Gaze-Tracking Analysis for Cognitive Screening and Assessment

by

Sarbari Sarkar

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Author .....
Department of Electrical Engineering and Computer Science
May 18, 2020

Certified by .....
Randall Davis
Professor
Thesis Supervisor

Certified by .....
Dr. Dana L. Penney
Director of Neuropsychology, Lahey Hospital & Medical Center
Thesis Co-Supervisor

Accepted by .....
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Gaze-Tracking Analysis for Cognitive Screening and Assessment

by

Sarbari Sarkar

Submitted to the Department of Electrical Engineering and Computer Science
on May 18, 2020, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Neurodegenerative diseases degrade the mental and physical capabilities of afflicted individuals around the world. Early diagnosis can make it possible to reduce the effects and progression of these diseases. The novel digital Symbol Digit Test (dSDT) is a new cognitive test that judges patterns of recall and cognitive associations, which can be used to differentiate between cognitive signs displayed by normal and neurologically impaired subjects. Our research identifies different strategies of learning and recall, and automates the process of analyzing the eye-tracking data collected from the dSDT to detect these patterns. This work paves the foundation for future studies to assess differences between healthy and impaired individuals, and model these features to detect and aid in the diagnosis of cognitive states of individuals.

Thesis Supervisor: Randall Davis
Title: Professor

Thesis Co-Supervisor: Dr. Dana L. Penney
Title: Director of Neuropsychology, Lahey Hospital & Medical Center

Acknowledgments

I would like to thank my advisor, Professor Randall Davis, for his continual support throughout the course of this project. His suggestions and encouragement were crucial in driving the research forward. I would also like to thank Dr. Dana Penney for her detailed observations and insightful explanations, and for always motivating me with her kind words. Professor Davis and Dr. Penney were both very generous with their time and advice, and I am so thankful for all of their guidance.

I would also like to thank my lab mates: Elizabeth DeTienne, Karunya Sethuraman, and Shang-Yun (Maggie) Wu. They were always willing to lend an ear, and were wonderful people to work with and around. Additionally, I would like to thank all the people who made my years at MIT so valuable: my professors, for teaching me to think outside the box and hone my technical skills; and my friends, who continually inspired me and cheered me on.

Last but not least, I would like to thank my amazing parents and older sister for always being so loving and supportive. Thank you for everything; I could not be where I am today without you.

Contents

1 Introduction **13**

2 Background **15**

2.1 Cognitive Screening Tests 15

2.2 Eye-Tracking in Cognitive Studies 16

2.3 Digital Symbol Digit Test (dSDT) 17

2.3.1 Tasks 17

2.3.2 Testing Procedure 19

2.3.3 Application 19

3 Prior Data Collection **21**

4 Data Processing and Visualization **23**

4.1 Fixation Detection 23

4.2 Data Aggregation and Interpolation 25

4.3 Base Observations 25

4.4 Visualization 27

5 Pattern and Behavior Analysis **29**

5.1 Visual Match Behavior 32

5.1.1 No Match Required 32

5.1.2 Concentrated Matches 34

5.2 Forward and Backward Fixations 35

5.3 Leftward and Rightward Scans 37

5.4 Back-and-Forth Fixations 39
5.5 Spatial Associations 41
5.6 External Fixations 42
5.7 Recall Metrics 44

6 Conclusion **47**
6.1 Contributions 47
6.2 Future Work 49

A Supplementary Information **51**

List of Figures

1-1 Layers of abstraction to derive behavior from raw data 14

2-1 An image of the SDMT form 16

2-2 A demo version of the test form for the translation task. 18

2-3 A demo version of the test form for the copy and delayed recall tasks 19

4-1 Fixations detected through the saccadic-velocity and dispersion-duration methods 24

4-2 A view of the spatially distributed test form 26

5-1 Matches vs. no matches in the translation task 33

5-2 Average of maximum cells gazed ahead across trials in the translation and copy tasks. Note: Cells 49-54 were dropped when comparing the translation and copy tasks since there was no copy task counterpart. . 36

5-3 A comparison of the aggregated percentages of the different types of scans during the translation task between trials 38

5-4 A comparison of the counts of back-and-forth fixations across participants and trials during the translation task 40

5-5 Side-by-side comparison of a common key pairing found in back-forth fixations: X and four-sided star 41

5-6 A comparison of percentage of spatial associations made in the translation task across participants and trials. 42

5-7 Aggregated fixation durations across participants and trials for each of the six cells in the delayed recall task 44

A-1 This is a still from the visualization tool, depicting the playback speed percentage, time since start of the video, the strokes made by the Anoto pen in blue, and the current gaze (or fixation) location as the gray dot 51
A-2 The Fixation Detector and Surface Tracker are plugins accessible from the Pupil Player 52

List of Tables

5.1 The following metrics were collected with respect to each cell for which the participant was filling out the associated digit for the symbol. . . 29

5.2 Behaviors observed and computationally detected from participants' test session recordings 30

5.3 Measurements of the frequency of observed behaviors, present in at least one participant. 31

5.4 Contingency table of aggregated cells in the translation task 33

5.5 Average match time per trial in the translation task 34

5.6 The correlation coefficients and associated p-values for average maximum cells looked ahead, compared between tasks and trials. 37

5.7 A comparison of the percentages of the types and total amount of external fixations across participants and trials in the delayed recall task 43

6.1 Summary of the layers of abstraction built to analyze patterns and behaviors from raw participant data 48

6.2 List of the behaviors and metrics for which detection was automated. 48

Chapter 1

Introduction

Every year, millions of people around the world are affected by different types of neurodegenerative diseases, such as Alzheimer's and Parkinson's [1]. These disorders result from the deterioration of neurons, nerve cells that are responsible for transmitting information throughout the body. Over time, impaired individuals slowly lose their cognitive or physical abilities, or a combination of both. Initially, the symptoms are generally mild, ranging from basic coordination problems to minor lapses in memory, but eventually they progress and drastically reduce the quality of life of afflicted patients. These diseases are generally incurable. It is beneficial for symptoms to be accurately detected as early as possible, as treatments can be administered to slow disease development.

While many cognitive tests have been developed to address the growing need for detection of cognitive decline, a major limitation in the current technology lies in the uniformity of its evaluation. Since the scoring depends heavily on the judgment of clinicians, the reproducibility of these tests is limited [2].

The digital Symbol Digit Test (dSDT) is a novel cognitive test that collects spatial and temporal information about the writing motions made on the test surface. We can also collect gaze data from individuals who wear eye-tracking headsets while taking the dSDT. This thesis has worked to automate the analysis of the eye-tracking data, in conjunction with the ink data, to determine the best ways to combine features and evaluate strategies of learning and memory, and discover new insights that enhance

our understanding of cognitive behavior.

In this work, we first manually identified types of participant behavior from viewing recordings of test sessions. Then, we computationally isolated eye movements of interest, known as fixations, from the gaze data. We merged the fixation data with the ink data and generated base observations in the context of the test form (e.g., the location of the gaze with respect to the current area being written in). Then, we implemented searches in a subset of participant data for the behaviors we initially identified, and commented on our findings with regards to our preliminary hypotheses. The levels of abstraction generated from our research, as shown in Figure 1-1, will serve as a foundation for later work to build on in differentiating between healthy and impaired participants.

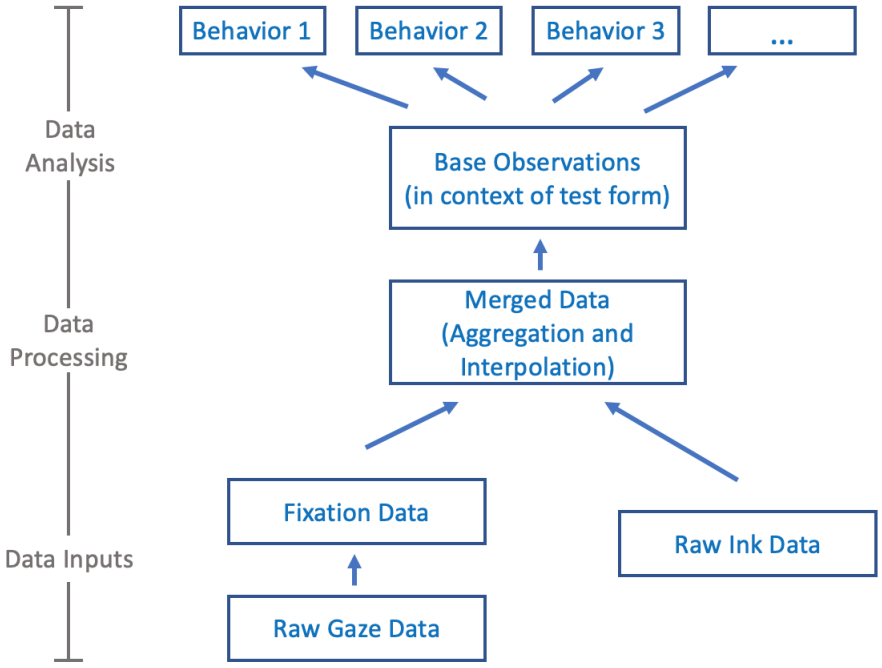

Figure 1-1: Layers of abstraction to derive behavior from raw data

Chapter 2

Background

2.1 Cognitive Screening Tests

Of the many types of screening tests for cognitive status, two of the most common are the Mini-Mental State Exam (MMSE) and Montreal Cognitive Assessment (MoCA). The MMSE is widely used in diagnosing Alzheimer's disease, among other forms of dementia. It screens a range of abilities, using tasks including counting backwards and identifying everyday objects. The other test, MoCA, is somewhat more comprehensive, particularly for memory and language domains. Both screening tests have been found to perform well in identifying cognitive issues in dementia including Alzheimer's and Parkinson's disease. The MoCA involves drawing, animal naming, and word list learning. Both of these tests are traditionally taken in pen-and-paper format, with oral portions noted by clinicians [3].

The traditional Symbol Digit Modalities Test (SDMT) is widely used to measure visuo-motor processing speed, and is sensitive to change in cognitive performance in subjects with combined cognitive and motor impairments, such as multiple sclerosis. It can also be used together with the MMSE and the MoCA to diagnose cognitive impairments [4]. The SDMT typically presents the subject with a view of nine assorted complex symbols, each paired with a digit between 1-9, as shown in Figure 2-1. Participants are directed to fill in blank cells under the provided symbols with their associated numbers in a specified time frame, often 90 seconds. This test seeks

to determine attention span, visual scanning abilities, and writing speed in a timed environment [5].

The dSDT explored in this work is a novel approach to the traditional paradigm.

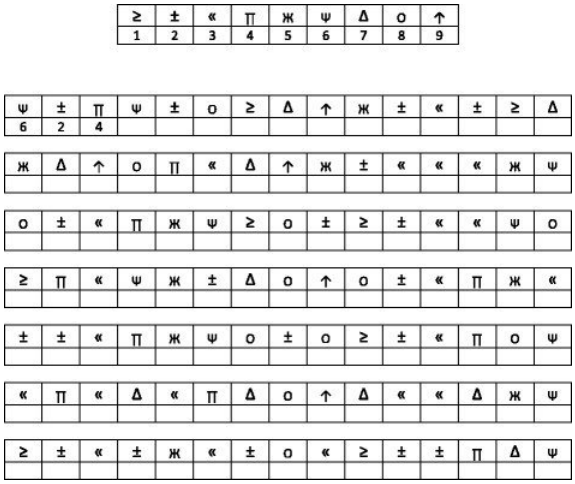

Figure 2-1: An image of the SDMT form

2.2 Eye-Tracking in Cognitive Studies

Eye-tracking has long been applied to cognitive studies, as analysis of eye movements can reveal information about a participant's cognitive ability. When viewing a static eye scene, the two most common types of eye movement are fixations and saccades [6]. During fixations, the eyes are focused in an area for an extended period of time (typically at least 100 milliseconds) so that the visual information can be processed. Saccades are rapid eye movements that occur between fixations.

In one study [2] durations of fixations in regions of interest were compared to distinguish individuals showing signs of dementia from healthy controls. It was found that healthy controls fixated for longer periods of time than participants with dementia. Another study focused particularly on analyzing fixation positions and their durations in participants taking the SDMT, concluding that there did not seem to be significant correlation between the number or duration of fixations and participants of three classes: healthy controls, Parkinson's disease (PD) with normal cognition, and PD with cognitive impairments [7]. Drawing inspiration from these studies, we

analyze fixation data to determine where participants direct their visual attention as they complete the test.

2.3 Digital Symbol Digit Test (dSDT)

The digital Symbol Digit Test (dSDT) was developed by Dr.Dana Penney and Professor Randall Davis to capture the dynamic aspects of learning and memory missed by traditional cognitive tests. The dDST consists of three tasks, and is administered twice consecutively.

2.3.1 Tasks

The first task is translation. In this task, participants are provided with a key of six symbols and associated numbers, as shown in Figure 2-2. Beneath the key, there are four rows of symbols, with blank cells underneath for individuals to fill in with the corresponding numbers by matching with the key. Individuals are given the opportunity to practice and demonstrate understanding of the test by filling in the first six symbols in the Sample section, after which they receive instructions to complete the rest of the translation test as rapidly and as accurately as possible without further interruption. The symbols are placed randomly and evenly distributed throughout the test, allowing for equal opportunity to learn pairings.

The second task is copying. The copy task is almost identical to the translation task in its layout in that there is a key in the upper middle of the page, and four rows of cells to fill in. The task once again is to hand-write the appropriate digit in each cell, but as the key makes clear, the digits are associated with themselves rather than symbols. As a result, participants do not actually have to look up anything in the key and can just copy the number into each blank cell. Similar to the translation section, participants are able to fill out the first six cells to understand the task and ask any questions, after which they are directed to complete the remainder of the test as quickly and accurately as possible, but without any time restriction.

The copy task does not aim to test any associative ability, particularly as the

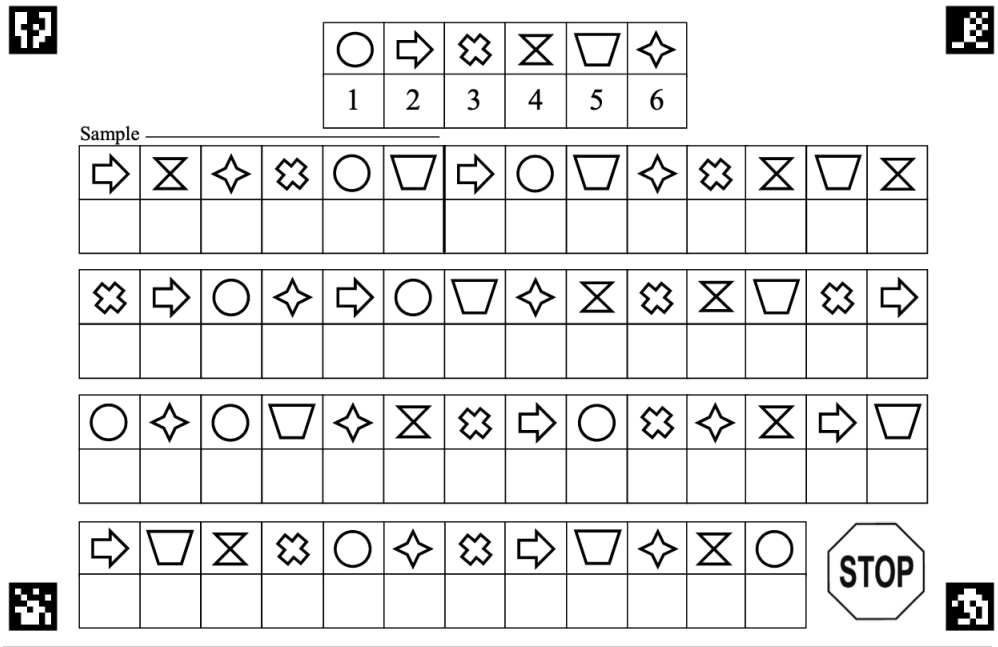

Figure 2-2: A demo version of the test form for the translation task.

presence of the key is unnecessary, but can provide information as to how individuals relate and respond to similar tasks, and set a basis for writing speed in relation to a task that requires low cognitive load. Unknown to the subject, the answer for each corresponding cell in both the translation and copy tasks are identical. By having the same responses in both tasks, we are able to measure the participants' performances under various cognitive loads. This also allows us to use each participant as their own control, enabling the separation of physical and cognitive components of performance.

The third task, delayed recall, presents the participant with the same six symbols as in the translation task, but in a different order from the original key. The participant is directed to fill in the blank cells with their associated numbers as best they can recall from memory, since the key from the translation task is not is view. The presence of the copy task on the same page allows participants to note which numbers are valid to use, as well as potentially identify patterns that individuals employ to make and call upon connections between symbols and numbers as they perform the final recall task.

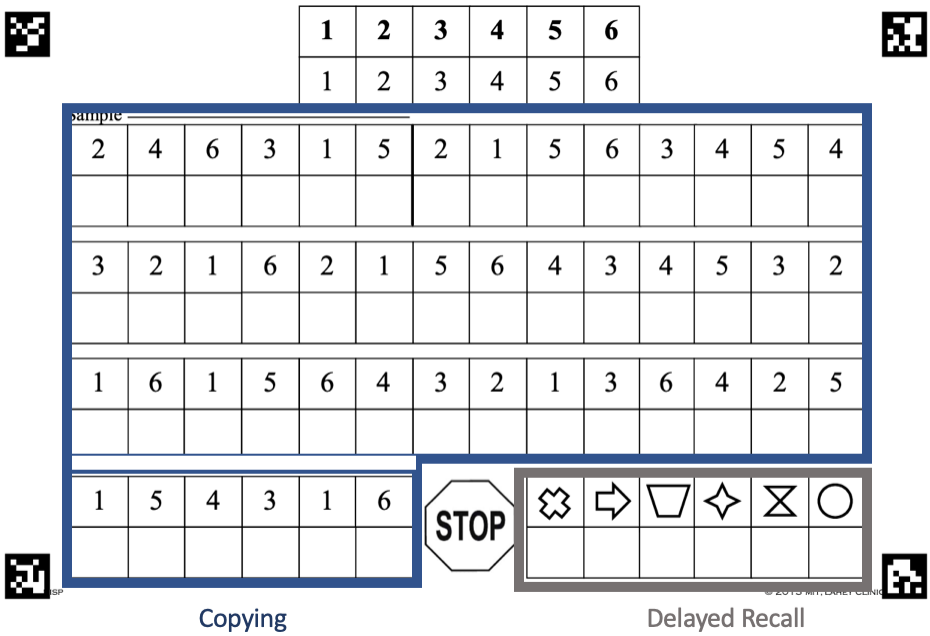

Figure 2-3: A demo version of the test form for the copy and delayed recall tasks

2.3.2 Testing Procedure

Participants are able to work at their own pace, self-balancing speed and accuracy. Upon completion of all three tasks, they are told that they will have to complete the tasks again. The second administration allows data to be captured about how people maintain or adjust their cognitive strategies over two administrations. Participants are not informed in the first administration that there will be a delayed recall task, allowing assessment of incidental learning. For healthy individuals, knowledge gained from the first administration is expected to inform behavior for the second administration. Information gained from experience may cause them to make additional efforts to learn the symbol-digit associations, i.e. to adjust their cognitive strategy.

2.3.3 Application

The combination of gaze and ink data in digitized format opens avenues for computational analysis of behavior. Much of the research on the dSDT to date has been concentrated on writing behaviors, constructing features from the digital ink data for machine learning models to predict diagnostic classifications for healthy and im-

paired participants, including those with clinical Alzheimer's or Parkinson's disease [8]. Here we explored gaze analysis in conjunction with the digital ink data, to further enhance our understanding of cognition and to identify novel, relevant markers of differentiation.

Chapter 3

Prior Data Collection

In prior IRB-approved work, a sample of healthy student volunteers were recruited from the Massachusetts Institute of Technology campus and administered the digital Symbol Digit Test under controlled eye-tracking conditions. These subjects served as healthy controls [9].

All testing occurred in a windowless room (for consistency of illumination), and procedures were followed to reduce glare to facilitate eye tracking data capture. Subjects were requested for consent and fitted with a Pupil Labs headset, after which the calibration process was carried out.

Eye-tracking data was collected from the headset, which used a cameras per eye to track pupil location, as well as a front-facing "world" camera to record the wearer's view. Samples were collected at a rate of 240 Hz and marked by timestamps through the Pupil Capture software. Ink data was collected from the Anoto pen that participants used to complete the tests; the pen measured its position on the page every 12.5 milliseconds.

Subjects received tests in succession. Three of these tests were the digital Maze tests for a different study, also related to cognitive state detection. Then, there were two administrations of the digital Symbol Digit Test. Following administration instructions reported earlier, participants were not informed beforehand that the two tests would be identical. To maintain confidentiality, the data for each subject was anonymized and stored in encrypted format per IRB-approved procedures.

Of the data collected, three subjects provided useful information for our work. In this work, we built upon the components of the past research [10] to explore a variety of patterns aimed at detecting the cognitive traits and strategies displayed by subjects. Novel contributions of this work include the definition of key features, and the automation of the measurement of features to detect and quantify interesting problem solving strategies.

Chapter 4

Data Processing and Visualization

The data analyzed in this work is composed of the collected ink and gaze information of three subjects who took the dSDT with fiducial markers,[1] so that their gaze location on the page could be determined.

The Pupil software includes a fixation detector and the ability to export the world video. The exporter produces video recordings of the test augmented with gaze and/or fixation points indicated graphically. The Pupil software can additionally generate spreadsheets listing timestamps of collected gaze points, normalized x and y positions of the gazes on the test form based on the fiducial markers, the confidence level of detecting the pupil, and whether or not the gaze was contained within the test surface.

4.1 Fixation Detection

We used fixations to determine where participants were focusing their visual attention. The fixation detector in the Pupil software identifies fixations on the basis of the dispersion and duration: to be a fixation, the gaze points in the focused area must be contained within a maximum distance (dispersion threshold) for at least a minimum duration.

The dispersion threshold is expressed in terms of visual angle between gaze points;

[1] Fiducial markers are graphical figures which are easily detected with computer vision technology, typically placed at the four corners of the test form to enable detection and localization of the form.

we used 1.5 degrees. Fixations are generally defined as at least 100 milliseconds long, and thus we followed the standard to set our minimum duration threshold. The detector also combines sequential fixations, within a maximum duration, into a single event. We empirically set the threshold for maximum duration to 220 milliseconds.

Another simple method for separating fixations from saccades is through analysis of gaze speed, as researched previously in the lab [9]. While the dispersion-duration principle identifies fixations amid saccades, this method classifies saccades amid fixations. Any gaze samples above a set speed threshold are considered saccades. Following the Pupil Labs guidelines, we kept only gaze points that were detected above a confidence threshold of 0.8. We additionally filtered out any gaze points that were labeled "off-surface", such as glances on the table or surroundings rather than on the test form. Then, speeds were calculated for gaze points by using a window three samples wide; any samples with speeds lower than 3 inches per second were classified as fixations.

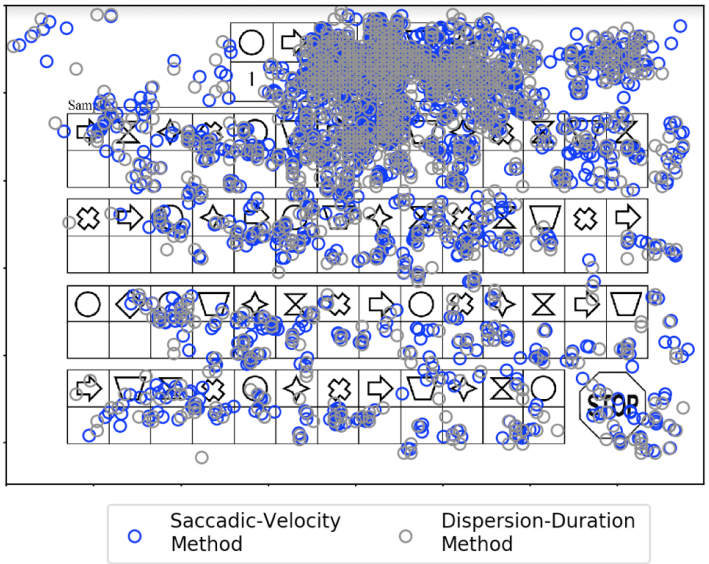

Figure 4-1: Fixations detected through the saccadic-velocity and dispersion-duration methods

Many of the same points were classified as fixations by both the Pupil Fixation detector and the gaze speed method, as shown in Figure 4-1. However, several fixations on the left side of the form were detected only by the Pupil Labs detector, and

confirmed by viewing the video recordings. The analysis for the remainder of this work is therefore carried out on fixations detected through the Pupil Labs software.

4.2 Data Aggregation and Interpolation

We aggregated the gaze points with the ink stroke information, combining the data points based on their Unix timestamps. The minimum fixation duration of 100 milliseconds was used as the sample interval. Since the ink data measurements were made every 12.5 milliseconds, we had to downsample the data. We chose the first measurement that appeared every 100 milliseconds, because the locations of the ink strokes did not change significantly within each interval.

We then interpolated the ink data to account for discrepancies in the data collection. Within every window of 3 ink samples, if there was missing data between the first and last ink data samples, ink positions were interpolated through averaging of the normalized x- and y- positions of the available samples in the window.

The Pupil software reports the start timestamp, location, and duration for each fixation. Based on this information, we replicated the normalized x- and y- positions on the test form every 100 milliseconds from the start time for the duration of each fixation. The periods of time between fixations were assumed to be representative of other types of eye movements, mainly saccades. Further in-depth classification of the gaze data can reveal whether other types of eye movements were indeed present, or if there was missing data due to the limitations of the equipment. For the purposes of this work, we focused on the available fixation data.

4.3 Base Observations

The test form was categorized into cell positions (e.g., Key 1-6 and Cells 1-54 for the translation task). Each cell consists of two slots: there is a symbol, which can either be a geometric figure or numeric digits, in the upper slot, and associated digits in the lower slot. We categorized the cells by their spatial positions, as shown in Figure 4-2.

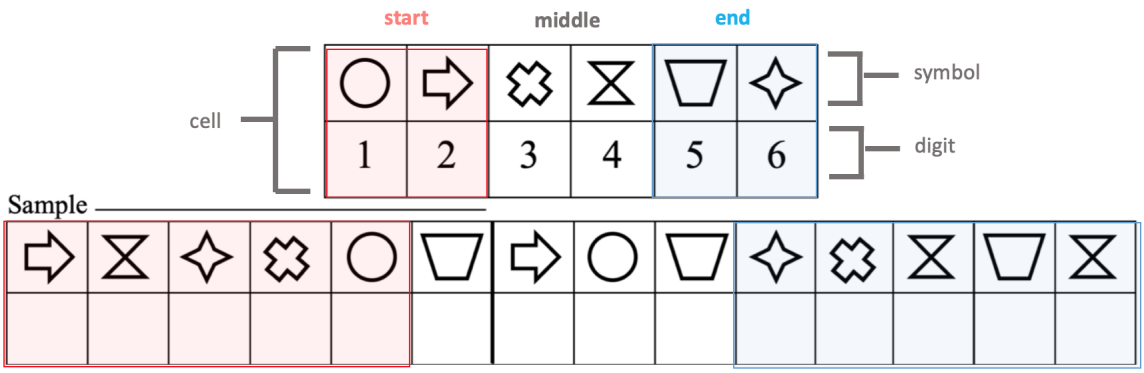

Figure 4-2: A view of the spatially distributed test form

We then classified the gaze and ink positions based on their locations on the paper, with respect to the cell positions. Many of the participants' fixations were located on the surface of the test form, but outside of bounds of the cell positions. This could be attributed to participants resting during transitions between cells or fixating their eyes on blank spaces beyond the task area during moments of thought.

We generated base observations on the aggregated datasets. We separated the data by event, based on task (e.g., translation) and test administration (e.g., Trial 1). Then, we further separated the data based on which cell was being currently worked on. Ink and gaze data was collected for each cell in the time between the end timestamp of the last ink stroke in the previous cell (or the start time of the first stroke, if there was no previous cell) and the end timestamp of the last ink stroke in the current cell.

Based on the symbols present in the current cells that were being gazed at or written in, we determined whether or not participants glanced at the key or other cells on the test form.

Among the data we collected were the number of gaze points or fixations while working to fill out a cell and the symbols gazed upon in the duration. This information built the foundation for detailed pattern search in the analysis phase.

4.4 Visualization

We developed a visualization tool using the Python matplotlib library to show the rapidly evolving gaze positions over the course of the test, with viewing speed set as an adjustable parameter. The tool plots the gaze data, either raw or of generated fixations on the test surface, based on the chosen speed. An image of the tool in action while processing data is provided in the Appendix. We used this tool mainly to identify positional shifts in the gaze points due to calibration inconsistencies, in order to adjust the data accordingly, as well as search for any patterns at reduced viewing speed. This tool can also be used to visually validate any computational observations generated from the gaze data.

Chapter 5

Pattern and Behavior Analysis

We identified behaviors displayed by participants throughout the administrations of the dSDT. We generated metrics, listed in Table 5.1, the first three of which were used to facilitate behavior detection. Metrics 4 and 5 were collected to be used as potential future measures of differentiation between participants. We implemented the detection of the behaviors detailed in Table 5.2.

Table 5.1: The following metrics were collected with respect to each cell for which the participant was filling out the associated digit for the symbol.

Metric	Description
Cell Completion Duration	Time spent from the last stroke of the previous cell to the last stroke of the current cell
Pre-Cell Writing Delay	Time between the end of the last stroke in the previous cell and the first stroke in the current cell
Average Fixation Duration	Average duration of the aggregated fixations during the stimulus cell duration
Average Ink Speed	Average writing speed based on the ink measurements, calculated for each stimulus cell
Gaze/Fixation Distance	Total visual distance covered during the stimulus cell duration, calculated from the gaze point/fixation data

Table 5.2: Behaviors observed and computationally detected from participants' test session recordings

Behavior	Description	Applicable Tasks	Types
Visual Match (General)	Fixation on another cell with the same symbol as the cell that has to be filled out	Translation Copy	*key-match* *cell-match* *key-and-cell-match* *no-match-required*
Visual Match (Concentrated)	Specific visual match with respect to other cells	Translation Copy	*concentrated* *non-concentrated*
Forward and Backward Fixation	Looking back at cells to the left or ahead at cells to the right in the same row	Translation Copy	*forward-fixation* *backward-fixation*
Scans	Consecutive fixations across three or more adjacent cells	Translation Copy Recall	*key-scan (left/right)* *same-row-scan (left/right)*
Back-and-Forth	Successive fixations in the format: x-y-x, where x and y are different cells	Translation Copy Recall	*key-key* *cell-cell* *key-cell/cell-key* *confirmation*
Spatial Association	First fixation in same spatial section as cell to be filled out	Translation Copy Recall	*spatial-association* *no-spatial-association*
External Fixation	Fixation in a portion of the test form that is not relevant to the current task	Copy Recall	*blank-fixation* *external-key-fixation* *external-cell-fixation*
Sequential Recall	Completion of the delayed recall task in sequential order	Recall	*sequential-recall* *non-sequential-recall*

Our initial work focused on patterns we observed in participant video recordings. Of those patterns, we chose the ones that showed up more than half the time in at least one participant. The behaviors described in Table 5.3 were observed more than 50% of the time in at least one task-trial combination for a participant. We then expanded the general Visual Match behavior to search for whether or not matches were generally concentrated, or occurred after exploring other cells.

Table 5.3: Measurements of the frequency of observed behaviors, present in at least one participant.

Behavior	Participant	Measurement	Additional Notes
Visual Match (General)	3	78% of all cells in the task	Translation task, Trial 1
Forward and Backward Fixations	1	74% of all cells in the task	Copy Task Trial 2
Spatial Association	2	56% of all cells in the task	Translation task, Trial 1
External Fixation	3	84% of all fixations in the task	Delayed recall task, Trial 1

We further focused on scans and back-and-forth, multi-cell behaviors. For all translation tasks among all our participants, Scans showed up in the range of 7-30 times, while back-and-forth behavior appeared in a wider range of 3-40 times. None of the participants in our sample demonstrated sequential recall behavior, though we observed the behavior in viewing videos of multiple participants who took the test without fiducial markers.

Once we finalized the behaviors, we categorized them into relevant types. In this chapter, we define our selected behaviors in greater detail, and explore the implications of their detection in the data. We aim to identify both variability between participants and differences between trials on aggregated measures.

5.1 Visual Match Behavior

We start by defining relevant terms. The *stimulus cell* is the cell for which the participant is trying to associate the correct digit with the symbol. We defined a *visual match* as an instance in which the participant focuses on another cell that has the same symbol as the stimulus cell. Match behaviors can be found in the translation and copy portions of the test, as the participant attempts to find the answer to fill in the current cell. The presence of match behavior is most meaningful in the translation task, since it is not necessary to look at other cells to copy in the copy task.

We initially expected that healthy participants would need to reference only the key to find the matching symbol and associated digit for the stimulus cell. However, we soon found that some participants began to search for matches in previously completed cells, either in the same or previous row(s). These strategies suggest that components of working memory (spatial and temporal factors) are considered in the decision to look back at previously translated cells. More specifically, the recognition that the symbols were being repeated, and the decision that it could require less visual displacement and time to match a symbol with a former cell rather than the key.

We defined four kinds of match behavior: *key-match* - looking at the key, *cell-match* - finding the stimulus in a previously completed test cell, *key-and-cell-match* - both, and *no-match-needed* - did not need to look up at the answer at all.

5.1.1 No Match Required

The no-match-required classification is particularly interesting and may have at least two possible implications. First, it may be construed as a metric of learning the symbol-digit association. Second, it may also serve as a sign of working memory, if the participant was able to remember the associated digits for the current symbol from a prior match.

We assumed that attention is required to visually process matches, and used fixation data to search for that behavior. For the translation task, we hypothesized that healthy participants would have a higher percentage of cells with no-match-

required in the second trial than the first, as participants would have longer exposure to the symbol-digit associations. Healthy subjects would display minimal match behavior in both trials of the copy task.

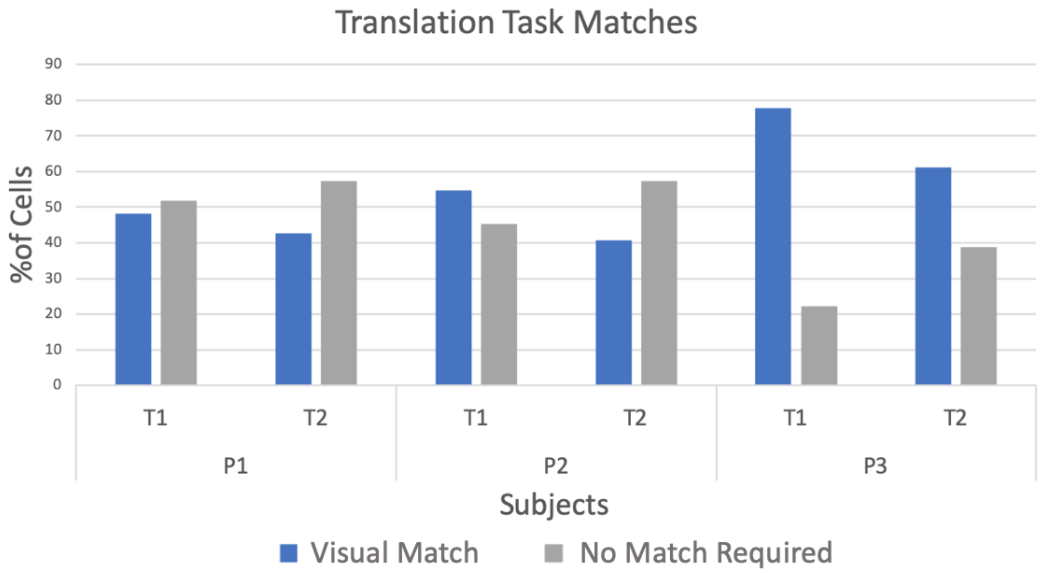

Figure 5-1: Matches vs. no matches in the translation task

Figure 5-1 shows the percentages of cells with match behavior found (key-match, cell-both, or key-cell-match) and no-match-required from participants across both trials. As hypothesized, it appears that the percentage of no-match-required increased in the second trial for all participants. To further explore this, we aggregated the data for all participants and conducted a chi-squared test to determine if there was a significant difference in match behavior between trials, as shown in Table 5.4[1].

	Match	No Match	Total
Trial 1	97	64	161
Trial 2	78	83	161
Total	175	147	322

Table 5.4: Contingency table of aggregated cells in the translation task

[1]Although there are 54 cells in the translation task, one of the participants had a test with a misprinted symbol and only filled out 53 cells for each of the two trials. Other participants filled out all 54 cells.

A chi-squared test of the data in Table 5.4 produced a p-value of 0.04, indicating that the difference in match behavior in the translation task between trials is significant.

As hypothesized, we also found that there were minimal match behaviors (shown in fewer than 5 cells out of 48) in the copy section of the test for the participants across both trials.

5.1.2 Concentrated Matches

We defined a concentrated match as one in which the first fixation in the key, or previously completed cells, while advancing to the next cell, was a match. We expected that the percentage of concentrated matches in the second trial would be higher than the first trial, reflecting that participants started to learn the spatial positions of the symbols in the key and previous cells over time.

We found that the percentages of cells with concentrated matches ranged between 13%-25%. Contrary to our hypothesis, we found that the percentages decreased from the first trial to the second trial for all participants during the translation task. This may have been associated with the general decrease in match behavior from Trial 1 to Trial 2 that we previously discussed.

We hypothesized that even if participants did not display concentrated match behavior, the average time to find the match should decrease between trials for healthy participants.

Participant	Avg. Match Time (s): Trial 1	Avg. Match Time (s): Trial 2
1	0.13	0.09
2	0.18	0.17
3	0.21	0.14

Table 5.5: Average match time per trial in the translation task

The data varies by participant, as shown in Table 5.5, but the trend between trials is clear. The results validate the hypothesis, as the average match time in cells within the translation task decreased in the second trial for all the participants. This finding

matches the intuition that healthy participants are able to recall the spatial locations of symbols better over time, leading to quicker matches in the second trial.

5.2 Forward and Backward Fixations

There are several reasons why participants might focus on cells other than the one they are currently working on. They may choose to look back in completed cells to match symbols, or simply view how much of the test they have completed. Under the assumption that the cells in the tasks are filled out in order, participants are less likely to look ahead for the purpose of matches, but may be trying to prepare for oncoming to fill out. The maximum number of cells looked ahead has the potential to reveal the limits of working memory for the participants.

Backward fixations are conducted in cells in the same row and to the left of the current cell, while forward fixations are to the right in the same row. These definitions allow our analysis to focus on the number of cells that can be preserved in memory. We hypothesized that forward fixations would rise from the first trial to the second for healthy subjects on the translation task, due to increased confidence in remembering symbol-digit associations over the progression of the test.

Figure 5-2 shows the average of the maximum number of cells gazed ahead across participants, plotted per cell for both trials of the translation and copy tasks. It appears that the same general trends are followed in the tests for both trials: there are sharp delayed drops at the end of the row (Cells 14, 28, and 42) or task (Cell 48 for the copy task and Cell 54 for the translation task), and sharp rises in the beginning of each row (Cells 1, 15, 29, and 43). This is expected, as participants cannot look further ahead at the end of the row, and may be interested in looking ahead again at the start of the next row after taking a short mental break.

There was also variability within participants, as expected, since different strategies were followed. One participant gazed ahead at the beginning of each row of the translation task, but otherwise did not look ahead. Another consistently looked at most 1-2 cells ahead. Yet another participant looked several cells ahead, then did not

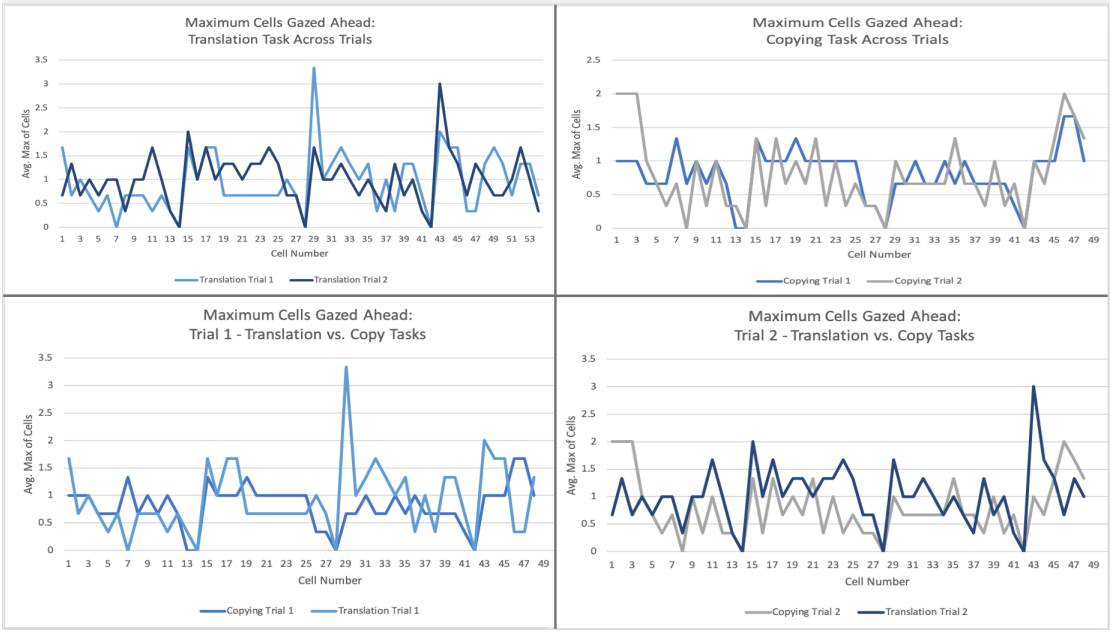

Figure 5-2: Average of maximum cells gazed ahead across trials in the translation and copy tasks. Note: Cells 49-54 were dropped when comparing the translation and copy tasks since there was no copy task counterpart.

look ahead at all for a few cells, and continued the pattern.

Manual analysis of the gaze recordings for other participants who took the tests without fiducial markers revealed another strategy in the copy task. Several participants consistently looked several cells ahead during the first trial, and then did not look ahead at all in the second trial. This may be attributed to the realization that there was no additional benefit in looking forward, because less mental preparation is needed to copy rather than translate a cell.

We computed the correlation coefficients and associated p-values across pairs of trial and task data of the averaged maximum number of cells looked ahead. The results are shown in Table 5.6.

We found similar trends between translation tasks for Trials 1 and 2. This suggests that strategies used for the first trial were similar for those of the second trial. This finding may not hold true in larger samples due to individual variation, but holds promise for potential to detect subtle change in individuals who may not be cognitively healthy. There was high correlation between copy tasks for Trials 1 and 2, also

Comparison	Correlation Coefficient	Relationship Strength	p-value	Result ($\alpha = 0.05$)
Translation Task, Trial 1 vs. Translation Task, Trial 2	0.47	Moderate	$3 * 10^{-4}$	Highly Significant
Copy Task, Trial 1 vs. Copy Task, Trial 2	0.64	Moderate	$1.13 * 10^{-6}$	Highly Significant
Translation Task, Trial 1 vs. Copy Task, Trial 1	0.16	Weak	0.26	Not Significant
Translation Task, Trial 2 vs. Copy Task, Trial 2	0.31	Weak	0.03	Significant

Table 5.6: The correlation coefficients and associated p-values for average maximum cells looked ahead, compared between tasks and trials.

suggesting that healthy individuals are consistent in their approach to complete the task. However, this increased correlation when compared to the translation task suggests that there may be signals within the translation task that merits further investigation and potentially could reflect subtle individual variations in cognitive strategies that are not apparent in our small sample.

Even though the motor demands are identical for the copy and translation tasks, the two tasks were found to be only weakly correlated, demonstrating the difference in the cognitive demand for these seemingly identical motor tasks. The data for the first trial between the two tasks was not statistically significant. Considering the data from the second trial, even though the relationship was weak, the strength of the correlation was higher in relation to the first trial. This suggests that there could be signals indicating how the change in the understanding of the task by experience could inform the behaviors demonstrated during the test.

5.3 Leftward and Rightward Scans

We then considered leftward and rightward *scans* during the translation task. *Scans* are classified as consecutive glances across three or more cells in a single direction. We defined two classes of scans: *key scans* and *same-row scans*. These can occur in either the leftward or rightward direction. Leftward same-row scans may arise from

visual searching to find a previous instance of the symbol to match, while rightward same-row scans are likely to appear while mentally preparing for the cells ahead. Key scans are expected to show up more frequently than same-row scans in the translation task, as participants attempt to perform quick scans to form spatial associations for symbol positions in the key.

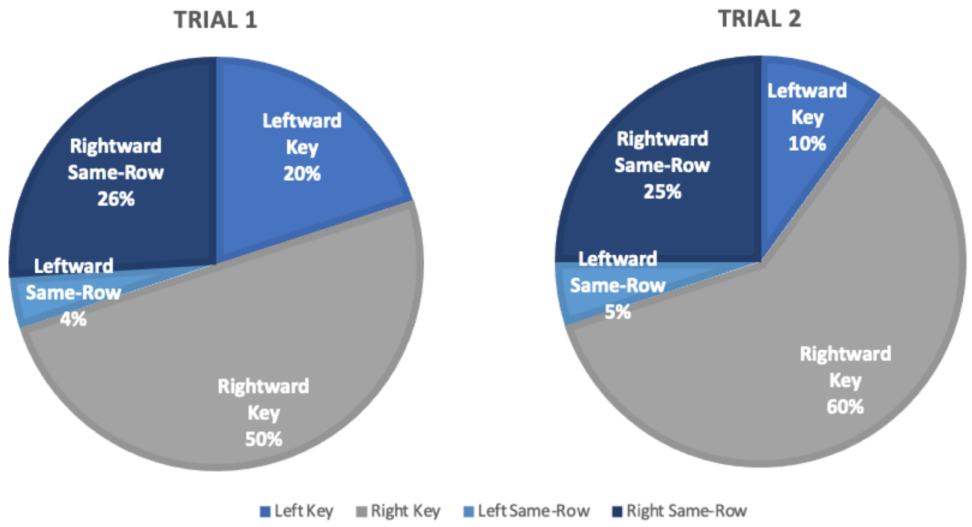

Figure 5-3: A comparison of the aggregated percentages of the different types of scans during the translation task between trials

The data across the participants was aggregated and visualized in Figure 5-3. It shows that key scans are more common than same-row scans, as expected. The relative percentages between same-row scans and key scans are the same across trials.

Within same-row scans, the percentages of leftward and rightward scans are also consistent across trials. In general, it appears that rightward same-row scans are more common, since perhaps participants like to scan in the same-row to prepare for upcoming cells. Leftward same-row scans occur far less frequently. This may be because participants conduct few symbol matches with previous cells in the same row, or know the general position of the past symbol and look to it directly rather than scanning. Furthermore, the percentages support the notion that participants performing the test look ahead more frequently than they look backward. We hypothesize that this behavior may be attributed to determine how much of the test remains to be

finished rather than backward to focus on how much they have completed, which can be explored in a future study.

Within key scans, the rightward scans are more common across both trials, but there is an increase in the percentage in the second trial. Many participants gaze up primarily to the middle of the key to match a symbol, and then scan rightward or directly jump to another symbol leftward, which may explain this percentage. It remains to be explored in further research why participants demonstrate this behavior, and whether there are any spatial-processing implications.

5.4 Back-and-Forth Fixations

Back-and-forth fixations are successive fixations involving looking at one cell, then another cell, and then back to the first cell. The categories of back-and-forth fixations include: *key-key* (key-key-key), *cell-cell* (cell-cell-cell), *key-cell* (key-cell-key or cell-key-cell), and confirmations. Confirmations are a special case of the back-and-forth behavior of key-cell, in which the cell is the current cell in which the subject is working.

The presence of back-and-forth fixations can be indicative of associations made or forming between symbols. Participants may look back and forth between similar symbols in order to confirm separating features. Alternatively, participants may also look back and forth between cells to confirm the spatial locations of symbols with respect to each other. Thus, back-and-forth behavior can perhaps suggest a way of further investigating how the symbols involved are processed.

We hypothesized that participants would show more back-and-forth behavior in the first trial, as they formed symbol-digit association for the first time. Additionally, we hypothesized that confirmation counts would be low in healthy participants with good working memory.

As shown in Figure 5-4, there is considerable variability between participants in the first trial, though the data is relatively similar in the second trial. There is indeed more back-and-forth behavior in the first trial for most participants, as hypothesized. Key-key and cell-cell gazes are most common, with cell-cell appearing

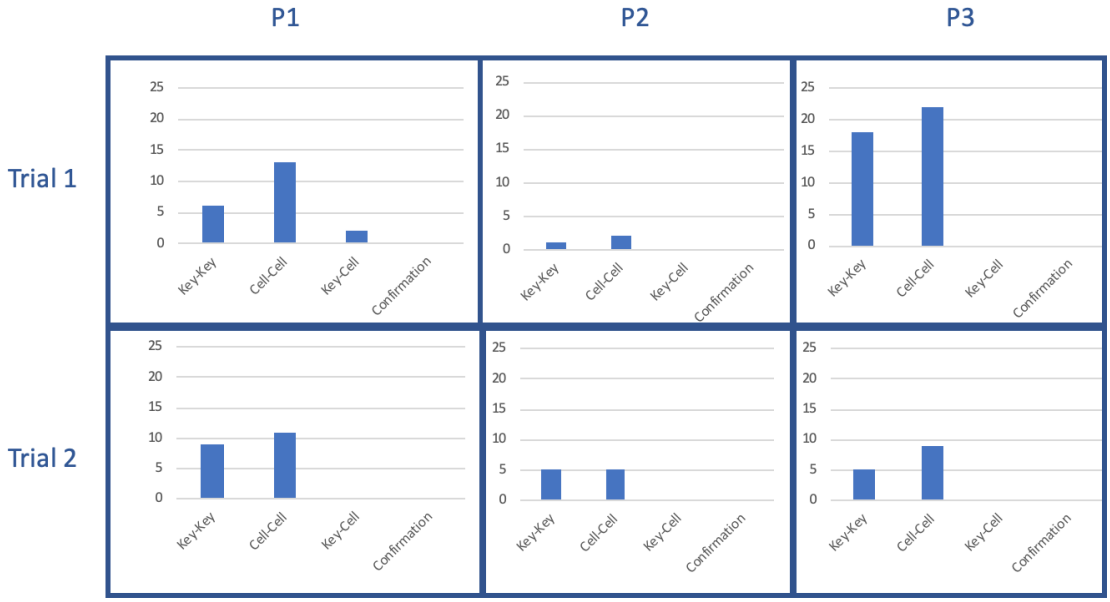

Figure 5-4: A comparison of the counts of back-and-forth fixations across participants and trials during the translation task

most frequently. Key-key gazes likely occur when the participant references the key to match a symbol, while cell-cell behavior may occur during both past-cell symbol matches and during mental preparations for cells ahead. Key-cell fixations show up rarely and confirmations do not appear at all in our participant data.

Within the key-key gazes, the most common pairings are between shapes that are right next to each other: Keys 1-2, 4-5, and 5-6. It is particularly interesting that another frequent pairing is between Keys 3-6, which are several cells apart. However, the symbols in those cells are an X and four-sided star, which appear similar when rotated, as shown in Figure 5-5. The back-and-forth between these symbols suggest a visual comparison process and could indicate intention to focus on the differences of these similar shapes for future recognition.

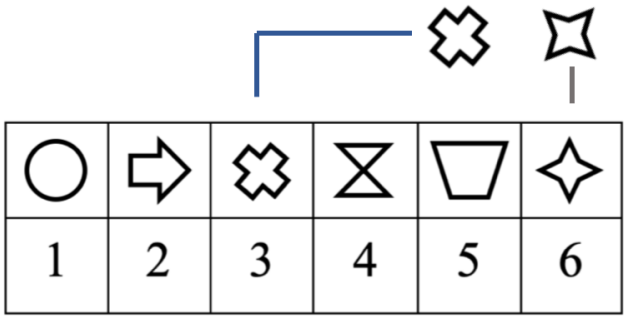

Figure 5-5: Side-by-side comparison of a common key pairing found in back-forth fixations: X and four-sided star

5.5 Spatial Associations

Other patterns to be analyzed within subjects relate to spatial perception. We separated the test form into: *start*, *middle*, and *end* sections (see Figure 4-2). It is interesting to note where participants first fixate after finishing a cell. If the first fixation is in the same spatial section as the current cell, we define this occurrence as a *spatial association*. We searched for spatial associations within the fixation data in the translation task, under the hypothesis that the percentage of spatial associations across cells for healthy participants should be high, particularly early on in the first trial before symbol positions in the key are learned.

As shown in Figure 5-6, there is considerable variability in the translation task between participants on the percentages of cells with spatial associations. Furthermore, the individual distributions of spatial associations also vary by rows and spatial sections of the page, both across participants and trials. There may be a confounding factor present in the symbols that are associated with each cell, as perhaps participants are able to recall the positions of symbols within the key or prior cells more quickly for some symbols, which affects where they first fixate. Furthermore, when combined with the look-ahead behavior, it may be possible that participants are searching for the symbol of another cell, which affects whether or not a spatial association is made.

Thus, we could not come to any significant conclusions from the data on spatial associations in the translation task. However, upon comparison with a larger sample

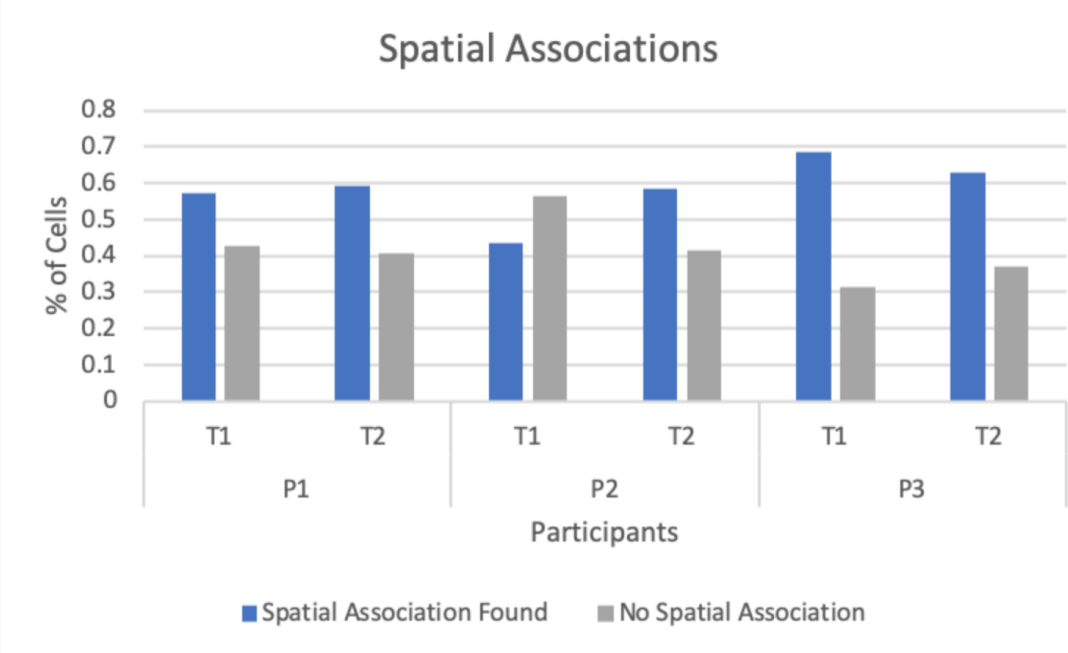

Figure 5-6: A comparison of percentage of spatial associations made in the translation task across participants and trials.

size, as well as another group of impaired participants, it may nonetheless be interesting to consider the differences in the presence of spatial associations, particularly upon keeping the discussed confounding factors in mind.

5.6 External Fixations

External fixations are classified as fixations in areas that are not relevant to the current task. This behavior is particularly analyzed in the copy and delayed recall tasks. The key and the delayed recall section are not necessary to copy digits, and are therefore considered external sections for the copy task. Similarly, the external sections for the recall tasks are the key and the copy section. Blank areas in the form are also considered external because they do not contain information required to complete the tasks. Visual attention in external areas during the copy task is not expected to be frequent among healthy participants, because of the irrelevance to the task, but external fixations during the delayed recall task can signal cognitive processes associated with recall.

As expected, most of the percentages of external fixations for the copy task were low (<15%). There were mainly fixations in blank areas, perhaps serving as a visual resting spot for participants as they completed the copy task.

Participant	Trial	% of External Digit Matches	% of External Blank Fixations	% of All External Fixations
1	1	0.00	2.15	3.86
1	2	0.00	9.65	14.04
2	1	25.85	0.68	31.29
2	2	4.71	0.00	19.37
3	1	5.95	51.38	84.33
3	2	5.14	0.00	23.36

Table 5.7: A comparison of the percentages of the types and total amount of external fixations across participants and trials in the delayed recall task

Table 5.7 covers the different types and total percentage of external fixations out of all fixations for three participants across two trials for the delayed recall section. We defined *external digit matches* as fixations in which the participant glanced at a digit in the key or copy section that matched the digit that should have been filled in for the stimulus cell in the recall section. *External blank fixations* are fixations in blank areas on the test surface.

We found the percentage of total external fixations to be variable, both between participants and trials. We initially assumed that participants would fixate on blank areas, without visual distractions, to concentrate and recall. However, we found the percentage of blank fixations to be quite low. It is possible that participants may have glanced too quickly at blank areas, fixated in areas outside of the test form, or simply followed other recall strategies that did not align with our assumptions.

Relatively few of the fixations were classified as external digit matches, particularly in the second trial. Nonetheless, the presence of high percentages of these types of fixations should be further investigated within individuals, to further shed light on the recall process of symbol-digit associations.

5.7 Recall Metrics

We integrated our programs with a symbol-digit recognizer developed in our lab that automates determining whether a hand-written delayed recall answer is correct. We observed that none of our participants completed the delayed recall section in sequential order, instead choosing to fill in cells seemingly based on personal ease of recall, since the first few cells were correctly completed more quickly than the later cells. The time taken before writing in each cell in the delayed recall section can be associated with patterns of recall shown by participants. We defined this measure as the pre-cell delay: the time between the end of the stroke in the previous cell and the start of the first stroke in the current cell. As such, the first cell that was filled in was not counted in the data. Our expectation was that the pre-cell delay would decrease, on average, from Trial 1 to Trial 2 for each participant.

However, we found that the pre-cell delay varied greatly between participants and across trials. Long delays, over 3 seconds, could be attributed not only to difficulty of recall of the cell to be filled out, but rather due to participants looking at different cells and determining which one to translate next. Since the task did not have to be carried out in sequential order, the possible presence of this thought process would be a misleading addition to our pre-cell delay metric.

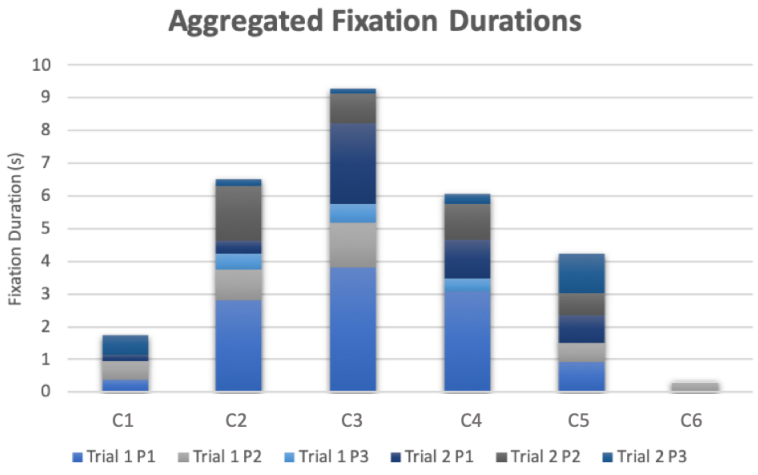

Figure 5-7: Aggregated fixation durations across participants and trials for each of the six cells in the delayed recall task

As a result, we expanded our analysis to include which cells were fixated on most frequently during the delayed recall task. We calculated the total duration of each fixation on any cell in the delayed recall section throughout the task. The aggregated fixation duration for all participants across the trials are shown for each cell (1-6) of the recall section in Figure 5-7. It is not surprising that the cell with the lowest fixation duration was associated with the circle symbol, potentially due to the common shape that may be remembered more easily than the other symbols. This finding can be further explored in a future study to determine whether some shapes are inherently more difficult to recall than others.

The maximum aggregated fixation duration value for a recall cell was slightly under 4 seconds. It remains to be seen whether impaired participants will fixate for longer periods of time, which may differentiate between healthy and unhealthy levels of recall. It will also be interesting to analyze the correlation between fixation duration and recall correctness with larger groups, particularly composed of both healthy and impaired participants, in future work.

Chapter 6

Conclusion

6.1 Contributions

In this work, we identified fixations from gaze data of participants who took the digital Symbol Digit Test. We combined the fixation measurements with ink data to create an aggregated dataset to analyze for each participant. The layers of abstraction we built are detailed in Table 6.1. We then analyzed video recordings of test sessions and identified behaviors and metrics that could be related to cognitive strategies. We implemented the computational detection of these behaviors and metrics, which are listed in Table 6.2, and analyzed the preliminary results on the data of our subset of participants.

We searched for similarities and differences, both between participants and across trials. We interpreted potential indicators of cognitive health, such as good working memory, from our findings, and proposed areas for further research. Our analysis was conducted on a small sample size, so our insights are preliminary but have high value for guiding future work.

Once more data is collected, statistical tests can be conducted to determine significance of patterns, particularly to determine effectiveness in predicting cognitive state. We further recommend that combinations of behaviors be analyzed, to better investigate the complexity of cognitive strategies.

	Data Source	Notes
Raw Inputs	Gaze Data	Generated using Pupil software
	Ink Data	Collected from THink software using the Anoto Pen
Transformed Inputs	Fixation Data	Generated using the Pupil Fixation Detector
Processed Inputs	Merged Ink and Fixation Data	- Data sources were aggregated, interpolated and merged - Available ink and fixation (x, y) locations were mapped to test rows and columns (e.g., Cell 5: Row 1, Col 5) - Participant data was separated by task and trial
Data Contextualization	Base Observations	- Ink and fixation locations within cell boundaries were mapped to the cell symbol - General observations - (e.g., glance in key, position of fixated cell with respect to current cell being filled out)
Data Analysis	Pattern/Behavior	- Behavior 1: Match Behavior - Behavior 2: Scan Behavior - (Additional Behaviors...)

Table 6.1: Summary of the layers of abstraction built to analyze patterns and behaviors from raw participant data

Metrics
Cell Completion Duration
Pre-Cell Writing Delay
Average Fixation Duration
Average Ink Speed
Gaze/Fixation Distance

Behaviors
Visual Match
Visual Match (Concentrated)
Forward and Backward Fixation
Scans
Back-and-Forth
Spatial Association
External Fixation
Sequential Recall

Table 6.2: List of the behaviors and metrics for which detection was automated.

6.2 Future Work

We suggest that it would be useful to emphasize improving eye-tracking calibration. Fiducial markers may be visually distracting to participants, so detection of the test form through computer vision techniques rather than the markers would be an improvement to the overall testing experience. We also focused on fixation data in this work, and suggest that the entire gaze dataset (e.g., including saccades) be analyzed to find more subtle patterns.

The tools and software developed in this work are intended to serve as a foundation for future research into identifying differences in cognitive strategies between participants, as well as similarities in healthy individuals that can serve as separating features from impaired participants. We hope our research will advance the goals of improving our understanding of human cognition and developing accurate cognitive tests sensitive to the early detection of neurological diseases.

Appendix A

Supplementary Information

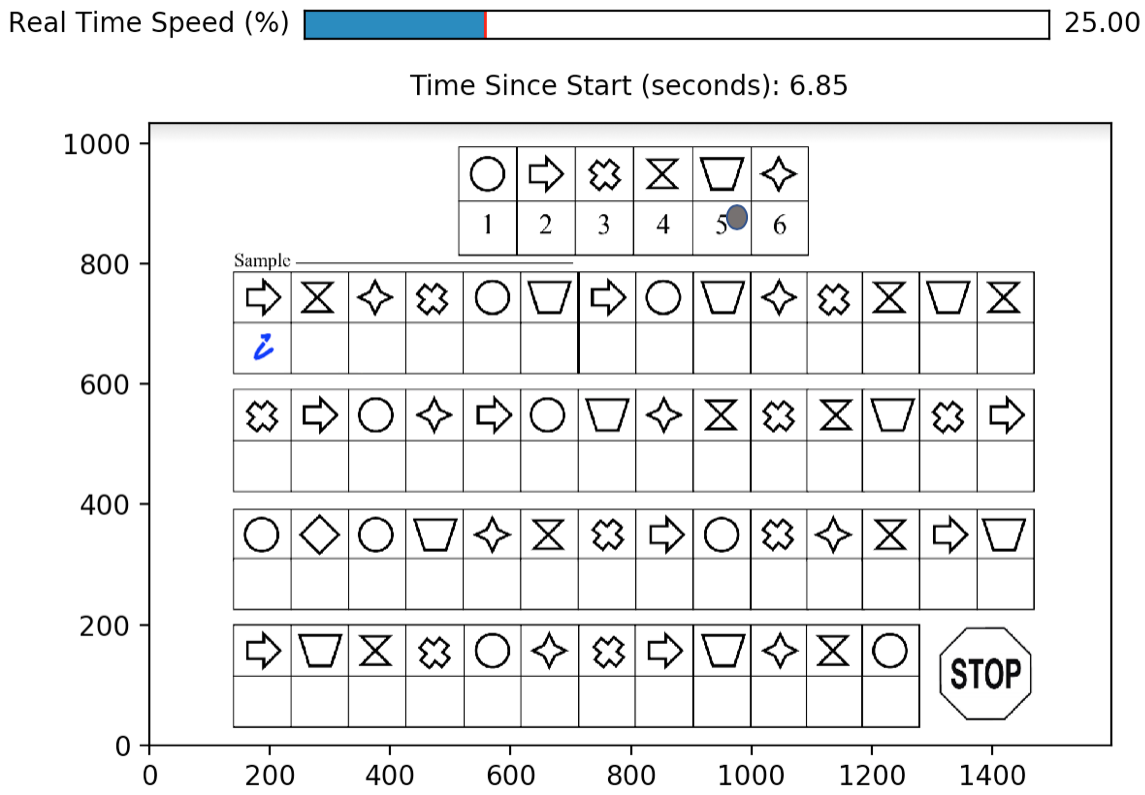

Figure A-1: This is a still from the visualization tool, depicting the playback speed percentage, time since start of the video, the strokes made by the Anoto pen in blue, and the current gaze (or fixation) location as the gray dot

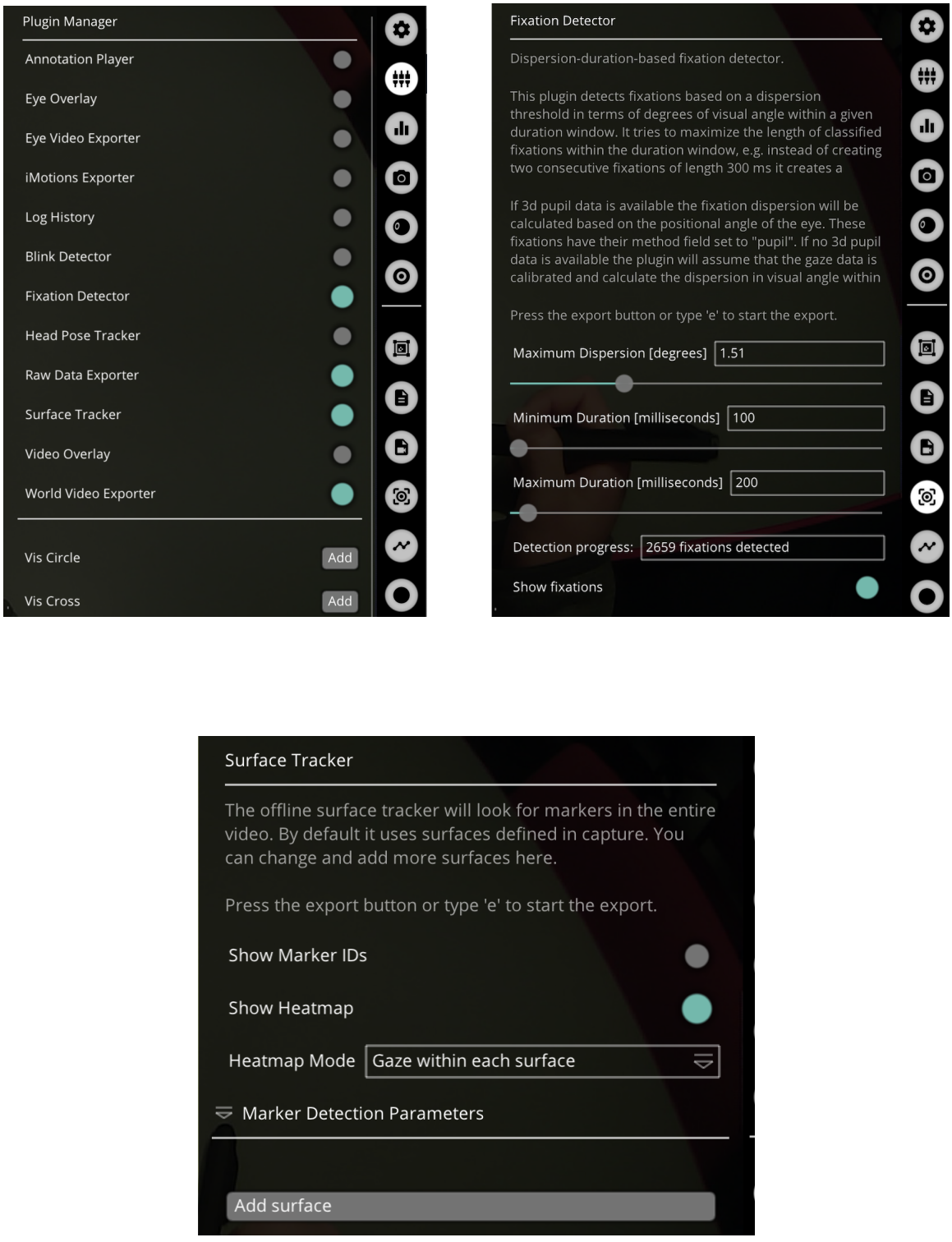

Figure A-2: The Fixation Detector and Surface Tracker are plugins accessible from the Pupil Player

Bibliography

[1] Neurodegenerative diseases, 2019. Retrieved online from https://www.niehs.nih.gov/research/supported/health/neurodegenerative/index.cfm.

[2] Akane Oyama, Shuko Takeda, Yuki Itoaand Tsuneo Nakajima, Yoichi Takami, Yasushi Takeya, Koichi Yamamoto, Ken Sugimoto, Hideo Shimizu, Munehisa Shimamura, Taiichi Katayama, Hiromi Rakugi, and Ryuichi Morishita. Novel method for rapid assessment of cognitive impairment using high- performance eye-tracking technology. *Nature*, 2019.

[3] Cognitive testing, 2020. Retrieved online from https://medlineplus.gov/lab-tests/cognitive-testing/.

[4] YanHong Dong, Melissa Jane Slavin, Bernard Poon-Lap Chan, Narayanaswamy Venketasubramanian, Vijay Kumar Sharma, Simon Lowes Collinson, Perminder Singh Sachdev, and Christopher Li-Hsian Chen. Improving screening for vascular cognitive impairment at three to six months after mild ischemic stroke and transient ischemic attack. *International Psychogeriatrics*, 2012.

[5] Ralph HB Benedict, John DeLuca, Glenn Phillips, Nicholas LaRocca, Lynn D Hudson, Richard Rudick, , and Multiple Sclerosis Outcome Assessments Consortium. Validity of the symbol digit modalities test as a cognition performance outcome measure for multiple sclerosis. *Sage*, 2017.

[6] Types of eye movement, 2020. Retrieved online from https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/types-of-eye-movements/.

[7] Maddie Pascoe, Yassar Alamri, John Dalrymple-Alford, Tim Anderson, and Michael MacAskill. The symbol-digit modalities test in mild cognitive impairment: Evidence from parkinson's disease patients. *European Neurology*, 2018.

[8] Lauren Huang. The digital symbol digit test: Screening for alzheimer's and parkinson's, 2017. MIT MEng Thesis.

[9] Jing Xian Wang. Tracking eye behavior in decision-making maze problems, 2019. MIT MEng Thesis.

[10] Stephen Li. The digital symbol digit test: Screening for alzheimer's and parkinson's, 2019. MIT MEng Thesis.