 161. <http://www.sfbtrr161.de/>

Risks and Rights

There are no elevated risks associated with this study. The use of a computer screen and keyboard is not harmful and does not entail any higher health risks, when compared to home use.

Your eye movement and gaze position on the screen will be recorded via remote cameras. They use low intensity infrared light to illuminate the outside of your eye balls. Infrared light carries less energy than visible light and it is around you every day in the form of heat or tv remote signals. It does not pose an elevated risk to your health. The remote eye tracker used in the experiment because it does not touch the eye. It is just sitting there at the bottom of the screen.

Your participation in this study is voluntary. By giving your informed consent you are under no obligations. You may revoke your consent at any time without any legal consequences. You may abort the study at any time without giving reasons. Doing so will not result in any legal consequences for you.

Declaration of informed consent

- I have read or have been read to the preceding explanation and understood it.
- I have been informed about the study by the research staff and all my question have been answered to my satisfaction
- I know that personal data will only be processed anonymously and that all authorized research personnel, which have access to my personal data, are obligated to confidentiality.
- I have been informed, that the obtained data is saved and processed on computers that are connected to the internet.
- I volunteer to participate in this study, and I am aware of the fact that I can discontinue my participation at any time.

Last name, first name

Place, Date and signature of participant

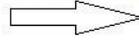
Date and signature of experimenter



Color Scale

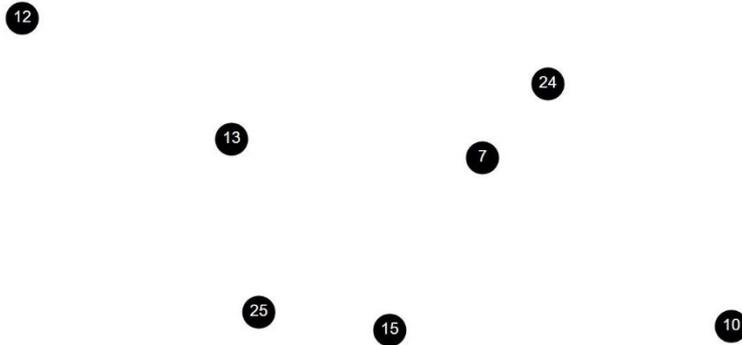


Lower Values

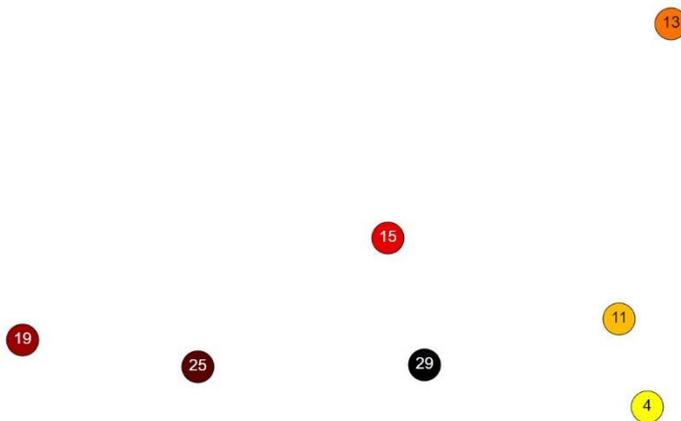


Higher Values

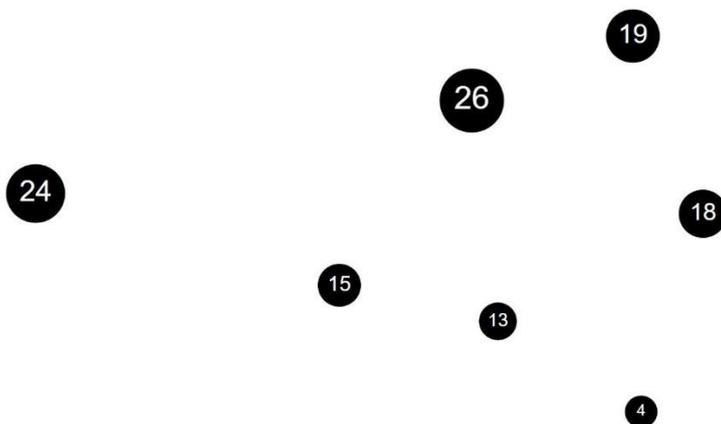
Game Mode "Numbers" Example:



Game Mode "Color" Example:



Game Mode "Size" Example:



Bibliography

- [AC99] T. J. Andrews, D. M. Coppola. “Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments”. In: *Vision research* 39.17 (1999), pp. 2947–2953 (cit. on p. 19).
- [BCHC09] J. Benesty, J. Chen, Y. Huang, I. Cohen. “Pearson correlation coefficient”. In: *Noise reduction in speech processing*. Springer, 2009, pp. 1–4 (cit. on p. 35).
- [Eps15] N. Epstein. *K-means Clustering*. <https://github.com/NathanEpstein/clusters>. [Online; accessed 15-May-2019]. 2015 (cit. on p. 35).
- [EW07] C. Ehmke, S. Wilson. “Identifying web usability problems from eye-tracking data”. In: *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1*. British Computer Society. 2007, pp. 119–128 (cit. on p. 20).
- [Gal89] N. Galley. “Saccadic eye movement velocity as an indicator of (de) activation. A review and some speculations”. In: (1989) (cit. on p. 18).
- [GK99] J. H. Goldberg, X. P. Kotval. “Computer interface evaluation using eye movements: methods and constructs”. In: *International Journal of Industrial Ergonomics* 24.6 (1999), pp. 631–645 (cit. on p. 20).
- [Ham+89] L. Hamers et al. “Similarity measures in scientometric research: The Jaccard index versus Salton’s cosine formula.” In: *Information Processing and Management* 25.3 (1989), pp. 315–18 (cit. on p. 34).
- [HH99] J. M. Henderson, A. Hollingworth. “High-level scene perception”. In: *Annual review of psychology* 50.1 (1999), pp. 243–271 (cit. on p. 20).
- [HNA+11] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, J. Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011 (cit. on pp. 11, 17–23, 39, 40).
- [JC80] M. A. Just, P. A. Carpenter. “A theory of reading: From eye fixations to comprehension.” In: *Psychological review* 87.4 (1980), p. 329 (cit. on p. 19).
- [JHML12] R. Johansson, K. Holmqvist, F. Mossberg, M. Lindgren. “Eye movements and reading comprehension while listening to preferred and non-preferred study music”. In: *Psychology of music* 40.3 (2012), pp. 339–356 (cit. on p. 19).
- [Jou15] P. Journalism. *Scale Color Perceptual*. <https://github.com/politiken-journalism/scale-color-perceptual>. [Online; accessed 15-May-2019]. 2015 (cit. on p. 45).
- [KNB+16] A. Kumar, R. Netzel, M. Burch, D. Weiskopf, K. Mueller. “Multi-similarity matrices of eye movement data”. In: *2016 IEEE Second Workshop on Eye Tracking and Visualization, ETVIS 2016, Baltimore, MD, USA, October 23, 2016*. 2016, pp. 26–30. DOI: 10.1109/ETVIS.2016.7851161. URL: <https://doi.org/10.1109/ETVIS.2016.7851161> (cit. on pp. 11, 13, 15).

- [KNB+18] A. Kumar, R. Netzel, M. Burch, D. Weiskopf, K. Mueller. “Visual Multi-Metric Grouping of Eye-Tracking Data”. In: *Journal of Eye Movement Research* 10.5 (2018), p. 11 (cit. on pp. 11, 13–15, 17, 19, 20).
- [LCHK91] C. Lueck, T. Crawford, H. Hansen, C. Kennard. “Increase in saccadic peak velocity with increased frequency of saccades in man”. In: *Vision research* 31.7-8 (1991), pp. 1439–1443 (cit. on p. 18).
- [MS96] D. K. McGregor, J. A. Stern. “Time on task and blink effects on saccade duration”. In: *Ergonomics* 39.4 (1996), pp. 649–660 (cit. on p. 18).
- [NH06] L. Nummenmaa, J. K. Hietanen. “Gaze distractors influence saccadic curvature: Evidence for the role of the oculomotor system in gaze-cued orienting”. In: *Vision research* 46.21 (2006), pp. 3674–3680 (cit. on p. 18).
- [NTS02] M. Nakayama, K. Takahashi, Y. Shimizu. “The act of task difficulty and eye-movement frequency for the ‘Oculo-motor indices’”. In: *Proceedings of the 2002 symposium on Eye tracking research & applications*. ACM. 2002, pp. 37–42 (cit. on pp. 20, 21).
- [Ols12] A. Olsen. “The Tobii I-VT fixation filter”. In: (2012) (cit. on pp. 39, 41).
- [PHG+04] B. Pan, H. A. Hembrooke, G. K. Gay, L. A. Granka, M. K. Feusner, J. K. Newman. “The determinants of web page viewing behavior: an eye-tracking study”. In: *Proceedings of the 2004 symposium on Eye tracking research & applications*. ACM. 2004, pp. 147–154 (cit. on p. 20).
- [PPKS02] G. Paliouras, C. Papatheodorou, V. Karkaletsis, C. D. Spyropoulos. “Discovering user communities on the Internet using unsupervised machine learning techniques”. In: *Interacting with Computers* 14.6 (2002), pp. 761–791 (cit. on pp. 11, 14, 15).
- [PW03] N. L. Port, R. H. Wurtz. “Sequential activity of simultaneously recorded neurons in the superior colliculus during curved saccades”. In: *Journal of Neurophysiology* 90.3 (2003), pp. 1887–1903 (cit. on p. 18).
- [Röt01] M. Rötting. *Parametersystematik der Augen-und Blickbewegungen für arbeitswissenschaftliche Untersuchungen*. Shaker, 2001 (cit. on p. 20).
- [Sne68] H. Snellen. *Probebuchstaben zur bestimmung der sehschärfe*. Peters, 1868 (cit. on p. 48).
- [TBG05] B. W. Tatler, R. J. Baddeley, I. D. Gilchrist. “Visual correlates of fixation selection: Effects of scale and time”. In: *Vision research* 45.5 (2005), pp. 643–659 (cit. on p. 21).
- [TVPK13] G. Tourassi, S. Voisin, V. Paquit, E. Krupinski. “Investigating the link between radiologists’ gaze, diagnostic decision, and image content”. In: *Journal of the American Medical Informatics Association* 20.6 (2013), pp. 1067–1075 (cit. on p. 11).
- [VLMJ01] K. F. Van Orden, W. Limbert, S. Makeig, T.-P. Jung. “Eye activity correlates of workload during a visuospatial memory task”. In: *Human factors* 43.1 (2001), pp. 111–121 (cit. on p. 21).
- [Wes13] M. West. *Pearson Correlation*. <https://gist.github.com/matt-west/6500993>. [Online; accessed 15-May-2019]. 2013 (cit. on p. 35).
- [WH07] G. A. Wilkin, X. Huang. “K-means clustering algorithms: implementation and comparison”. In: *Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007)*. IEEE. 2007, pp. 133–136 (cit. on p. 35).

All links were last followed on May 15, 2019.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

place, date, signature