Matthew Re - Research Log:

Current related papers:

- Teresa Busjahn, Carsten Schulte, Bonita Sharif, Andrew Begel, Michael Hansen, Roman Bednarik, Paul Orlov, Petri Ihantola, Galina Shchekotova, and Maria Antropova. 2014.Eye tracking in computing education. InProceedings of the tenth annual conference on International computing education research. 3–10.

- Arielle S Keller, Ido Davidesco, and Kimberly D Tanner. 2020. Attention matters: How orchestrating attention may relate to classroom learning. CBE—Life Sciences Education 19, 3 (2020), fe5.

- David Rosengrant, Doug Hearrington, and Jennifer O'Brien. 2021. Investigating student sustained attention in a guided inquiry lecture course using an eye tracker. Educational psychology review 33 (2021), 11–26.

- David A Slykhuis, Eric N Wiebe, and Len A Annetta. 2005. Eye-tracking students' attention to PowerPoint photographs in a science education setting. Journal of Science Education and Technology (2005), 509–520.

- Pnina Stern and Lilach Shalev. 2013. The role of sustained attention and display medium in reading comprehension among adolescents with ADHD and without it. Research in developmental disabilities 34, 1 (2013), 431–439.

- Jiahui Wang, Pavlo Antonenko, and Kara Dawson. 2020. Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. Computers & Education 146 (2020), 103779.

Current Usable Tech: Hardware and Software

Credit: https://github.com/pupil-labs/pupil

Hardware

- Pupil labs core eye-tracker
  - 1 world camera
    - 30Hz@1080p
    - **60Hz@720p (used settings)**
    - 120Hz@480p
    - 2 Lens (more detail later)
      - Wide angle
      - Narrow angle
  - 2 eye cameras
    - 200Hz@192x192px
  - Gaze accuracy
    - 0.60 degree accuracy
    - 0.02 precision
  - Latency
    - Camera - 8.5ms
    - Processing - typically > 3ms (based on cpu)

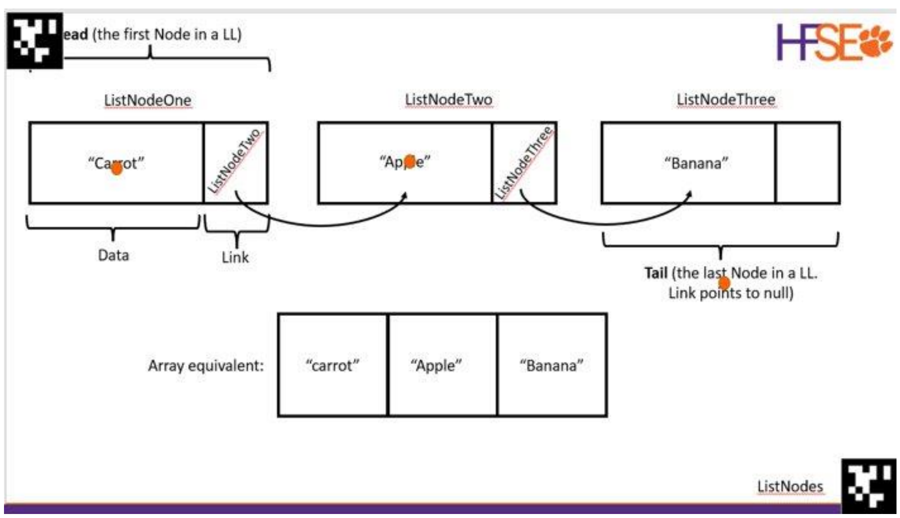
Software

- Pupil Labs Pipil Capture
- Easy to install and set up here: https://docs.pupil-labs.com/core/
  - Main recording tool
  - Multiple add-ons have been included in the study as it progresses…
    - **Fixation detector**
    - Annotator

        - Blink detector
    - These tools have lowered the work demand significantly.
- Pupil Labs Pupil Player
    - Visualizer
    - Also contains numerous plugins…
        - **Surface tracker (using AprilTags)**
        - **Fixation detector**
        - Polyline generator
        - Vis Circle

AprilTags (version 36h11)

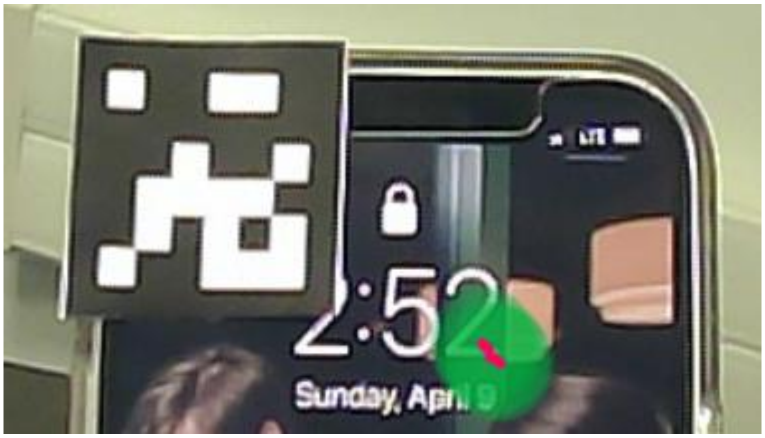
- Used for defining surfaces (for the surface tracker plugin) and AOIs.
    - Important surfaces to track:
        - PowerPoint
        - Phone
        - Laptop/notebook
        - Many more depending on scale…
- Really great to use during the initial testing of just playing with the equipment…



(Image of a PowerPoint with AprilTags in the corners to define it as a desired AoI)

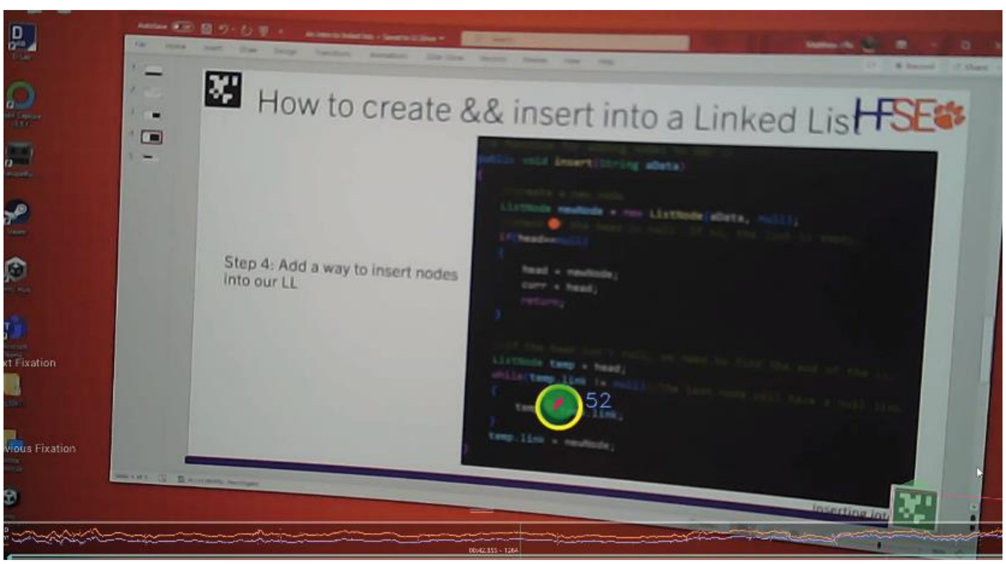
- Originally believed that "There's nothing these things can't do"...
- Apparently, there's a lot of things these things can't do…
    - Every part of the AprilTag must be visible at all times, including the entire "black border."

- If the AprilTags are on a digital screen, they must be very large in order to fight against contrast between the black border and any white backgrounds
- Sometimes, they just never get recognized…



(The unrecognized AprilTag)

- They must also be present at all times.
- Having more than one tag present can be awkward…
- They need to be in focus of the world camera
- Apriltag positioning must be done carefully, as you can only use 1 tag to define a surface; you can't use 2 tags to create a "region"



- 2 tags being used for only slide, only the bottom is being picked up and it thinks the region I want has a top-left corner of the active tag…

- However, if the AprilTags are working as expected, they are incredibly powerful. (more on this later)

Pilot to the pilot

- An initial run was conducted to get feet wet with the equipment…
    - No real goals, just learn how to use the hardware/software.
- Setup:
    - 3 surfaces/AoIs:
        - Powerpoint (digital, on a computer about 2 feet in front of the user)
        - Notes (physical, on table)
        - Phone
    - Main plug-ins:
        - Surface tracker
        - Fixation detector:
        - Vis circle
- How it went:
    - https://drive.google.com/file/d/1LX7YdMh3kNtSDhHu2YdlLNwI_LS3TRva/view?usp=sharing
    - Recordings
        - Not great…
            - The camera doesn't auto-focus, meaning that focusing has to be done before recording starts and you're stuck where you are…
            - The camera is also very sensitive to contrast in light/shadows
                - Any study should be done in a very well lit room with as little natural light as possible for consistency
            - The narrow angle lens is *very* narrow vertically
                - Sitting less than 2 feet away from your target gives you about a foot and some inches of vertical space.
        - Calibration is a nightmare…
            - (more on this later)
    - Data
        - Very useful, but very cluttered
            - Working with multiple different frames of time is weird…
                - pupil _timestamp (pupil_positions.csv) =/= gaze_timestamp (gaze_positions.csv) =/= world_timestamp (gaze_potisions_on_surface_<>.csv)

- There's' a lot of columns, but not a lot of them are helpful for this kind of project:

```
Columns present in pupil data:
['pupil_timestamp',
 'world_index',
 'eye_id',
 'confidence',
 'norm_pos_x',
 'norm_pos_y',
 'diameter',
 'method',
 'ellipse_center_x',
 'ellipse_center_y',
 'ellipse_axis_a',
 'ellipse_axis_b',
 'ellipse_angle',
 'diameter_3d',
 'model_confidence',
 'model_id',
 'sphere_center_x',
 'sphere_center_y',
 'sphere_center_z',
 'sphere_radius',
 'circle_3d_center_x',
 'circle_3d_center_y',
 'circle_3d_center_z',
 'circle_3d_normal_x',
 'circle_3d_normal_y',
 'circle_3d_normal_z',
 'circle_3d_radius',
 'theta',
 'phi',
 'projected_sphere_center_x',
 'projected_sphere_center_y',
 'projected_sphere_axis_a',
 'projected_sphere_axis_b',
 'projected_sphere_angle']
```

  -
  - Only the first 6-ish columns are helpful to varying degrees…

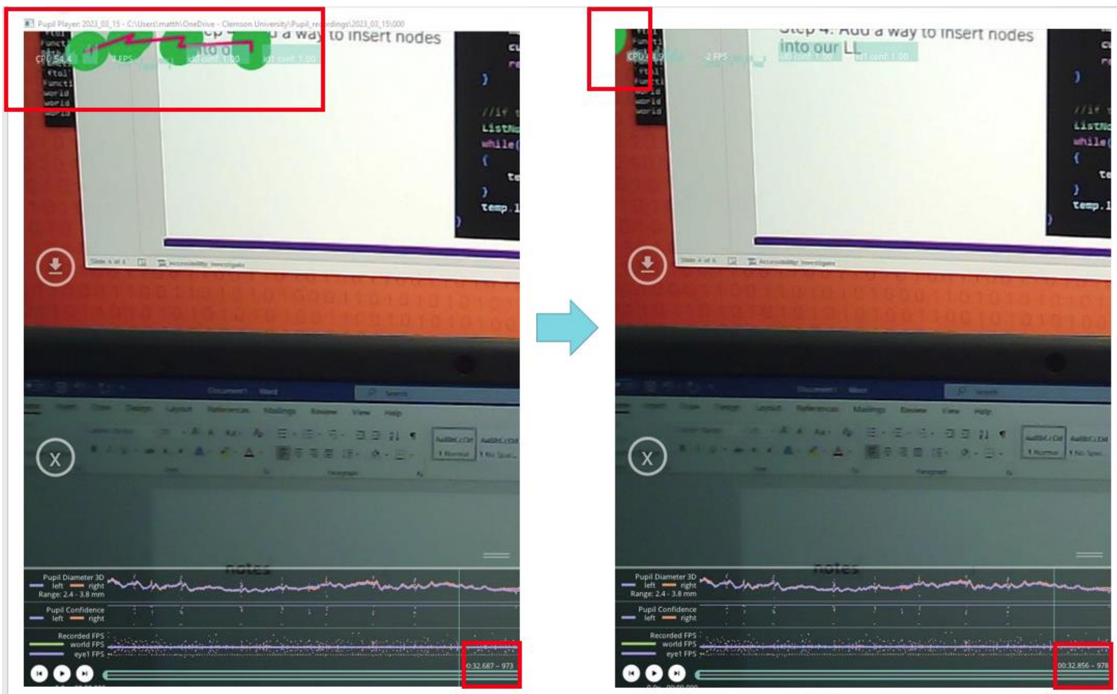
- Working with AoIs is great so long as the AprilTags work

Attention has moved off the lecture

	A	B	C	D	E	F	G	H	I
1	world_timestamp	world_index	gaze_timestamp	x_norm	y_norm	x_scaled	y_scaled	on_surf	confidence
8136	38689.57825	972	38689.58723	0.195192	0.536126	0.195192	0.536126	TRUE	1
8137	38689.57825	972	38689.59125	0.144983	0.55216	0.144983	0.55216	TRUE	1
8138	38689.57825	972	38689.59525	0.139489	0.564992	0.139489	0.564992	TRUE	0.995
8139	38689.6141	973	38689.59929	0.060833	0.588914	0.060833	0.588914	TRUE	0.995
8140	38689.6141	973	38689.60329	0.059814	0.607781	0.059814	0.607781	TRUE	0.995
8141	38689.6141	973	38689.60721	-0.03908	0.641141	-0.03908	0.641141	FALSE	0.99
8142	38689.6141	973	38689.61118	-0.03382	0.656932	-0.03382	0.656932	FALSE	0.99
8143	38689.6141	973	38689.61522	-0.12021	0.687605	-0.12021	0.687605	FALSE	0.99
8144	38689.6141	973	38689.61928	-0.11021	0.698308	-0.11021	0.698308	FALSE	0.99

Gaze_positions_on_surface_PowerPoint.csv

On the 973'rd frame, attention was lost.
What happened on this frame?



The participant just briefly shifted focus away from the PowerPoint, but it was still on the screen displaying the presentation...

  - This is the kind of analysis that I can do to answer some of the research questions
    - Namely, "When a student's gaze leaves an AoI, where is it going?" and "How long are students fixated on AoIs, and how long are they fixated on anything else?"

Pilot study

- Recording: https://drive.google.com/file/d/1BSsllf9pka1z5UVrbzojkN_37j4WKzml/view?usp=sharing
- What was done differently from the pilot to the pilot?
  - Better understanding of the headset and the FOV of the camera
  - More surfaces to track due to (a perceived) better understanding of the AprilTags
    - Slides
    - Phone
    - Table
    - Notes
  - More plugins were used:
    - Blink detector
    - Annotations
    - Better settings on the fixation detector:
      - 1.50 degrees Maximum dispersion
      - 150 milliseconds minimum duration
      - 500 milliseconds maximum duration

Data:

- Blinking is a problem, so data during a blink needs to be excluded.

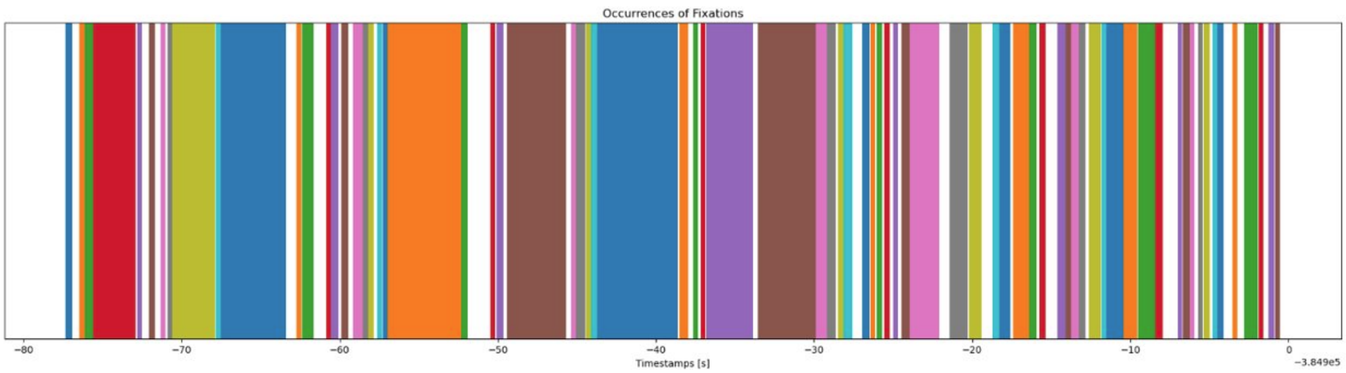
Blinks have been explicitly defined -> Exclude data at these frames

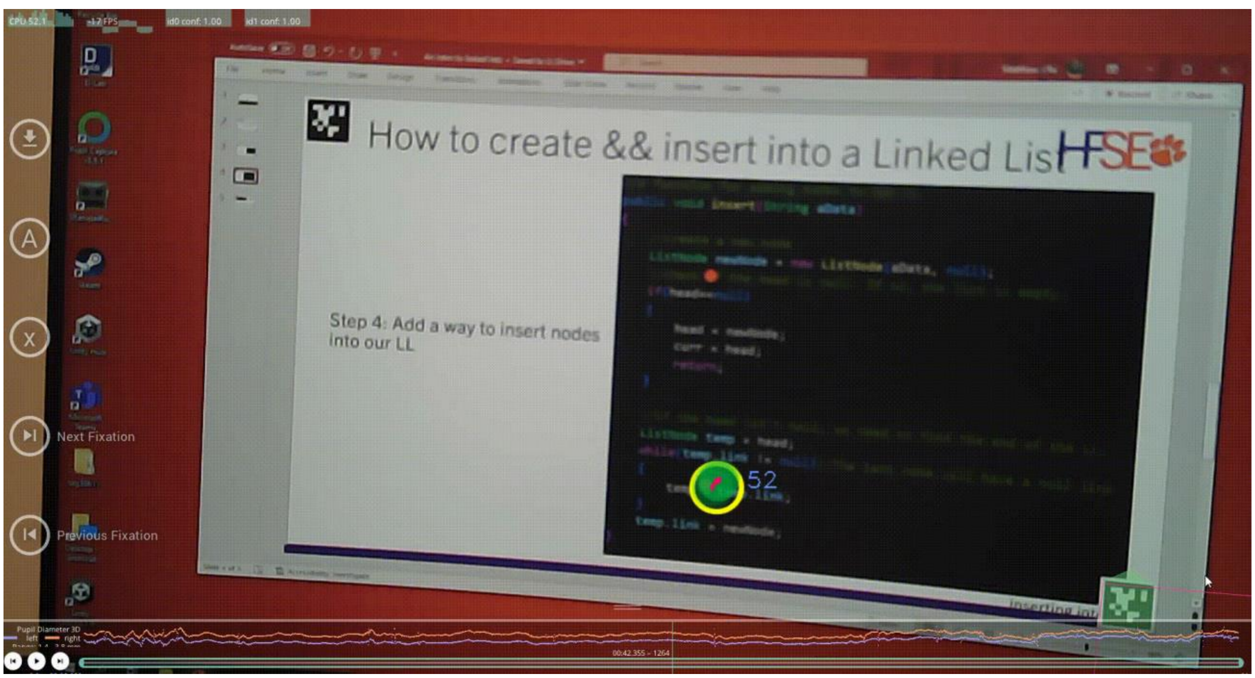
(through)

id	start_timestamp	duration	end_timestamp	start_frame_index	index	end_frame_index	confidence
1	-384981.9468	0.186098	-384981.7607	42	45	48	0.487414171
2	-384978.353	0.196334	-384978.1566	150	153	156	0.510629853
3	-384976.7602	0.231244	-384976.529	197	200	204	0.625189277
4	-384972.3489	0.193669	-384972.1553	329	332	335	0.465656464
5	-384959.3319	0.12668	-384959.2052	717	719	721	0.363278367
6	-384949.6491	0.194677	-384949.4544	1006	1009	1012	0.573030206
7	-384938.1489	0.239724	-384937.9091	1349	1353	1357	0.676121876
8	-384926.3648	0.18876	-384926.1761	1701	1704	1707	0.474344568
9	-384920.4325	0.228655	-384920.2039	1878	1881	1885	0.533404772
10	-384915.2122	3.387395	-384911.8248	2034	2084	2135	0.080450687
11	-384907.6412	0.176286	-384907.4649	2260	2262	2265	0.418373796
12	-384905.0291	0.171829	-384904.8572	2338	2340	2343	0.438174777
13	-384903.068	0.201502	-384902.8665	2396	2399	2402	0.515385808

- How do we visualize the fixations? In other words, how do we identify what was just a fixation due to the definition of a fixation and what was a fixation because there was something relevant to this project (attention focus) occurring?
  - We can visualize the length of all the fixations, helping us identify longer points of focus…
    - The longer the bar, the longer the fixation
      - Longest ones are near timestamp -384980, -384970, -384966, -384955, -384948, -384940…

Text(0.5, 1.0, 'Occurrences of Fixations')





Fixation near that occurred on timestamp -384940

- Getting better at parsing through the data
  - 3 surfaces were defined:
    - Powerpoint
    - Notes
    - Phone

- We can easily get the percentage of gaze points (of total) from surface_gaze_disrobution.csv:

total_gaze_point_count	20079	
surface_name	gaze_count	% Total
PwerPoint	5307	26%
Notes	1896	9%
Phone	4	0%
not_on_any_surface	12917	64%

- The surface_events.csv file also gives a list of enter/exit occurrences:

world_index	world_timestamp	surface_name	event_type
0	-384983.3511	PwerPoint	enter
18	-384982.7481	PwerPoint	exit
26	-384982.48	PwerPoint	enter
35	-384982.18	PwerPoint	exit
52	-384981.6079	PwerPoint	enter
163	-384977.892	PwerPoint	exit
173	-384977.5561	PwerPoint	enter
176	-384977.4542	PwerPoint	exit
[illegible]	[illegible]	[illegible]	[illegible]

- Annotations, while I still haven't gotten anything useful out of them from a "I want to answer a research question" standpoint, they are helpful for confirming calibration.

index	timestamp	label	duration	added_in_capture
240	-384975.3225	Orange dot fixation	0	TRUE
395	-384970.1279	Orange dot fixation	0	TRUE
502	-384966.5374	Orange dot fixation	0	TRUE
804	-384956.4246	Orange dot fixation	0	TRUE
1030	-384948.8352	Orange dot fixation	0	TRUE
1202	-384943.0705	Orange dot fixation	0	TRUE
1423	-384935.6769	Orange dot fixation	0	TRUE
1508	-384932.8325	Orange dot fixation	0	TRUE

Final Remarks on pilot study

- What did the pilot study yield?
    - Will consider running another quick run-through to determine if wide lens is better for classroom usage.

    - With better practice with the headset, the narrow lens is fine for up-close recordings.
    - Wide lens will likely be required if I want to use Densepose.
    - Data analysis is very beginner friendly with the Pupil Player software.
- Is this what I wanted?
    - Still need to examine the difference between full-scale classroom and up-close PowerPoint.
    - How can Densepose lead to more interesting results?
    - Other than scale and whether to use Densepose, the pilot test shows I can get the data I want
- What can be done better?
    - AprilTag usage needs to be done carefully. Only 1 tag per surface.
    - Practice focusing the world camera.
    - Find a better definition for fixations (min and max time).

Data collection proper

- The “final” data collection session was scheduled on 4-7-23
    - Takes place during an actual java review session.
    - Several AoI’s
        - Presented powerpoint (TV)
        - Whiteboard
        - “Student” laptop (notes)
        - “Student” phone
        - Presenter’s person (using densepose)
    - Unfortunately, the “final” data collection session was doomed from the start.
        - It was 44 **minutes** long.
            - The world video was 16GB large.
            - Even the individual eye videos were over a gig each.
            - Csv files were well over 400,000 rows large, which is way too much for even my desktop to handle.
        - The room for the final data collection was changed last second, so we were in a room that had a lot of obstacles between the “student” and the presenter.

        - The room was not well lit, meaning the digital AprilTags were constantly failing to register with the eye-tracker.
        - Used the wide-angle lens for the first time for proper data collection.
            - Great because everything fits in the frame.
            - However, the focus of different necessary elements that were not in the same 2D plane in space were inconsistent at best.
    - Even issues with the headset itself
        - Clearly isn't "one size fits all."
        - Getting both eyes into view of the individual eye cameras without letting the eye cameras just slip off the tracker was difficult.
            - Results in falling confidence levels for the eye captures.
- Overall, this data was considered **unusable…**

Final data collection (for real this time)

After the failure of the last final data collection, another study was conducted on 4-9-23.

Fundamentally, what was done differently?

- Different room with nothing in-between the "student" and the presentation
- Switched back to the narrow lens
    - for better focus at the cost of a worse FOV
- AprilTags were printed on physical paper
    - To avoid contrast issues
- Instead of 1, longer presentation -> use 4, short presentations
    - Good vs bad
        - 1 "good" presentation
            - Black text, white background
            - Easy to read font
            - Subtle use of figures
        - 1 "bad" presentation
            - Same material as the "good" presentation
            - Dark grey text, blue background
            - Horrid, but still readable font

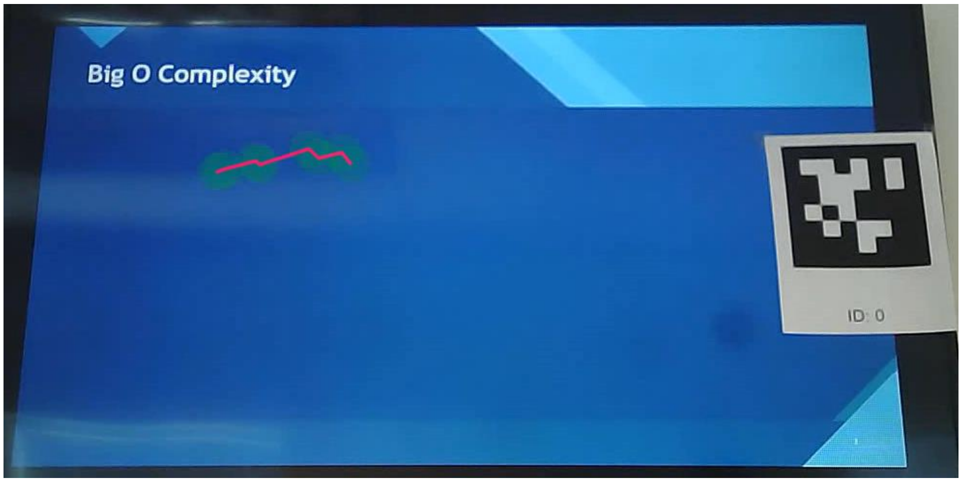
        - Figures were much less subtle
    - Figures vs no figures
        - 1 “figure heavy” presentation
            - If it could be described using a picture, use one
        - 1 presentation with no figures
            - All text

Final data collection design:

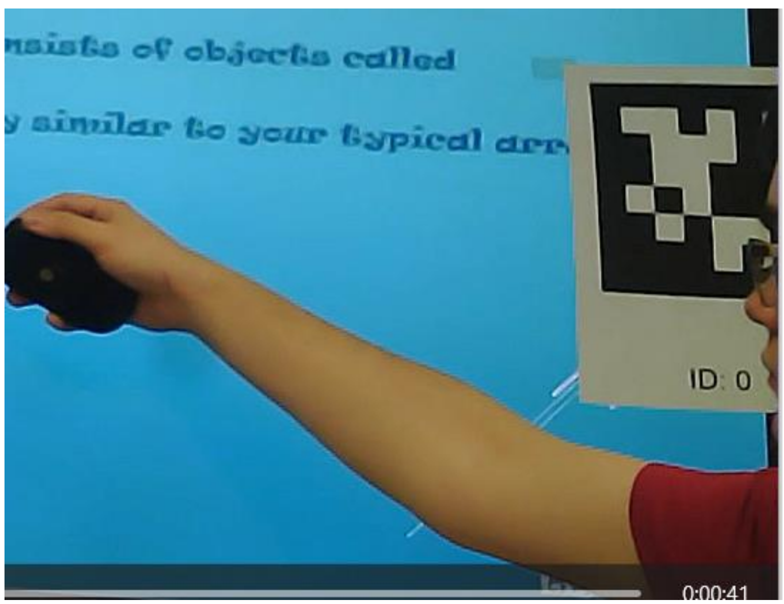
- Once all the presentations were ready, they were presented to the student one at a time in a research lab in McAdams, not a real classroom as to avoid more last minute scheduling conflicts and obstacles between the student and the presentation.
- Because the goal was to just prove that this methodology can yield usable data, a lot of things were “rigged” to see what the data output would look like
    - At one point during the presentation, the student was told to “stare off into space for about 5 seconds” and an annotation was used to mark that time.
- Other than instances like the one above, the student was told to behave “naturally” and shift between the presentation or the notes when needed.
- In-between each presentation, the recording was cut off and the eye-tracker was recalibrated.
- There was a presenter for the first two presentation, who read the slides standing next to the projector screen and occasionally pointed to things on the slide. This was to be used for Densepose, but as noted later in this log, Densepose ended up not being compatible with the Pupil Core headset.
- The last two presentations did not have a presenter standing in front of the TV. The presenter just read the slides from outside of view. This was to see if smaller details that were pointed to in the first two presentation were missed due to the lack of an in-person/visible presenter.

What issues did we still run into?

- AprilTags
    - Had to be massive to ensure focus wouldn’t be an issue…

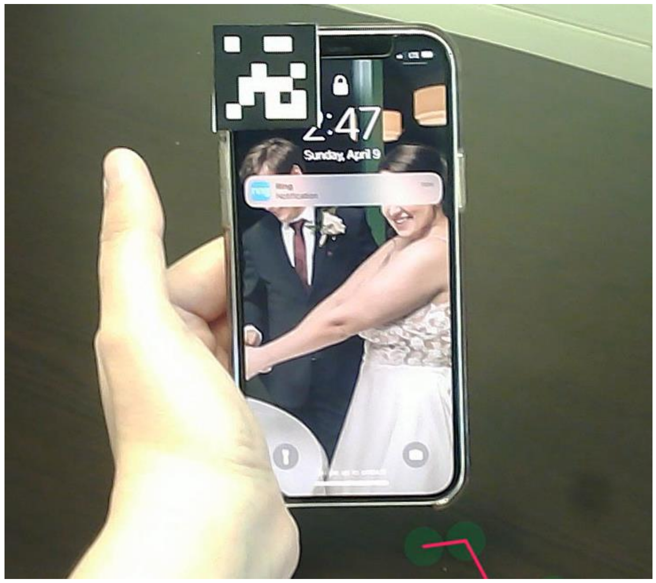


- Stopped having issues with it going out of focus
- But now a massive AprilTag takes up a large chunk of the view…
- During the presentation where I was a presenter (for densepose), if I blocked any of the AprilTag, it was rendered useless.



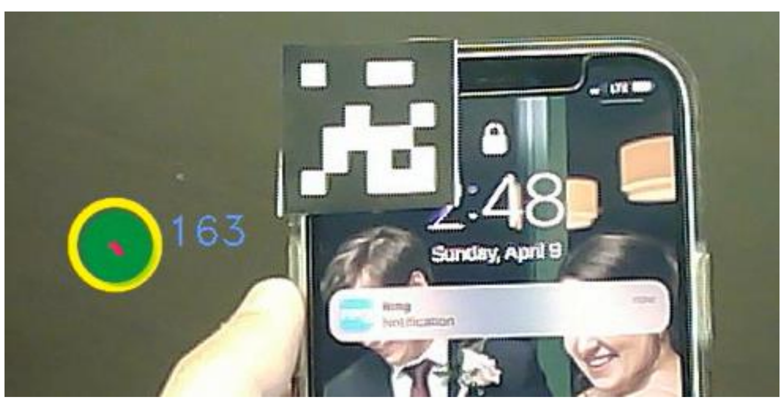
(blocked by presenter)

- Sometimes, AprilTags just never got picked up…



(AprilTag in view, but not picked up by tracker)

- Densepose
  - Ultimately, Densepose ended up not being supported by the pupil labs **core** eye-tracker.
  - All the documentation references Densepose for the pupil labs **invisible.**
  - Unfortunate, but it only means we drop one of the RQs ("Are students giving direct attention to the instructor?"
- Calibration
  - The eye tracker required **frequent** calibration, even during short data collection periods.
  - Sometimes, calibration was off even immediately off being calibrated

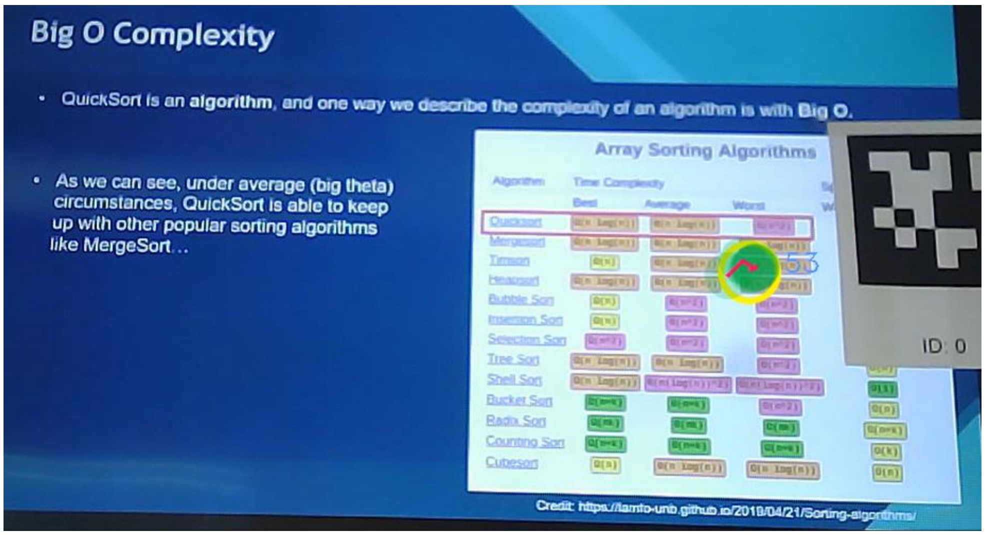


("Student" is looking at the ring notification, but calibration is off 40 seconds into recording)

■ This made identify the **true** values of whether the participant was on or off a surface very difficult, but still workable.

Data collected.

- It looks promising…
- We want to look at whether students' attention can be retained via the use of figures, and if so, for how long?



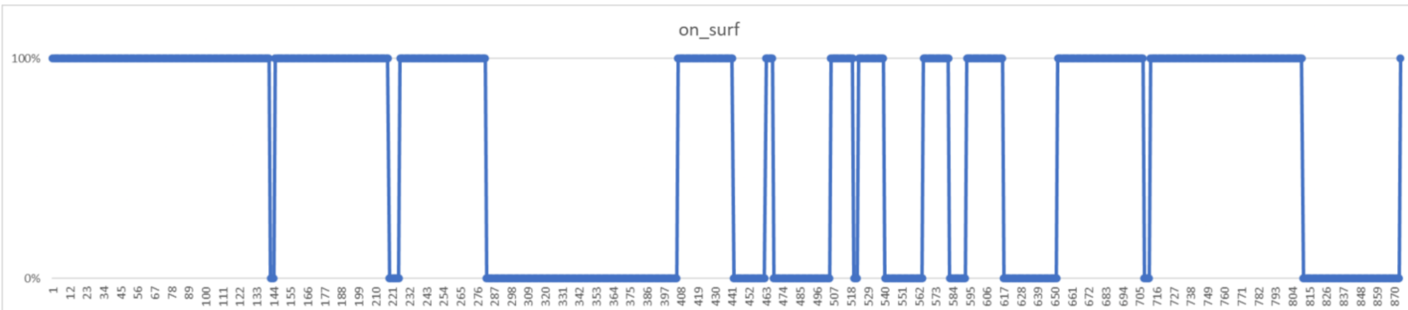
- One question would be how long is the student looking at the figure?
  - We can use the one AprilTag to define 2 regions:
    - Presentation
    - Figure
  - The figure is only present for about 1150 frames out of the total ~3500
    - We can separate the region of code where the figure is present…

1	world_timestamp	world_index	gaze_timestamp	x_norm	y_norm	x_scaled	y_scaled	on_surf	confidence
1418	-755752.8193	169	-755752.8224	0.267145	0.959867	0.267145	0.959867	TRUE	1
1419	-755752.8193	169	-755752.8184	0.268276	0.959207	0.268276	0.959207	TRUE	1
1420	-755752.8193	169	-755752.8144	0.272367	0.963442	0.272367	0.963442	TRUE	1
1421	-755752.8193	169	-755752.8104	0.273193	0.957785	0.273193	0.957785	TRUE	1
1422	-755752.8193	169	-755752.8064	0.272498	0.958182	0.272498	0.958182	TRUE	1
1423	-755752.7873	170	-755752.8024	0.272275	0.951245	0.272275	0.951245	TRUE	1
1424	-755752.7873	170	-755752.7983	0.276024	0.955293	0.276024	0.955293	TRUE	1
1425	-755752.7873	170	-755752.7943	0.274372	0.95417	0.274372	0.95417	TRUE	1
1426	-755752.7873	170	-755752.7903	0.277336	0.951018	0.277336	0.951018	TRUE	1
1427	-755752.7873	170	-755752.7843	0.27908	0.947514	0.27908	0.947514	TRUE	1

    - The figure appeared on frame (world index) 170.
    - Separate this data where the figure is present and get basic stats…

Total Frames:	1147
Number of Frames Looking at Figure:	540
Percent (Frames) on Figure:	47%
Total Time - Figure on screen (seconds)	38.436
Time on Figure (seconds):	25.367
Time off Figure (seconds):	13.069

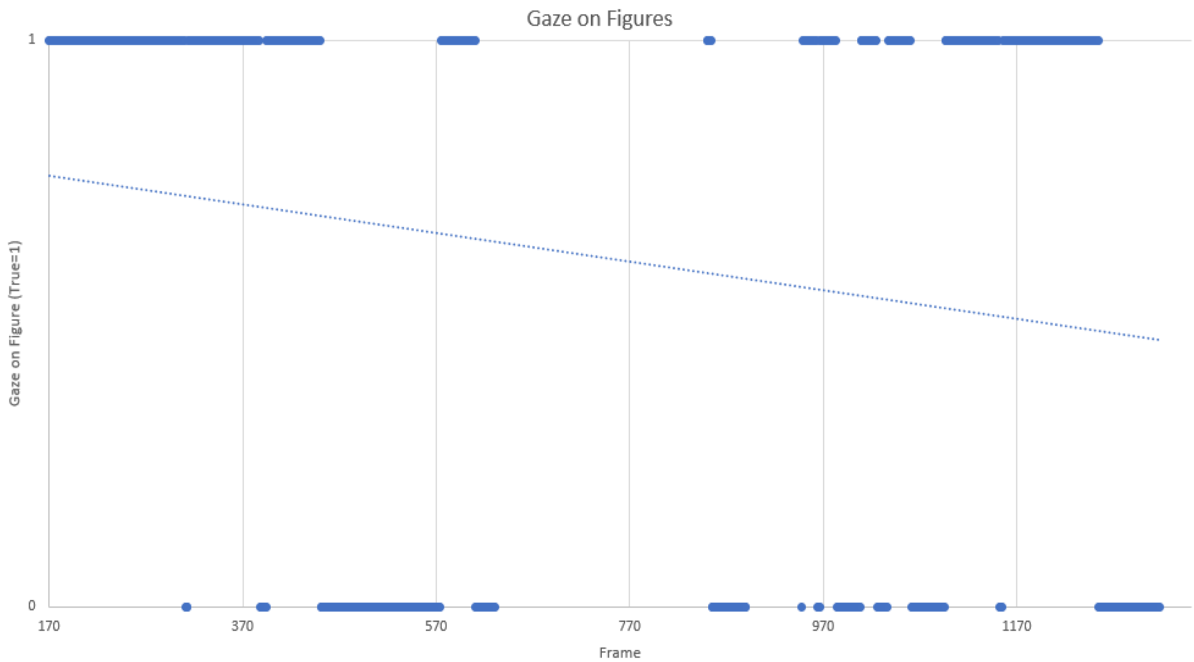
▪ Produce a visualization in reference to time…



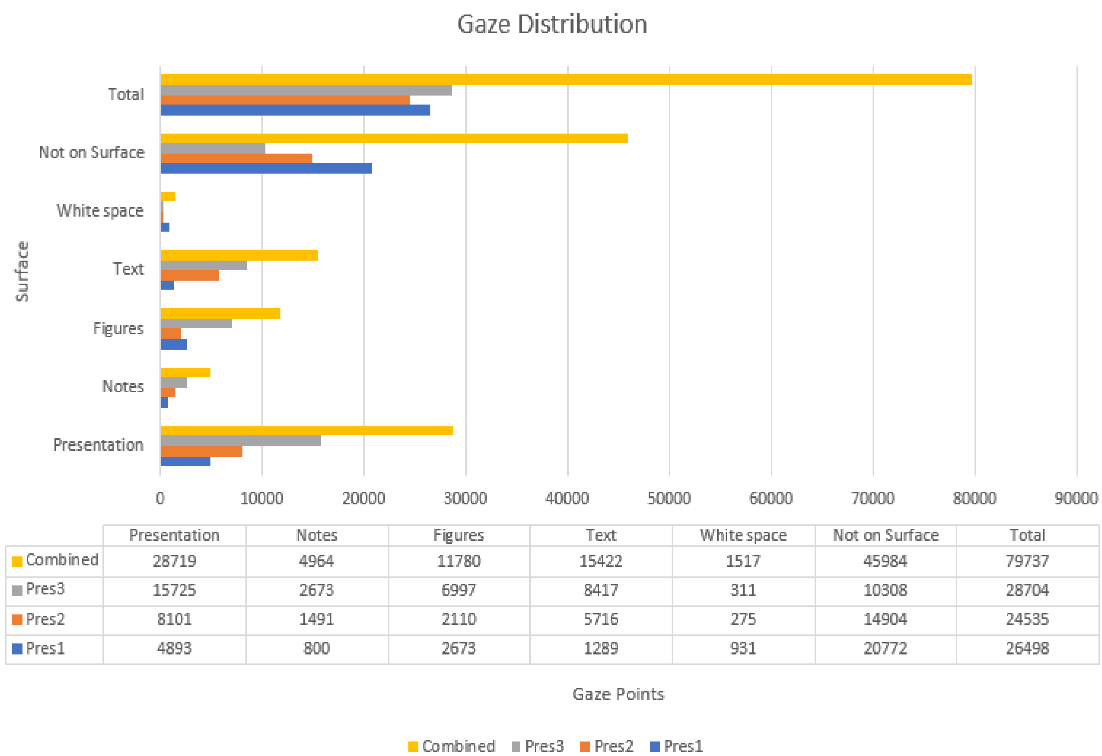
(100% means on surface, 0 means off)

▪ We can see that after about 400 frames, the focus of attention on the figures starts to waver a lot more compared to when they first appeared.
    - Perhaps that means that there's a "cut-off point" for the usefulness of a figure?

- Overall, the goal of this project was to evaluate how well this methodology could be used to investigate some points of interest:
    - What components of the presentation are students focused on?
    - How long are students fixated on Areas of Interest (AoI)?
    - When a student's gaze leaves an AoI, where is it going?
    - Are there areas outside of AoIs that attract student gaze?
    - Is there a defined pattern that appears to explain why students move to and from the different dimensions of attention outlined in Keller et al.?
- Of these, the first three were rather simplistic to investigate…



(A better mock-up of what gaze points across time could look like, with a trendline that could help investigate the 2nd point of interest.)



	Presentation	Notes	Figures	Text	White space	Not on Surface	Total
Combined	28719	4964	11780	15422	1517	45984	79737
Pres3	15725	2673	6997	8417	311	10308	28704
Pres2	8101	1491	2110	5716	275	14904	24535
Pres1	4893	800	2673	1289	931	20772	26498

(A visualization of the breakdown of the gaze distribution, which helps investigate point 3.)

- An improvement over the use of AprilTags as well as a thorough qualitative analysis of the world recordings post-export (to view vis circle/polyline) would allow for the same methodology to help investigate point 4, but as project stands in its current design this point would be very tedious to investigate.
- For the final point, this methodology could help explain *part* of the trends regarding student attention. The referenced paper briefly discusses the use of eye-tracking for investigating these trends, but because behavior and habits rely so much on psychology and neuroscience, there would need to be more components (such as the named EEG) of this study to adequately assess the point of interest.

Usage of This Methodology in Future Iterations

Overall, if I wanted to turn this project into a proper research study, it would likely take a similar form to what was outlined with the pilot but with a few points of improvement.

- First and foremost, the distinction between "attention" and "visual attention" would be must more explicitly defined.
    - The study of just all-around attention requires a lot more than just eye-tracking and exceeds outside the scope of a single computer scientist.
- Obviously, the number of participants would be greater than 1, and they would not be related to the researcher and would not already a very specific understanding of the project.
- I would recruit about 5-10 participants for each presentation. PowerPoints 1 & 2 and 3 & 4 use the same information/text, so a between-subject design would be more appropriate to ensure that focus is not being lost due to familiarity with the material during presentations 2 and 4.
- This study would take place in a "natural classroom setting" instead of a private, quiet research lab with no potential for outside distractions.
    - I would even consider getting some "intentional distractions" like the sound of someone coughing from the hallway to see if that produces any interesting results with attention.
- AprilTags would be a combination of digital and physical markers.
    - Each figure should receive its own, unique AprilTag for tracking as to keep its surface separate from the presentation itself.

    - The presentations themselves could use larger, physical AprilTags since the space the presentation is on is consistent.
    - Anything physical that we would want to track, like the students' computers or phone, should also have physical AprilTags.
- The idea of a visible vs non-visible presenter is *interesting,* however that variable should probably be left consistent across all four presentation to reduce the risk of overloading the data.
- The timing for each presentation, which was about 2 minutes per, was good and would be left as-is.
- Once all of the recordings are collected, it should be visualized and exported using Pupil Player.
    - We need surface trackers
    - Vis circles & polylines
    - Fixation detector
        - There is no consistency with what the research defines as the "perfect length" so 150ms to 350ms would be used.
- Once all of the data is exported and the world videos are ready, most of what we need is going to be in either the gaze_positions.csv file or within the surfaces folder, so for ease of use those files would be separated.
- The data between the individual surface files (gaze_position_on_surface_PowerPoint.csv) would be joined with the gaze_positions.csv file on the gaze_timestamp columns to relate the individual surface data with the world data.
- Once joined, it's easier to begin working with the gaze distributions, fixations, and everything outlined in the pilot study section.
- Qualitative data, if necessary, could also be collected by viewing the exported world videos alongside the surface_events.csv file. Notes could be taken about indicators of lost or gained visual attention based on how the vis circle and polylines move within the presentation.