Eye Tracking for Cognition

by

Stephen Li

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

Masters of Science in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Author . . . .
Department of Electrical Engineering and Computer Science
May 24, 2019

Certified by . . . .
Randall Davis
Professor
Thesis Supervisor

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

Eye Tracking for Cognition

by

Stephen Li

Submitted to the Department of Electrical Engineering and Computer Science
on May 24, 2019, in partial fulfillment of the
requirements for the degree of
Masters of Science in Computer Science and Engineering

Abstract

Early detection of neurodegenerative diseases can lead to slower disease progression, as well as possible symptom reduction. Existing research has studied how cognitively impaired subjects solve tests such as the clock-drawing test and the Digital Symbol-Digit Test differently compared to healthy subjects. While subjects in previous work used a digitized pen in solving the Digital Symbol-Digit test, our research focuses on having the subjects wear eye-tracking glasses as well. These glasses bring a significant improvement in mobility over computer-mounted or headframe eye trackers, but also may come with its reliability issues. After these issues are solved, the gaze data provided brings a wealth of information on learning rate and clues to what the subject is thinking.

Thesis Supervisor: Randall Davis
Title: Professor

Acknowledgments

I would like to thank Professor Randall Davis for his constant words of wise guidance and clear insight into my research project. I would also like to acknowledge Dr. Dana Penney's invaluable contributions via interesting observations and questions that only an experienced doctor could provide.

Assisting me in my project as well was Jing Wang, who served as a test subject at times and helped me run subject experiments. Steve Giles helped us set up the system of software that we used and answered any questions that I had.

Contents

List of Figures

Chapter 1

Introduction

As life expectancy across the world increases and populations gradually age, neurodegenerative diseases such as Alzheimer's Disease and Parkinson's Disease will begin to affect more and more of the world population. With these diseases having such a large impact on the quality of life of individuals as well as placing a large burden on caretakers and the health care system, mitigating the severity and prevalence of these diseases is crucial.

There are numerous benefits from being able to detect disorders such as Alzheimer's early [3]. Thus, increasing the use and effectiveness of early diagnostic tests should be emphasized. One such class of diagnostic tests are cognitive tests. Cognitive tests check for problems involving cognition, and often involve asking the patient to perform simple tasks. Patients may be asked to draw a clock, copy sentences, or take a multiple choice exam.

Existing tests are easy and simple to administer but may suffer from having only qualitative and sometimes lengthy grading schemes [8]. Automating test evaluation can save the grader precious time, as well as create consistency in the testing and evaluation process.

In this work, we will be evaluating a new testing procedure for the Digital Symbol-Digit Test (abbreviated as dSDT). The dSDT, a relative of the Symbol Digit Modalities Test (SDMT), is designed to identify neurodegenerative diseases such as Alzheimer's and Parkinson's.

Subjects taking the test were equipped with an off-the-shelf digitizing ballpoint pen and eye tracking headset. While testing 28 healthy subjects, we encountered eye-tracking issues with a majority of the participants, mostly regarding reliability in pupil detection. However, for the subjects that did not encounter such issues, we were able to collect high-frequency and high-accuracy gaze data. The combination of eye and pen tracking data gave us a multitude of metrics to measure real-time learning as well as cognitive performance on attention and scanning speed.

The reliability issues are likely solvable via a combination of updated testing procedures and slightly modified headset. We expect that, in the future, this analysis can be extended to a large majority of subjects and thereby become an accurate and robust process in evaluating cognitive health.

Chapter 2

Background

2.1 Digital Clock Drawing Test

For many years, clinicians have been giving variations of the Clock Drawing test as a screening test to detect neurological disorders [1]. In one variation, the test occurs in two phases. In the first "command" phase, the subject draws a clock showing 10 minutes after 11 from a blank sheet of paper. In the second "copy" phase, the subject is asked to copy a clock showing 10 minutes after 11. Performing the test requires verbal understanding, memory skills, and spatial reconstructive skills [4].

In prior work performed by William Souillard-Mandar, Professor Randall Davis, and others, the clock drawing test has been moved to digitized form, now referred to as the digital clock drawing test. The test is still conducted with paper/pen, but the use of the Anoto digital pen allows for automated analysis of the pen strokes, as well as the development of human-interpretable machine learning classifiers.

2.2 Symbol Digit Modalities Test

The inspiration for the dSDT comes from the Symbol Digit Modalities Test (SDMT). The SDMT first presents the subject with 9 different symbols, each associated with the numbers 1-9. The subject is then presented with an assortment of the symbols, and empty cells below the symbols. The subject is asked to fill in as many corresponding numbers below the symbols as possible within 90 seconds, and the number of correct responses is then tallied up. The test is an economical way to screen for cognitive impairments since administration usually takes less than 5 minutes.

The SDMT's aim is to assess key neurocognitive functions such as attention, visual scanning, and motor speed [7]. Existing research has demonstrated the SDMT's reliability and validity in detecting multiple sclerosis [9].

2.3 Digital Symbol Digit Test

The Digital Symbol Digit Test can be seen as a digital analogue to the SDMT. However, substantial changes were made to the design in order to capture additional information about learning and memory recall that the SDMT does not [4]. Developed by Dr. Dana Penney and Professor Randall Davis, the dSDT consists of three major tasks. Instructions given to the participants are included in Appendix B.

The first task involves translation (similar to the SDMT). A key on top of the page presents associations between six symbols and their corresponding digits. The subject is expected to use the key as a guide to fill in the empty cells below a series of the same symbols.

The second task involves simple copying. The key for the copy task is straightforward - the symbol and associated digit are one and the same. Therefore, the subject is simply made to copy the number from above into the empty cell.

The last task involves delayed recall. The subject is presented with the six symbols from the previous task in permuted order and is asked to translate them into numbers. Since the key from the translation task is no longer visible, the patients must rely on their memory to complete the section.

After all three tasks are finished, the subject is presented (without warning) another identical, blank copy of the dSDT. After the subject completes the three tasks again, the dSDT examination is finished. The total duration of both tests ends up being 10 minutes or less for most patients.

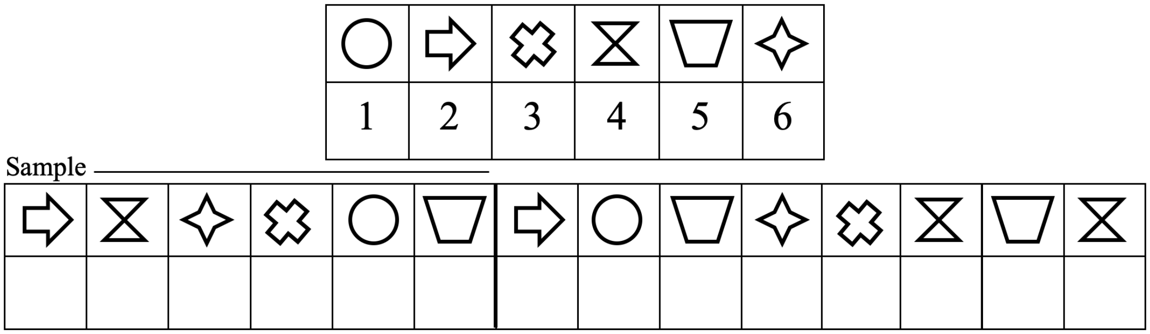
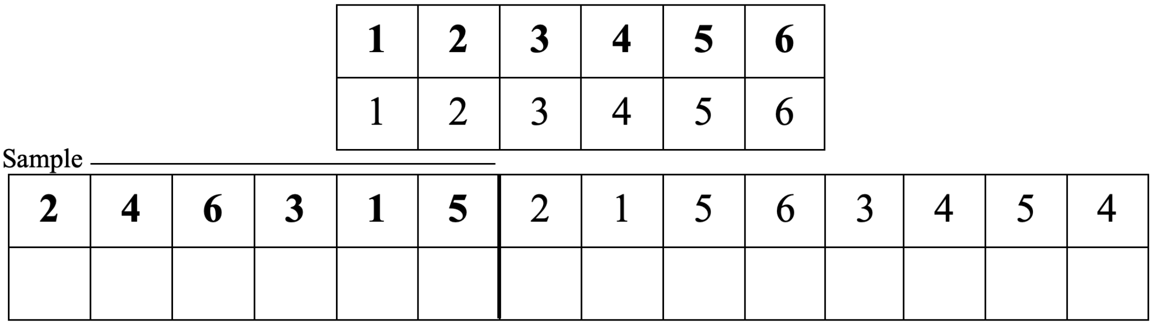


Figure 2-1: A sample of the translation task of the dSDT, with the key and first row displayed.



1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6

Sample

2	**4**	**6**	**3**	**1**	**5**	2	1	5	6	3	4	5	4

Figure 2-2: A sample of the copy task of the dSDT, with the key and first row displayed.

Each task checks different aspects of a subject's cognitive function. The translation task, similar in nature to the SDMT, tests for the same cognitive functions - attention, visual scanning, and motor speed. The copy task provides a control for the subject's writing speed, allowing us to disentangle writing speed as a confounding factor during our analysis. The recall task explicitly checks a subject's learning speed and memory. The quicker a subject learns, the more symbol-digit pairings he is likely to recall correctly.

Having a subject complete the identical test a second time allows us to explicitly measure a subject's learning speed. Checking timing results across the two tests and the number of correct responses in the delayed recall section gives us a wealth of information otherwise unattainable just by taking an exam once.

2.4 Existing Work on the Digital Symbol Digit Test

The dSDT, in prior work, has been run and analyzed on subjects using the Anoto pen alone [4]. ML classifiers have been produced from the pen data, which prove to be a promising tool in screening for and identifying patients with Alzheimer's, mild cognitive impairment, and Parkinson's Disease.

The addition of eye-tracking data to our classifiers adds a great deal of information about a subject's current activity and learning patterns. Since the dSDT key is in such a distinct location (the top of the page), eye-tracking software can immediately note the frequency and duration of gazes at the key. Individual glances at other items on the page (e.g. the current symbol the subject is working on) can be recorded as well. Most of these behaviors can only be guessed at or inferred with just pen data alone, but the additional eye data makes these behaviors explicit and analyzable.

2.5 Relevant Eye Tracking Studies

An existing study had participants take the SDMT, using paper and pencil, while wearing an eye-tracking headset [6]. The study recruited Parkinson's patients with or without mild cognitive impairment, as well as normal healthy participants. The experiment did not find a correlation between the SDMT score and proportion of fixations on the key area to working area, nor between the SDMT score and fixation duration on the key area or working area.

This paper presents a good starting point in analyzing eye-tracking data by looking at key and working area gazes. However, since the authors did not incorporate pen and eye data together - which would allow segmentation of data cell-by-cell - their range of analysis is limited.

Another study recruited subjects with schizophrenia to take the SDMT on a computer [2]. The monitor was equipped with an eye tracking device, and subjects used a chin rest to maintain a fixed distance from the monitor. Participants' answers were verbalized. The study concluded that schizophrenic patients averaged 22% more visits to the key area, as well as each visit being 134ms longer on average. In addition, there was a significant negative correlation between the number of key visits per cell and SDMT score. The authors suggested that the lower performance by schizophrenic patients can be attributed to less efficient visual search, as well as poorer memory retention.

This paper goes more in-depth with eye tracking analysis and breaking down gaze data cell-by-cell. The more in-depth analysis may in part explain the inconclusiveness of the gaze data found in the first study. However, since the SDMT was conducted on a computer and subjects were made to answer verbally, the paper misses out on picking up a subject's writing patterns, which already prove to be a promising technique in screening for impaired patients [4]. Our paper expands on the work by [4] and [2] by fully incorporating pen and gaze data in a more natural setting.

Chapter 3

Equipment

3.1 Physical Equipment

From the perspective of the subject, the Anoto pen functions just as an ordinary ballpoint pen.

A picture of the Pupil Labs headset is shown in Figure 3-1. The three functional components of the headset that are relevant to eye tracking are the world view camera (sharing the same perspective as the subject), as well as two eye cameras (each camera pointed towards one eye).

All three cameras are adjustable. The world view camera can be adjusted up or down. The eye cameras rest on a ball joint and have a great deal of freedom moving around the joint. In addition, two sets of sliding adjusters for each of the eye cameras can move the camera forward or backward (relative to the subject's face). The sliding adjusters end up modifying both the distance and the angle of the camera to the user's eyes.

The headset is wired and connected to a laptop computer operated by the examiner during the experiment.

For the eye tracker calibration process, we've attached a calibration marker to the end of a long, thin metal rod. The marker consists of a target with concentric rings of different radii, as well as a plus in the center where the user directs his gaze. The marker is moved around using the rod in the calibration process. Lastly, a light meter

Figure 3-1: The Pupil headset on display.

is also used to measure the room lighting during the experiment.

3.2 Relevant Software

Both capture and playback software (Pupil Capture and Pupil Player, respectively), are provided by Pupil Labs as well.

To make the testing process simpler for the experimenter, we've built our own custom version of Pupil Capture. However, the differences between the public version are simple - most of the software functionality not relevant to our experiment has been cut out. Pupil Capture has a number of useful functions while recording. It allows us to see the subject's perspective through the world camera, as well as the eye camera's view of the eye and pupils.

The most crucial function of the Pupil software is calibration. During the calibration process, the subject keeps his gaze at the center of the calibration marker at all times while the marker is moved around. The software keeps track of the marker locations (using the world camera), as well as the center of the pupil (using the eye cameras). In the end, the process produces a mapping between the pupil and gaze coordinates, allowing the subject's gaze to be displayed relative to the world camera at all times. The extent of the mapping is overlaid over the world camera in Pupil Capture.

The Pupil software can keep track of surfaces that appear in the world camera field of view, as long as the surface's edges are bounded by fiducial markers, one for each corner of the surface. If at least two fiducial markers are present in the camera's view, the software can make a reasonable guess regarding the boundaries of the surface. Here, the test forms are the surfaces that we want to keep track of.

Once both the surface and gaze positions are known, the Pupil software can produce the location of the subject's gaze on the surface.

Most of the experiments reported here were performed without fiducial markers on the edges of the test, so gaze data on the dSDT surface was not available for these experiments. Instead, a future version of the Pupil Player software we develop will use the test's features itself as a fiducial marker. This change was made to accommodate for issues when a subject would unwittingly mask two or more fiducial marks from the world camera (thus making the surface position unknown), and to lessen the extraneous cognitive load on a subject as well.

Chapter 4

Experimental Procedure

4.1 Setup

The subject takes the test while sitting at a desk. The examiner sits with a laptop computer in a separate chair on either side of the table, but ideally not the same side. The subject should sit in an adjustable-height chair to allow for optimal calibration conditions, with the elbows of the subject reaching approximately tabletop height.

4.2 Calibration Procedure

The experiment begins by putting the Pupil headset on the subject. The subject is given a blank piece of paper on the table, as well as the capped Anoto pen to hold. The subject is then asked to position his head and the piece of paper as if he were taking an exam.

Following this, all three cameras on the Pupil Camera are adjusted. The World Camera is tilted so that the piece of paper is roughly centered, or slightly below the center of the camera. The eye cameras are adjusted so that the entire eye lies within the vision of the camera. Next, the pupil detection algorithms (provided in the software) are adjusted to capture the pupil of each eye accurately.

Once all the cameras and algorithm parameters are modified, the calibration procedure can begin. During the calibration process, the participant is asked to keep his posture (including the head) in the exam-taking position and follow the calibration marker with only the eyes.

The calibration marker is moved using the long metal rod, and the subject is to keep his eyes on the center of the cross. The rod is moved slowly and carefully in all directions, with the goal of covering as much of the world camera view as possible.

4.3 Calibration Standards

The pupil detection algorithm has its own constantly running confidence measure of pupil location for each eye. The numbers range from 0.0 (no confidence in pupil location) to 1.0 (complete confidence). Once the calibration procedure finishes, only data points with a confidence of 0.8 or greater are used in the mapping process. The final calibration region - gaze positions through the world view camera that can be accurately viewed - is bounded in green.

If the pupil camera does not have a good view of the subject's pupils, the confidence measure frequently drops below the 0.8 threshold during calibration, and thus those data points are dropped. This issue ends up limiting gaze accuracy and gaze positions that can be detected by the headset.

Examples of high quality versus low quality calibrations are shown in Figure 4-1. Higher quality calibrations have large calibration regions as well as more defined interiors. Low quality calibrations have large amounts of dropped points, thus resulting in small regions and sparse interiors.

The results of the calibration are then manually checked afterwards by moving the calibration marker and verifying the reported gaze position matches the marker location. If the calibration region is too small, or the confidence threshold of either eye drops below 0.8 frequently, or gaze accuracy is checked to be low, then the eye cameras are readjusted and the calibration procedure repeated.

If repeated attempts of camera adjustment and calibration fail, the experiment may still continue, but the eye data is noted as having low accuracy.

Occasional gaze accuracy issues may still be ever present with a good calibration.

The calibration procedure is repeated at the end of the exam because Pupil Capture allows any calibration to be retroactively applied afterwards to the exam recording, thus overwriting previously reported gaze positions. The more accurate of the two calibration is chosen as the final one.

Best practices for setup and the calibration procedure are described in Appendix A.

4.4 Exam Procedure

The testing procedure consisted of the dSDT as well as a series of three maze tests, which were analyzed using the same Anoto pen and Pupil tracker. While the maze exams are not relevant to this paper, they are also tests of a subject's cognitive status. The three maze tests together roughly take the same time to complete as the two dSDT.

After finishing calibration, the subject begins taking the five paper tests. The dSDT is administered twice in back-to-back sessions. To discourage purposeful memorization, the subject is not informed of the second session while taking the first. The three maze tests are administered back-to-back as well. We randomize the order of presentation of the dSDT and maze test.

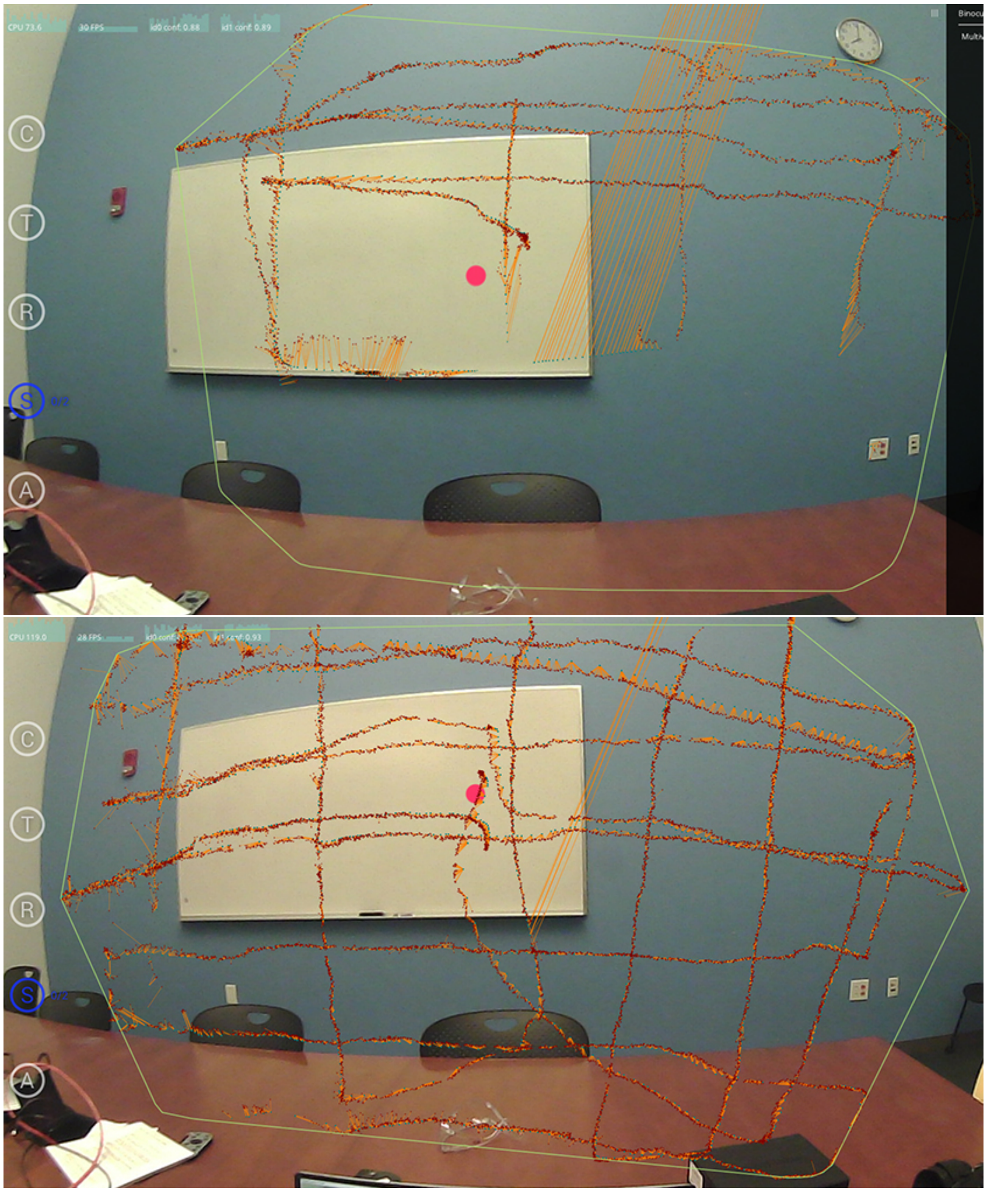


Figure 4-1: Examples of a bad calibration (top) and a good calibration (bottom). Note the bad calibration features a lack of mapped points near the bottom of the screen, indicating that the headset was unable to detect the pupil when the subject was looking down.

Chapter 5

Experimental Results

Subjects were recruited on the Massachusetts Institute of Technology campus. All subjects were either undergraduate or graduate students between the ages of 18 and 31 and presumed to be in good cognitive health. Out of the 28 subjects that were tested for the study, only 9 matched the criteria for having a good calibration and good-quality recording of the dSDT.

- 7 subjects were disqualified for having poor calibrations due to the eye camera's view of the pupil constantly being obstructed.
- 3 subjects were disqualified due to the exam forms being constantly out of view from the world camera.
- 3 subjects were disqualified for slippage of the Pupil headset.
- 2 subjects were disqualified for the pupil being undetectable due to low contrast between the pupil and the rest of the eye.
- 4 subjects were disqualified for other reasons, such as constant pupil detection dropout or recent eye surgery.

Although we were unable to collect effective gaze data from more than two-thirds of our subjects, most of the reasons for disqualification can be worked through with modified hardware and testing procedure.

The biggest issue we found when using the Pupil tracker was pupil obstruction. When subjects glanced downwards (below the plane of the horizon if facing straight), the eyelids would move down with the subject. If the subject had long or thick eyelashes, the eyelashes would often begin to obstruct the eye camera's view of the pupil, leading to weak pupil detection. This problem would be compounded if the headset sat naturally high on a subject's nose due to facial structure.

The solution would be moving the eye camera lower so that it would appear at a lower angle relative to the subject's point of view. This would lead to less pupil obstruction due to the eye camera being able to peek "under" the eyelashes. However, since the camera position can only be adjusted on the Pupil headset with mostly horizontal sliders, its range of vertical motion is limited. Since the original headset was 3D printed, new parts can be 3D printed as well to allow the eye camera a greater deal of freedom. This should solve most, if not all the issues of pupil obstruction we encountered with the subjects.

Another resolvable issue is the exam forms constantly moving out of the world camera's view. Some subjects have a tendency to slouch or hunch over while taking the test. This issue can be partially solved by giving subjects reminders to maintain their posture. Another concurrent solution would be to adjust the height of the table (or rather, of the subject's chair). The lower the table is relative to the subject, the more ground the world camera can cover. Thus, the camera's view of the testing forms would not be as susceptible to subject posture. We found that a good table height would be at the level of the subject's elbows.

The other issues noted above will require more investigation to solve.

We observed a pattern in the gaze tracking where as the test continued on, some subjects' reported gaze positions would drift upwards from their expected position. Upon performing a calibration at the end of the study, the reported gaze positions would once again match expectations, but applying the second calibration to the beginning of the study would result in gaze positions at the beginning being too low. The most logical explanation for this would be slippage in the Pupil headset. Even though the headset is fastened with glasses tighteners, the data suggests that this solution doesn't stop all Pupil headset motion. Eye videos are also produced by Pupil Capture, but the slippage seems too subtle to identify visually.

Low pupil contrast might be caused by the subject's own glasses. However, there were cases where the detection algorithm had low confidence in pupil location even though the pupil view was completely unobstructed. The algorithm may have trouble detecting large pupils - increasing the room lighting proved to be the solution for one participant. IR light emitted from near the eye camera may be another factor. Depending on the eye geometry, the IR may reflect directly back from the pupil into camera, thus causing issues with pupil contrast.

Fortunately, the majority of issues that were encountered seem to be solvable with modified hardware and improved testing procedure. We expect that future iterations of the experiment will be likelier to produce high quality data.

Chapter 6

Analysis of Gaze Data

6.1 Breaking Down Gaze Data

As discussed in Section 3.2, the Pupil Player software outputs gaze positions relative to the testing surface up to 120 times a second. All of the following visualizations were derived from a small subset of subjects who took the dSDT with fiducial markers on the edges of the page. There were 7 subjects with good raw tracking data collected without fiducial marks, and as soon as the software development of a future build of Pupil Player is completed, that data will be analyzed as well.

6.1.1 Defining Cell Boundaries

To make sense of the gaze data, we decided to break it up temporally into cell-by-cell chunks. Our rationale for this was that at each point in time while taking the dSDT, the subject is likely working on a specific cell. Either the subject is currently writing in a cell, looking for the symbol's association in the key (or perhaps looking for the symbol in an already completed cell), or distracted by another stimulus while working on the cell. To analyze the subject's thinking and learning habits, we'd like to look at the pattern of gazes that led the subject to produce the answer to an individual cell.

The first question that immediately comes up is how to define each cell-to-cell boundary. One possible solution is to define the start time of, say cell 10, as the moment the subject finishes writing (i.e. lifts the pen) in cell 9. The end time of cell 10 would then be the moment the subject finishes writing in cell 10. However, this might not be a good solution for all test takers. Some test takers might choose to leave the pen tip resting on cell 9 while simultaneously moving on to view the next symbol or key.

We decided the best solution to the cell-to-cell boundary would be to define the end time of a cell as the time of the last functional pen stroke in the cell. A pen stroke ceases to be functional when it effectively covers no ground on the page. One complication is that while the pen may appear to be stationary with the pen tip resting, the Anoto software still reports minute vibrations and position changes on the page, thus making absolute velocity thresholds unusable. As a result, we define a bounding box for pen positions across 10 pen observations. If the combined length and width of the bounding box drops below 2mm, those pen positions do not represent a functional stroke.

6.1.2 Cell Events

With our temporal boundaries defined, we can now analyze the gaze positions for each cell. One aspect of interest is gaze fixations. Gaze fixations represent the current point of interest of the subject. The timing and location of these gaze fixations can often elucidate the subject's thought process. If the subject fixates his gaze at a new symbol, he is likely storing the symbol inside his working memory so he can later cross reference the symbol across the key. If the gaze fixates on the relevant symbol and digit on the key, the subject is likely memorizing the association so it can be written into the empty cell.

Our high-frequency data allows us to detect times when the subject's gaze fixates around regions of interest on the test. The regions of interest include the key and all cells in the working area. As displayed in Figure 6-1, some regions are also specific to the cell the subject is currently working on. We pay attention to the cell and its neighbors, as well as the corresponding symbol-digit pairing in the key.

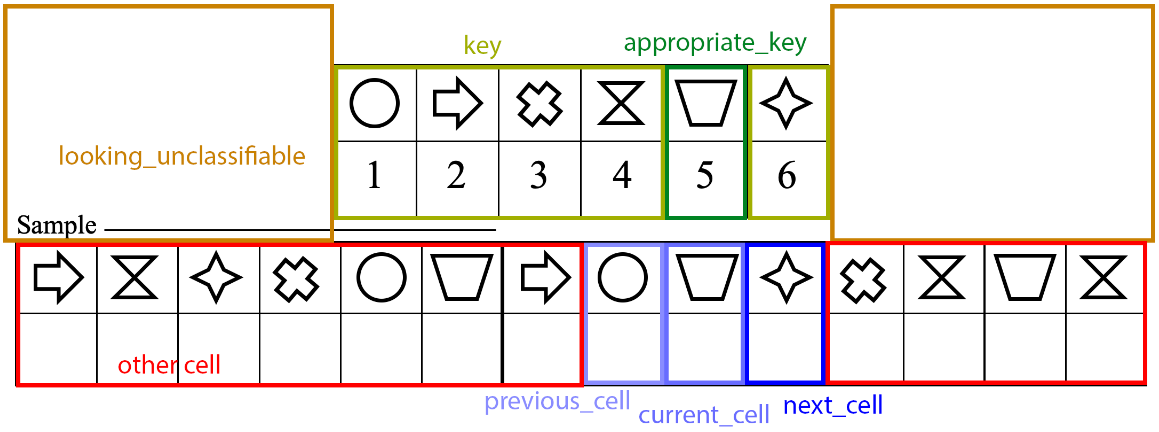


Figure 6-1: Examples of regions corresponding to the current cell - the second upside down trapezoid on the bottom row. The looking_unclassifiable region extends to all non-cell regions. The appropriate_key and current_cell region are made slightly larger than otherwise indicated since the gaze locations reported are not perfectly accurate, and these regions are the most important in our analysis.

All gaze activity is categorized into events. Events are marked with a start time, duration, and event type. When the subject's gaze lingers for at least 50ms in a region, we group all consecutive gazes in that region into a region-specific event. If the gaze duration is under 50ms, the event is marked as a transition event instead. If no vision data was provided for that time range, a no-data event is created instead.

6.2 Visualizing Gaze Data and Events

We've developed software that plots the gaze data and pen strokes together on top of the dSDT. The plot can either be played back on matplotlib [5] directly or exported as an mp4 file. The playback speed and the time duration of the gaze trails can all be configured. An example is shown in Figure 6-2.

Events can also be visualized in a multitude of ways. From Figure 6-3, we can see a typical pattern of events emerge from a cell. A subject typically first glances at the cell symbol and then moves his gaze up into the key area. After a period of searching, the subject finds the corresponding symbol in the key. The subject's gaze then drops back down into the cell as he writes his answer down. We found that this pattern

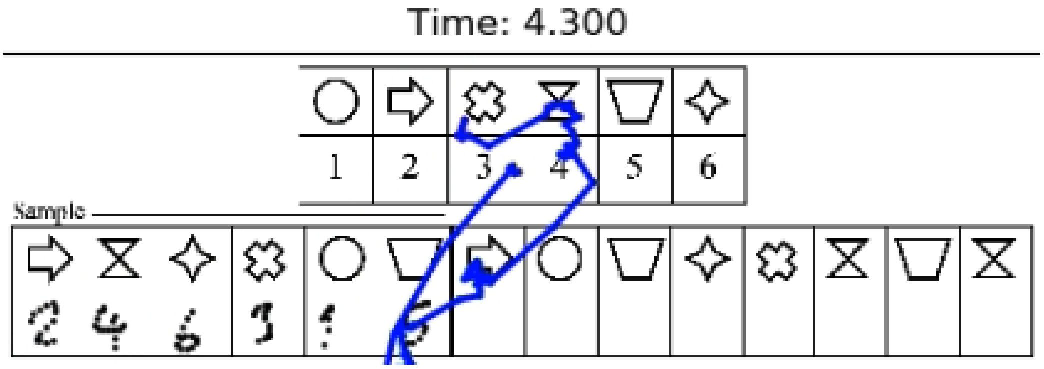


Figure 6-2: A screenshot from a video of pen and gaze playback played together.

was repeated across many subjects and seemed to be the most common approach to taking the test.

Not all cells from a subject, however, follow the standard formula. Some subjects may glance multiple times at the key while working on a cell, or not look at the key at all. A few participants were able to memorize parts of the key on-the-fly and work through 4-5 cells without glancing at the key again. Other participants preferred to look at previous cells in the working area, rather than the key, to find the corresponding symbol. Since this behavior is more difficult to track, as well as less common among participants, we decided not to integrate the looking backwards behavior into event analysis for the time being.

For the copy task, we would expect the subject to rarely, if ever glance at the key. Figure 6-4 confirms our intuition - the subject focuses only on the working cells, and never on the key area. Whether this behavior is maintained in impaired patients will be interesting to determine.

6.3 Metrics

6.3.1 Speed Metrics

We want to first be on the lookout for metrics that indicate the state of a subject's cognitive functions such as attention and scanning speed. Comparing these metrics

Event #	Description
1	looking at previous cell
2	looking at appropriate cell
3	looking at next cell
4	looking at other cell
5	looking at appropriate key
6	looking at key
7	looking unclassifiable
8	transition
9	no data

Table 6.1: A mapping between the event number and its description.

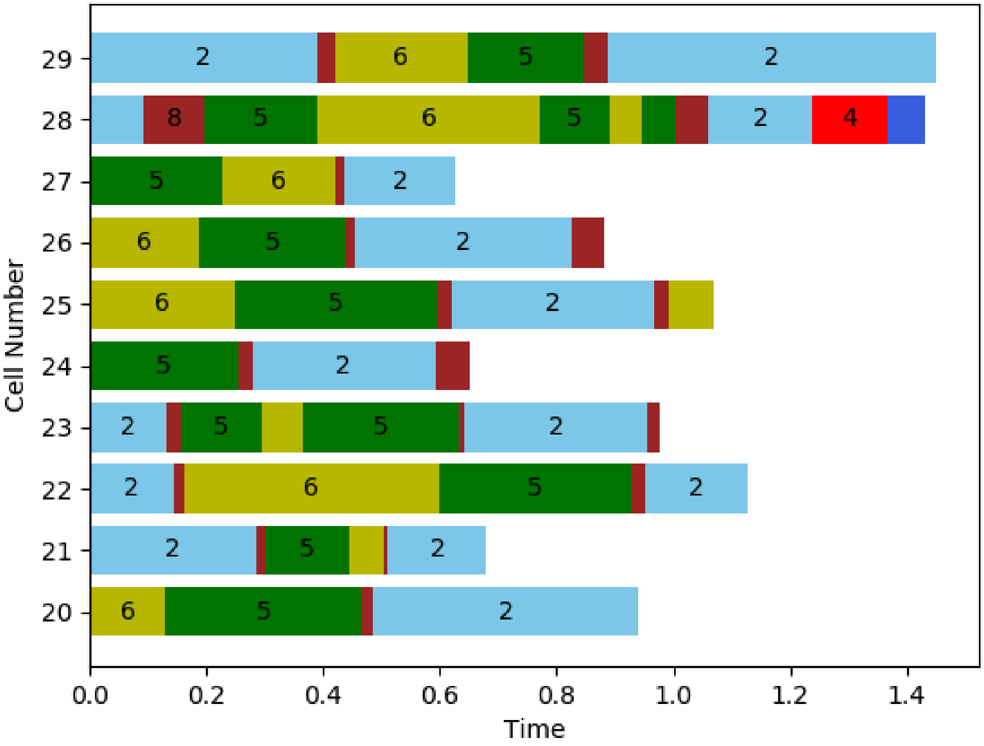


Figure 6-3: A time range plot of the events associated with 10 cells featured in the translation task. Event numbers inside the bar can be linked to the table above.

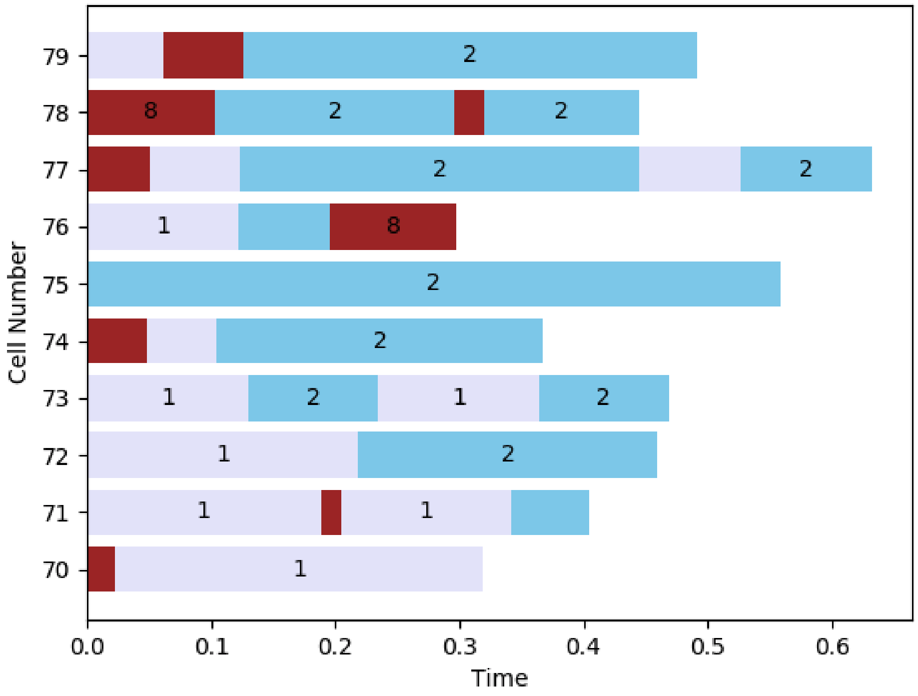


Figure 6-4: A time range plot of the events associated with 10 cells featured in the copy task. Event numbers inside the bar can be linked to Table 6.1.

subject-by-subject may lead to clues of cognitive impairment. With the event data provided, many of these metrics can be easily extracted. First of all, let us observe the total time elapsed per cell in Figure 6-5.

Using pen data collected from the 20 subjects who completed both dSDTs, we found that the translation task took a median time of 67.6 seconds to complete, and the median speedup between the translation task of the two dSDTs was 5.51 seconds. As a side note, the copy task took a median time of 27.1 seconds to complete, with a speedup of 0.99 seconds.

The gaze data from our example, as well as other subjects, subtly corroborates this finding - subjects are more familiar with the symbols and test layout, so naturally they would work slightly faster. When we begin testing cognitively impaired patients, we expect the speedup to be less pronounced since they will likely have a more difficult time with acquiring familiarity.

However, there doesn't seem to be any appreciable patterns within a test itself,

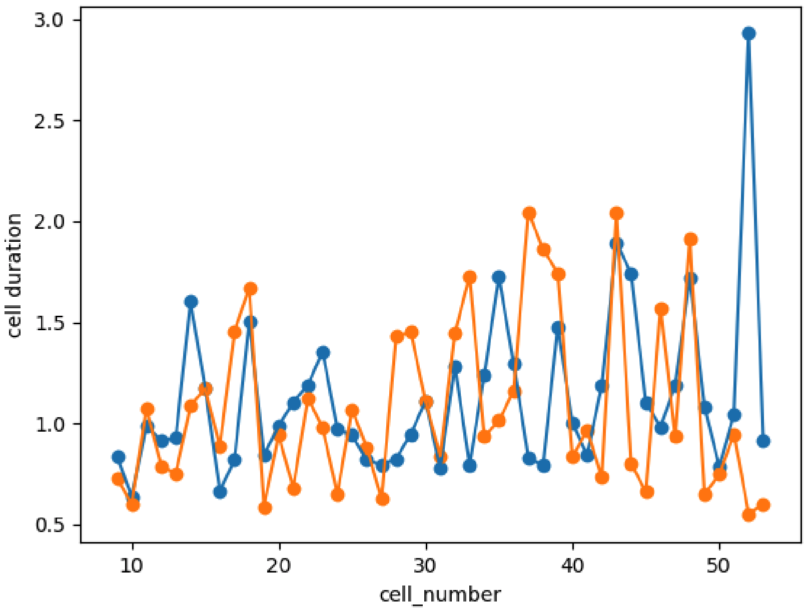


Figure 6-5: Graph of time spent on each cell for a subject. The blue (darker) line represents the subject's first time taking the dSDT, and the orange (lighter) line the subsequent.

at least from the sample gaze data we collected. While subjects may work through the cells more quickly in the later rows, they will also likely take more time to shift their gaze to the key because of the greater distance between the key and the rows. Once again, the greater distance may pose a problem for impaired patients, as they may have trouble visually navigating between the various rows to find the key.

Another metric that can be collected is writing delay. The writing delay measures how much time elapses between a subject finishing looking at the key and writing his first stroke in a cell. In the time elapsed, the subject must scan for the appropriate cell within the working area and then commit the pen onto the paper. An example is shown in Figure 6-6.

From our example, the median writing delay does drop in between tests from 0.401s to 0.316s, but we don't see any appreciable patterns within a test itself. We would expect the writing delay within a test to rise as the distance between the key and working cell increases, but that has not manifested yet. Writing delay for cognitively impaired patients may be much greater than the healthy subjects displayed here,

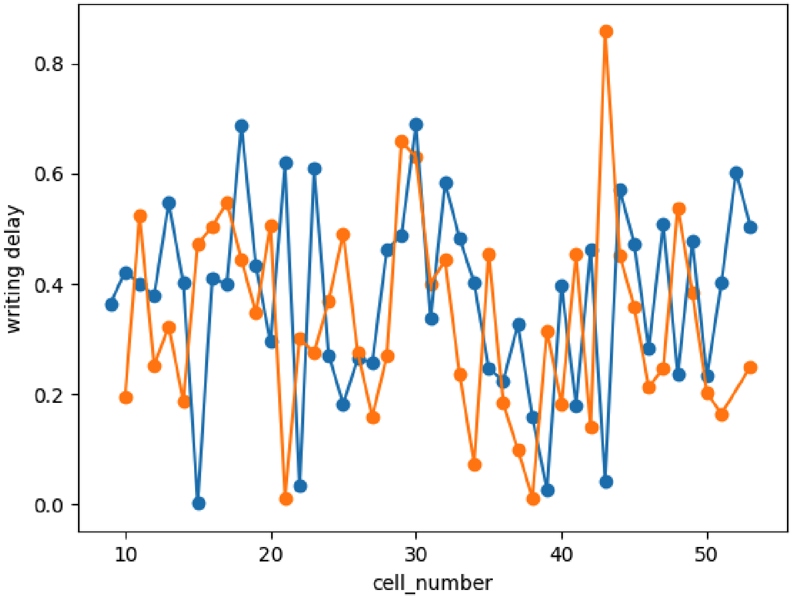


Figure 6-6: Graph of the delay between a subject finishing looking at the key and beginning to write.

since the task requires scanning speed and motor control.

Other related metrics, such as the duration of time between the subject finishing a cell and looking at the key for the next association, can also be collected. This metric shows similar properties to writing delay.

6.3.2 Metrics of Learning

We also want to look for key metrics that demonstrate a subject's learning ability. Learning can be most aptly demonstrated when we look at differences between a subject's first and second test. We'd like to also try and find the source of the 5.51s speedup between tests, as the source may underlie a key metric for learning.

As the first step, we observe the total time spent in the key per cell in Figure 6-7. The median time spent in the key per cell has decreased in this example from 0.508 seconds to 0.428 seconds, likely a product of test familiarity. As mentioned previously, work reported in [2] found that healthy subjects averaged fewer visits to the key area, as well as a shorter duration per visit. We should expect to see similar results once we test impaired patients - impaired patients will likely have higher key

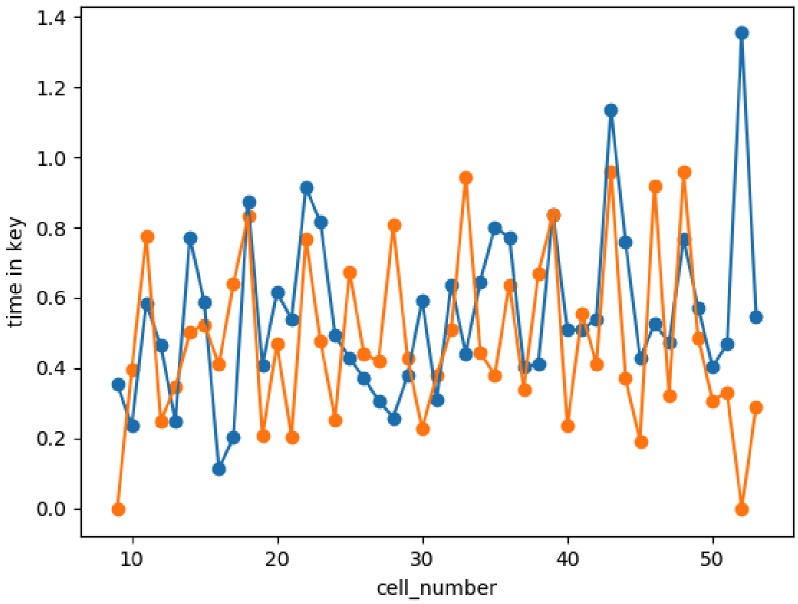


Figure 6-7: Graph of the total time spent in the key per cell.

gaze times in general, as well as a smaller dropoff between the two tests.

From where does the decrease in key duration arise? One source may be how long it takes for the subject to find the symbol in the key corresponding to his current cell. The timer starts the moment the subject's gaze lands inside the key and ends when the subject's gaze lands inside the appropriate symbol-digit pairing.

From our intuition, we would presume that as the subject becomes more familiar with the structure of the key, it would take less time for him to find the correct symbol-digit pairing. The example in Figure 6-8 suggests that this may be the case. The median time to the key pairing in this example was 0.202 seconds for the first test and 0.129 seconds for the second.

The increase in the number of data points in the second test in which the subject's gaze instantly snapped to the appropriate pairing may imply that the subject learned the physical location of the mapping on that key.

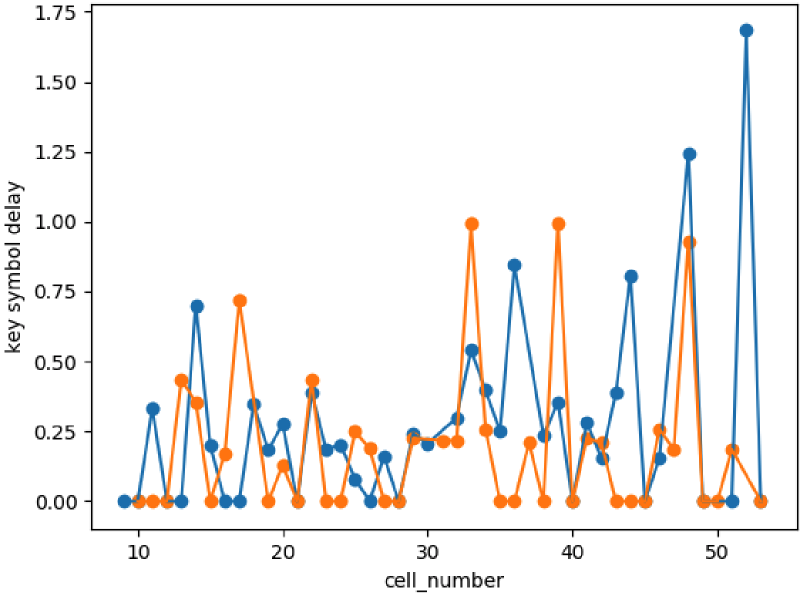


Figure 6-8: Graph of the delay between the subject's first look at the key and finding the appropriate symbol-digit pairing for his current cell. A data point at 0.0 means that the subject's first glance in the key was already in the appropriate location. Subsequent glances to the key following the timer end are ignored. Cells are skipped if the subject never glances at the appropriate symbol digit pair.

Chapter 7

Suggestions for Improvement

7.1 Hardware Suggestions

Various modifications to the Pupil headset can dramatically increase the reliability of calibration and eye tracking for future experiments.

Glasses retainers must always be used to ensure that the Pupil headset stays snug on the subject's head. Otherwise, even a slight slippage of the headset on the user's head can cause gaze offset issues. The cause of this is in Pupil software's pupil detection. Because the position of the pupil is reported in pixels coordinates relative to the eye camera, adjusting the eye camera slightly will alter the pupil coordinates and subsequently throw the previous calibration off.

Lowering the angle of the eye cameras can also make a large impact in reducing the issue of the pupil being constantly obstructed. Since the entire Pupil headset is 3D printed and the eye cameras are slid onto the frame, adjustments to the camera geometry should be straightforward.

7.2 Future Avenues of Analysis

Given the nature of the pen and eye data collected, the potential range of analysis is immense. The immediate focus for the future should be extracting gaze data from the rest of the subjects, as well as testing new subjects. With our current small sample of gaze data, finding patterns in metrics and drawing definitive conclusions is very difficult. A large sample size makes it clear which metrics solidly demonstrate cognitive ability and learning.

While the key is a focal point for the subject in obtaining associations, it is not the only location subjects base their answers on. As previously discussed, some subjects prefer to use nearby entries or the first row of entries as a reference point, instead of the key. Our event analysis needs to be updated for this use case, since our current code lumps any gazes not on the key or working area into one generic cell event. The subject's gazes on symbols that match the current working symbol should be specially categorized.

The last recall task in the dSDT is another potential source of analysis. How do average key durations for particular symbols correspond to the subject's answers in the recall task? We would expect subjects who recalled an association correctly to have spent less time looking for that symbol in the key or gazed less at the key overall. The recall task is a good indicator of a subject's memory capacity.

The speedup we observed on the translation task between the first and second test can also be further investigated. We've uncovered one potential factor in the search time related to finding the appropriate symbol/digit pair. Could the speedup also be attributed to faster visual search of all elements, less time spent in transition, or fewer glances at the key?

In addition, only healthy subjects have been tested thus far, so we can only analyze how normally functioning subjects think and acquire knowledge from their gaze data. Cognitively impaired patients may exhibit substantially different gaze patterns from normal subjects that don't fit the standard procedure we've seen from normal subjects. Testing impaired patients will allow us to see which features of the gaze data are effective in predicting impairment. So far, we only have educated guesses based on our intuition and past papers related to the dSDT/SDMT.

One confounding factor that may cause differences in task performance and our calculated metrics is age. Age has been shown to affect SDMT performance [7], and we would expect the dSDT to be similarly impacted. The copy task allows us to at least control for writing speed, but introducing new metrics more resistant to age-related impacts may be helpful.

Chapter 8

Conclusion

Answering all of these questions and exploring further requires gathering accurate gaze data from more subjects, healthy and impaired. So far in our work, we've built an end to end procedure that sets up subjects to take the dSDT while collecting their pen and eye data, presents the data in a readily accessible format, and creates a system where subjects' gaze patterns and learning habits can easily be explored and analyzed.

While some reliability issues with the Pupil tracker have been uncovered, most of these issues should be addressable with better testing practices and modified hardware. Once a large sample of subject data is collected, we can easily create new gaze/pen features that can help differentiate healthy from impaired patients. Automated scoring systems utilizing these new feature sets may significantly improve both the prevalence and accuracy of these cognitive tests.

Appendix A

Best Practices for Calibration

A.1 Pre-Calibration

- The height of the exam table should be low.
  - The lower the table is, the more that can be captured by the world camera.
  - Having a high table means the FOV of the camera is severely restricted. This makes the calibration boundaries smaller, and also makes it easier for the testing page to go out of view (by the subject lowering/raising his head).
  - A good table height would be having the table at the at the level of the subject's elbows.
- Encourage the subjects to wear contacts if they have them. Ask them if they can comfortably take the exam without glasses - the experiment is not affected by a slight nearsightedness or farsightedness.
  - The success rate of calibration is lower in people with glasses.
- The glasses tightener should be snug with the subject's head. The Pupil headset will slip without the tighteners, even if the subject doesn't wear glasses.
- The eye camera should face the subject at the lowest angle possible.

  - If the pupil cameras are too high on the face, when the subject looks down, the eyelashes are likely to get in way of the camera and block the pupil from being seen.

- Make sure the pupil camera is at an appropriate distance from the user's eyes.
  - For subjects with glasses, the contrast between the pupils and the rest of the eyes is often low - the glasses reflect some of the infrared light shining from the pupil camera.
  - This means patients with glasses should have the camera closer to the subject's eyes.
  - Having the cameras too close to the subject may be uncomfortable for the subject. The FOV of the eye camera must also cover the entire eye.

- Settings for both sliding adjusters on the frame:
  - Both sliding adjusters on each side of the pupil camera control distance of the camera to the eye. The first sliding adjuster (closest to the headset's main body, or further from the pupil camera) also moves the camera angle downwards the further out it is, while the second adjuster moves it slightly upwards.
  - Good starting settings might be having the first adjuster 3/4 of the way out, and the second adjuster 1/4 of the way out. The angle of the camera should be low.
  - The ball joint on the eye camera itself can also be rotated and moved along all 3 dimensions (to an extent). It can also be pushed slightly downwards as well to lower the angle.

- Software settings:
  - For each of the pupil windows in the software (Eye 0, Eye 1), hit 'a' on the keyboard to see a view of what the pupil detection algorithm sees.

  - Have the person look left, right, up, down (or have them follow the calibration marker) to cover all directions of pupil movement. Adjust the parameter settings (Intensity, Pupil Size) so only the pupil is colored no matter where the person's looking.
  - The Anoto software should be open and running on the desktop before uncapping the pen and starting the test. Both the Pupil and Anoto software sync their internal clocks with the computers. Since the pen and eye data are combined, they should have the same frame of reference.

A.2 Actual Calibration

- When starting the calibration, have the subject imitate the posture that he would use to take a test. This includes the head and body position/angle.
  - As the test continues, the subject will tend to lean in more, moving the test at a higher position (in the world camera) than expected. When setting the initial world camera angle, have the subject look at the test in his test posture, and try to leave more open camera space on top of the paper than below.
- Keep the laptop pointed away from the subject during calibration. If the world camera sees two copies of the calibration marker, it will not know which one to use.
- Move the calibration marker slowly during calibration. The subject may have trouble following an erratic movement or a fast-moving target. Try to cover as much of the FOV of the world camera as possible.
- If the green circle around the calibration marker disappears, it's no longer being detected by the software. Try pulling the marker back in. The entirety of the calibration marker must be inside camera view, and the calibration marker must not be too small.

A.3 Tips During the Exam

- Make sure to mark the starting times and ending times of each test in the Pupil software.
  - To mark the starting time, click on the ‘Hips Surface Tracker’ field on the right, select the appropriate form on the ‘Select Form‘ menu, and click ‘Begin Form’. Click ‘End Form’ to mark the ending time.
- If the person is leaning forward too much or slouching, the test form may go out of view of the camera. Remind them to maintain their posture.

Appendix B

Instructions Given to Subjects

The dSDT is printed on one sheet of US letter size paper (8.5 by 11 inches), in the standard upright position. The first task is positioned on the top half of the paper, and the second and third tasks on the bottom half. The test is always folded in half, so only one half of the paper is visible at a time.

The first task consists of six sample items, arranged in a permutation of the key's six items, as well as 50 main items. When starting the first task, the subject is first given guidance on completing the sample items. Any mistakes the subject makes are immediately pointed out by the examiner. After the subject completes the sample, the subject is instructed to work through the rest of the items as quickly and as accurately as possible, without skipping any items.

When the first task is completed, the examiner immediately turns the page and begins the second task. The instructions for the second task are identical to the first task, except the subject is instructed to stop at the bottom of the page, where the third task begins. Until the subject completes the second task, the examiner does not point out the third task in order to minimize attention and hopefully discourage purposeful recall from the subject.

Unlike the first two tasks, the items in the third task can be filled in any order. Since the subject may often express doubt in being able to recall the associations correctly, we simply instruct him to remember as best as possible, or guess if necessary.

After finishing the first dSDT, the subject is immediately handed another dSDT without prior warning. We make it known to the subject that the second test is an identical copy.

Bibliography

[1] B. Agrell and O. Dehlin. The clock-drawing test. 1998. *Age Ageing*, 41 Suppl 3:i41–45, Nov 2012.

[2] A. Elahipanah, B. K. Christensen, and E. M. Reingold. What can eye movements tell us about Symbol Digit substitution by patients with schizophrenia? *Schizophr. Res.*, 127(1-3):137–143, Apr 2011.

[3] S. G. Gauthier. Alzheimer's disease: the benefits of early treatment. *Eur. J. Neurol.*, 12 Suppl 3:11–16, Oct 2005.

[4] Lauren Huang. The digital symbol digit test: Screening for alzheimer's and parkinson's. unpublished, 2017.

[5] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, 9(3):90–95, May 2007.

[6] M. Pascoe, Y. Alamri, J. Dalrymple-Alford, T. Anderson, and M. MacAskill. The Symbol-Digit Modalities Test in Mild Cognitive Impairment: Evidence from Parkinson's Disease Patients. *Eur. Neurol.*, 79(3-4):206–210, 2018.

[7] Laura Sheridan, Hiram Fitzgerald, Kenneth M Adams, Joel Nigg, Michelle M Martel, Leon Puttler, Maria Wong, and Robert Zucker. Normative symbol digit modalities test performance in a community-based sample. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, 21:23–8, 02 2006.

[8] K. I. Shulman. Clock-drawing: is it the ideal cognitive screening test? *Int J Geriatr Psychiatry*, 15(6):548–561, Jun 2000.

[9] J. Van Schependom, M. B. D'hooghe, K. Cleynhens, M. D'hooge, M.C. Haelewyck, J. De Keyser, and G. Nagels. The symbol digit modalities test as sentinel test for cognitive impairment in multiple sclerosis. *European Journal of Neurology*, 21(9):1219–e72, 2014.