Practical Considerations for Low-Cost Eye Tracking: An Analysis of Data Loss and Presentation of a Solution

Ciara Sibley[1(✉)], Cyrus K. Foroughi[1], Tatana Olson[2],
Cory Moclaire[2], and Joseph T. Coyne[1]

[1] Naval Research Laboratory, Washington, D.C., USA
{ciara.sibley, cyrus.foroughi.ctr,
joseph.coyne}@nrl.navy.mil
[2] Naval Aerospace Medical Institute, Pensacola, FL, USA
{tatana.m.olson.mil, cory.m.moclaire.ctr}@mail.mil

Abstract. This paper presents data loss figures from three experiments, varying in length and visual complexity, in which low-cost eye tracking data were collected. Analysis of data from the first two experiments revealed higher levels of data loss in the visually complex task environment and that task duration did not appear to impact data loss. Results from the third experiment demonstrate how data loss can be mitigated by including periodic eye tracking data quality assessments, which are described in detail. The paper concludes with a discussion of overall findings and provides suggestions for researchers interested in employing low-cost eye tracking in human subject experiments.

Keywords: Eye tracking · Data quality · Data loss · Supervisory control

1 Introduction

Several commercial off-the-shelf low-cost eye trackers have emerged on the market in the last few years, providing researchers the opportunity to inexpensively and unobtrusively collect eye tracking data across a variety of experimental protocols. Specifically, the Gazepoint GP3, Eye Tribe and Tobii EyeX have been available for under $500. Unfortunately, however, Eye Tribe is no longer selling their system due to its recent acquisition by Oculus [1] and the Tobii EyeX user agreement prohibits data from being recorded. Mobile or glasses-worn eye trackers are another low-cost option, but these are notoriously uncomfortable to wear for extended time periods. As such, at this point in time, the Gazepoint GP3 is the only off-the-head low-cost tracker truly available for research purposes.

Likely due to their nascence, only a limited number of studies have assessed the viability of low-cost eye trackers for research purposes. One team of researchers investigated the fixation accuracy and precision of the Eye Tribe system and generally found its performance acceptable for their research purposes [2, 3]. Another study concluded that Eye Tribe's pupillometry measurements were comparable to those of a high-quality tracker, when participants were exposed to black and white screen

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC 2017, Part I, LNAI 10284, pp. 236–250, 2017.
DOI: 10.1007/978-3-319-58628-1_19

backgrounds [4]. This ability to capture pupillary responses to changes in screen luminance was confirmed by the authors, who also included assessment of the Gazepoint GP3 system, and furthermore found both trackers capable of identifying pupillary responses to cognitive workload [5].

Researchers from the Air Force Research Laboratory conducted a more in depth performance comparison of two low-cost eye tracking systems (Eye Tribe and Tobii EyeX) to three more expensive alternatives [6]. They primarily assessed accuracy and precision of each system during a 9-min fixation task, but also provided data quality measures, as defined by the amount of data samples dropped by each system. Their analysis revealed that the low-cost trackers experienced more data loss than the higher cost-systems, with percentages of useable data at approximately 78% for both low-cost systems and between 90 to 100% useable data for the higher-cost systems.

The few evaluations that have been conducted to date have utilized short and visually simple tasks, involving either static images or fixation points. Data loss was typically reported in each study, but not discussed at length. This paper focuses exclusively on data quality, or data loss, since the authors believe this is an issue that is often overlooked but critical in determining whether low-cost eye tracking systems are appropriate for use in research and across a variety of experimental protocols, including longer tasks within visually complex environments.

Specifically, this paper presents data loss figures from the Gazepoint GP3 eye tracking system across three separate experiments. Tasks across each experiment varied in visual complexity and duration. The next section, Sect. 2, presents data from an experiment comprised of several short and visually simple tasks. Section 3 presents data from a longer, more visually complex experiment. Section 4 presents a technique used to mitigate data loss and shows improved results after its implementation. The final section discusses overall findings and provides suggestions for researchers interested in using low-cost eye tracking.

2 Experiment 1: Data Loss Across Visually Simple Tasks

2.1 Method

Participants. Eye tracking data was collected from 25 participants (24 male, 1 female) who were Naval and Marine Corps student pilots. They ranged in age from 22 to 29 (M = 23.76, SD = 2.24). An error occurred with one of the data files, so data from 24 participants are presented here.

Equipment. The Gazepoint GP3 eye tracking system was used to collect data from participants. This system is recommended for use with single displays up to 24″ and provides data at a 60 Hz sampling rate. Data recorded includes a user's left and right pupil diameter (in pixels, corresponding to a fraction of the camera image size) and left and right point-of-gaze (x and y-coordinates on the screen). The software also enables capture of the location of each eye in 3D space, with respect to the camera, as well as pupil size, all in meters. Fixation data (x and y-coordinates and duration) is also available. The system provides binary "validity" values for the following measurements:

left pupil size; right pupil size; left eye point-of-gaze (x and y screen coordinates); right eye point-of-gaze; average point-of-gaze; and fixation point-of-gaze. The validity parameter is coded as “1 if the data is valid, and 0 if it is not.” [7]

Each eye tracking unit was centered immediately below a 17 inch monitor (1280 × 1024 resolution), using Gazepoint’s tripod set up, as shown in Fig. 1. Eye trackers were placed at approximately arm’s length distance from the participant, as instructed in Gazepoint’s user manual. The appropriate distance is also verified using the native calibration software controller, discussed below.

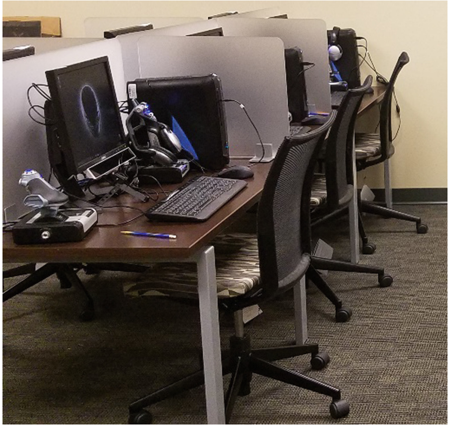


Fig. 1. Laboratory set up showing Gazepoint GP3 beneath a 17 inch monitor

Procedure. This experiment took place in a group setting in which participants were seated at their own station, but beside other participants, as seen in Fig. 1. Data collection occurred over two sessions. Upon arrival, participants were provided informed consent documents. After giving consent, participants completed a brief demographic survey and then began Gazepoint’s set up and calibration process. During set up, the user is shown a screen that verifies the camera is well positioned to track both eyes (see Fig. 2) and that the user is sitting at an appropriate distance. Distance is assessed by the dot shown above the image of the face; the dot moves horizontally across the top of the screen, shifting from red on the far left (user is positioned too far away from the camera) to green within the middle of the screen (user is positioned well) to red on the right (user is positioned too close to the monitor).

Each participant was verbally instructed to verify that their eyes were centered in the images and that the distance dot was green and positioned close to the center of the screen. If either was not true, they were told to move the camera and/or their body position. Experimenters then verified each participant’s settings, after which participants were instructed to continue to the calibration. During calibration, participants tracked a white dot around the screen to nine different locations, which were presented in a 3 × 3 grid pattern. At the end, participants were able to see their eye gaze rendered on the screen in real time in order to qualitatively verify the accuracy of their

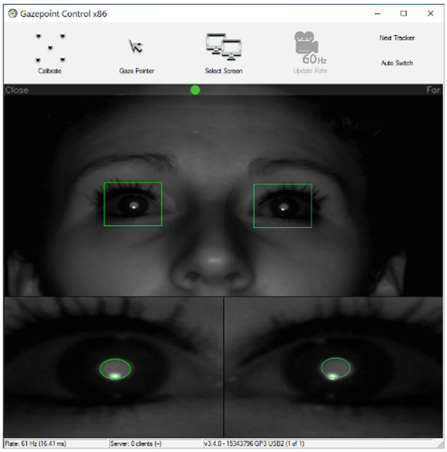


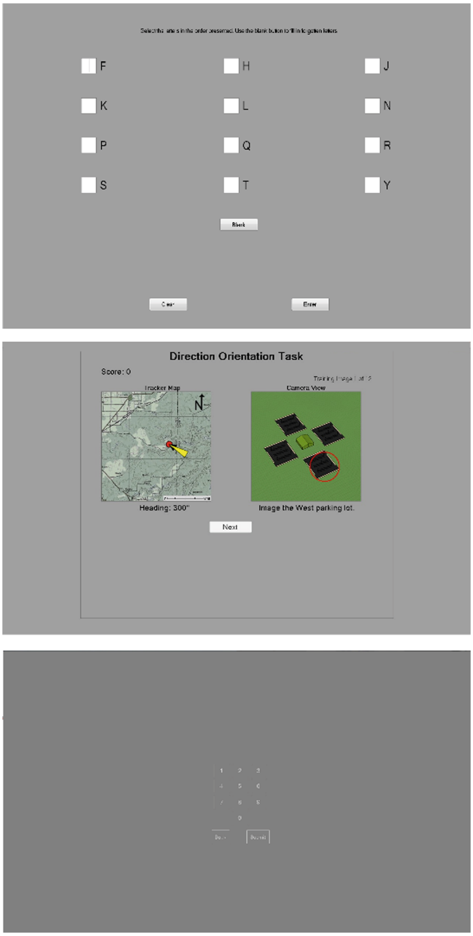
Fig. 2. Gazepoint GP3 setup screen showing a user correctly positioned in the camera's view and seated at an appropriate distance (Color figure online)

calibration. Participants were told to re-calibrate if their results were poor. Once calibration was successful participants were asked to be aware of body position relative to the tracker, however they were not reminded throughout experimentation.

Tasks. After calibration, participants were instructed to put headphones on and then engaged in three consecutive tasks in the following order: Operation Span (OSPAN), Direction Orientation Task (DOT), and Digit-Span Task. See: [5, 8, 9]; respectively, for comprehensive descriptions of these tasks. Most importantly, each of the three tasks required the participant to focus his/her attention in the center of the screen and all input was provided by mouse clicks on the screen, so participants did not have to divert visual attention away from the screen, to the keyboard. Each task took a variable length of time to complete, depending on how quickly participants input their responses: approximately 15 min for OSPAN; 6 min for DOT; and 14 min for Digit-Span. See Fig. 3 for screen grabs of the response screen for each task. All three tasks had a limited area in which relevant information was displayed and for purposes of this paper are considered to be low in visual complexity.

2.2 Results

As previously mentioned, the data presented here will only address data loss. Table 1 shows the proportion of point-of-gaze quality samples that Gazepoint marked as invalid. The correlation between pupil and point-of-gaze quality was very high, but point-of-gaze quality was used for analysis, since it is the slightly more conservative figure. Overall data loss represents the percentage of data where valid data from both eyes were not available. Note the high variance in the average data loss percentages, showing that some participants suffered much higher loss and others did much better.



Select the letters in the order presented. Use the blank button to fill in forgotten letters.

Fig. 3. Screen grabs from the OSPAN (top), DOT (middle) and Digit-span (bottom) tasks, demonstrating visual simplicity of each task

Table 1. Percentage of left, right, and overall data loss across tasks

Task	Task duration	Monitor size	Left pupil % data loss		Right pupil % data loss		Overall % data loss	
			Average	St. Dev	Average	St. Dev	Average	St. Dev
OSPAN	~15 min	17″	20.59	20.79	22.07	21.50	23.20	20.40
DOT	~6 min	17″	22.69	27.14	29.11	29.19	31.60	27.40
Digit-Span	~14 min	17″	30.88	31.35	30.83	30.22	32.80	31.10

3 Experiment 2: Data Loss During Visually Complex Tasks

3.1 Method

Participants. Eye tracking data was collected from 19 participants (18 male, 1 female) who were Naval and Marine Corps student pilots, ranging in age from 22 to 29 (*M* = 24.4, *SD* = 2.3). Each experiment had a unique set of participants; no participant took part in multiple experiments.

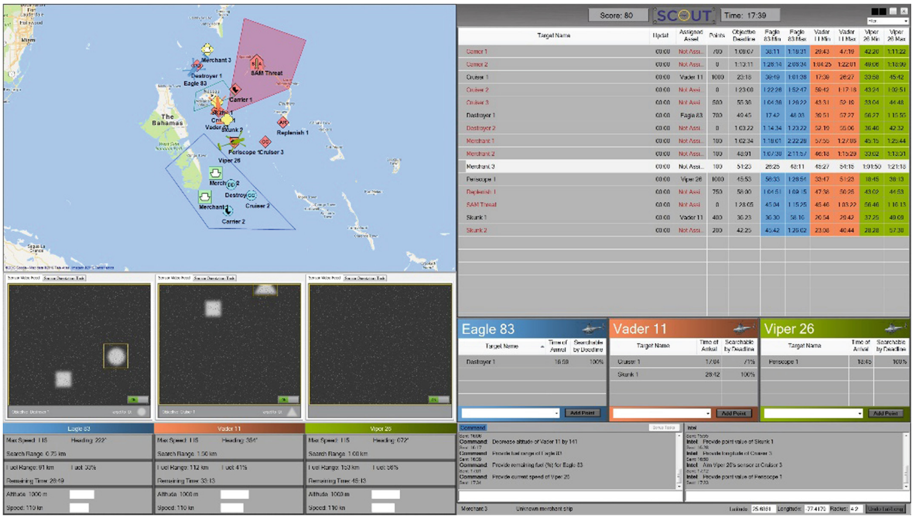
Equipment. The Gazepoint GP3 eye tracking system was again used to collect data from participants with the same set up as Experiment 1, except a 25 inch monitor (2560 × 1440 resolution) was used for this experiment.

Procedure. This experiment took place in a group setting over two sessions. Upon arrival, participants were provided informed consent documents. After giving consent, participants completed a brief demographic survey and then began Gazepoint's set up and calibration process. Participants were given the same instructions for set up and calibration as described in Experiment 1.

After calibration, participants were instructed to put headphones on and then began a self-paced training session, which took approximately 35 min, and instructed them how to interact with the Supervisory Control Operations User Testbed (SCOUT$^{\text{TM}}$). After completing training, participants completed one twelve minute practice mission followed by two thirty-minute missions. Half the participants received one mission scenario first, while the other half received the other first.

Task. The U.S. Naval Research Laboratory developed SCOUT to investigate future challenges operators will experience while managing missions involving multiple autonomous systems. SCOUT contains representative tasks that a future UAV supervisory controller will likely perform, assuming advancements in automation. The Gazepoint GP3 system is integrated with SCOUT in order to gather a more complete understanding of a user's state, including attention allocation, mental workload, and situation awareness, throughout a mission. SCOUT is available in a dual or single screen version, but the single screen version (see Fig. 4) was used in this data collection since the Gazepoint system does not yet reliably support use with multiple screens. See [10] for an overview of SCOUT functionality.

Throughout a mission the participant's primary responsibility was to determine how to dynamically assign unmanned assets to different objectives. Specifically, operators had to decide where to send each of their three UAVs to search for targets with different priority levels, uncertainty and deadlines. In addition, operators had to respond to requests for information and commands by typing in chat boxes. Finally, they had to monitor and click on sensor feeds when potential targets were present, and request access if they needed to fly through restricted airspace. Participants gained points for finding targets and providing timely and accurate information, and lost points for violating restricted airspace and missing potential targets on the sensor feeds. All tasking was driven by pre-scripted scenario files. See [11] for more information on research completed in SCOUT.



Target Name	Updat	Assigned Asset	Points	Objective Deadline	Eagle 83 Min	Eagle 83 Max	Vader 11 Min	Vader 11 Max	Viper 26 Min	Viper 26 Max
Carrier 1	00:00	Not Assi...	700	1:09:07	38:11	1:18:31	29:43	47:19	42:20	1:11:22
Carrier 2	00:00	Not Assi...	0	1:13:11	1:28:14	2:08:34	1:04:25	1:22:01	48:08	1:18:09
Cruiser 1	00:00	Vader 11	1000	23:18	39:49	1:01:38	17:39	28:27	33:58	45:42
Cruiser 2	00:00	Not Assi...	0	1:23:00	1:22:28	1:52:47	58:42	1:17:18	43:24	1:02:51
Cruiser 3	00:00	Not Assi...	500	55:38	1:04:38	1:28:22	43:31	52:19	33:04	44:48
Destroyer 1	00:00	Eagle 83	700	40:45	17:42	48:03	39:51	57:27	56:27	1:15:55
Destroyer 2	00:00	Not Assi...	0	1:03:22	1:14:34	1:23:22	52:19	55:06	36:40	42:32
Merchant 1	00:00	Not Assi...	100	1:02:34	1:18:01	2:22:28	57:55	1:27:05	45:15	1:25:44
Merchant 2	00:00	Not Assi...	100	48:01	1:07:30	2:11:57	46:18	1:15:29	33:02	1:13:01
Merchant 3	00:00	Not Assi...	100	51:23	28:25	48:11	45:27	54:15	1:01:50	1:21:18
Periscope 1	00:00	Viper 26	1000	45:53	58:33	1:28:54	33:47	51:23	18:45	38:13
Replenish 1	00:00	Not Assi...	750	58:00	1:04:51	1:09:15	47:38	50:25	43:02	44:53
SAM Threat	00:00	Not Assi...	0	1:28:05	45:04	1:15:25	45:46	1:03:22	56:46	1:16:13
Skunk 1	00:00	Vader 11	400	36:23	36:30	58:16	20:54	29:42	37:25	49:09
Skunk 2	00:00	Not Assi...	200	42:25	45:42	1:26:02	23:08	40:44	28:28	57:30

Eagle 83

Target Name	Time of Arrival	Searchable by Deadline
Destroyer 1	16:59	100%

Vader 11

Target Name	Time of Arrival	Searchable by Deadline
Cruiser 1	17:04	71%
Skunk 1	28:42	100%

Viper 26

Target Name	Time of Arrival	Searchable by Deadline
Periscope 1	18:45	100%

Eagle 83
Max Speed: 115 Heading: 222°
Search Range: 0.75 km
Fuel Range: 91 km Fuel: 33%
Remaining Time: 28:49
Altitude: 1000 m
Speed: 110 kn

Vader 11
Max Speed: 115 Heading: 354°
Search Range: 1.50 km
Fuel Range: 112 km Fuel: 41%
Remaining Time: 33:13
Altitude: 1000 m
Speed: 110 kn

Viper 26
Max Speed: 115 Heading: 072°
Search Range: 1.00 km
Fuel Range: 153 km Fuel: 56%
Remaining Time: 45:13
Altitude: 1000 m
Speed: 110 kn

Command
Command Decrease altitude of Vader 11 by 141
Command Provide fuel range of Eagle 83
Command Provide remaining fuel (%) for Eagle 83
Command Provide current speed of Viper 26

Intel
Intel Provide point value of Skunk 1
Intel Provide longitude of Cruiser 3
Intel Aim Viper 26's sensor at Cruiser 3
Intel Provide point value of Periscope 1

Merchant 3 Unknown merchant ship Latitude: 25.8381 Longitude: -77.4170 Radius: 4.2

Fig. 4. Interface of single screen SCOUT on a 25 inch, 2560 × 1440 resolution monitor

3.2 Results

This analysis will focus on data from the two thirty minute mission scenarios, and not training or the practice scenario, since this is the data that would be of most interest for all other analyses. Table 2 presents the average proportion of data loss that occurred across all participants, broken out by the order in which they completed the SCOUT mission scenario and by eye. Again, overall data loss represents the percentage of data where valid data from both eyes were not available.

Table 2. Percentage of left, right, and overall data loss across tasks

SCOUT mission order	Task duration	Monitor size	Left pupil % data loss		Right pupil % data loss		Overall % data loss	
			Average	St. Dev	Average	St. Dev	Average	St. Dev
#1	30 min	25″	50.4	26.5	46.1	26.7	58.5	27.5
#2	30 min	25″	50.2	28.4	50.1	28.4	57.3	28.9

Figure 5 decomposes this data further into one minute increments, in order to consider the impact of data loss over time. Here, the percentage of good quality data (both eyes are being tracked), as opposed to data loss (in Table 2) is shown across the two 30-min SCOUT missions.

Figure 6 shows a representative sample of data from eight participants, binned into one minute increments, and across the two SCOUT mission scenarios. One can observe that there are many instances where the data drops out for long periods of time and sometimes reemerges for periods of time. There is also large variability across

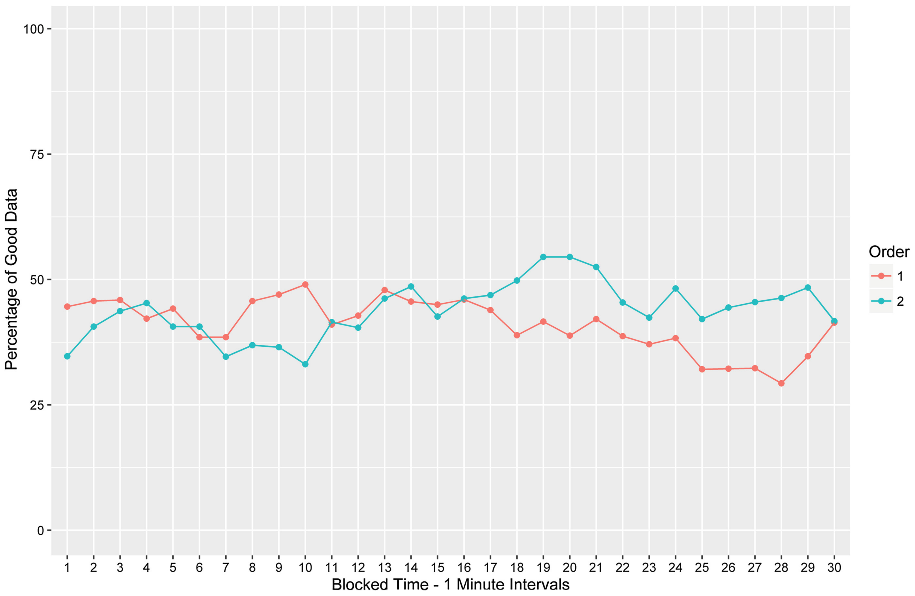


Fig. 5. Percentage of valid data each minute during the first and second SCOUT missions

individuals and does not appear to be an effect of time, where quality either improves or deteriorates. We hypothesized these large fluctuations were attributable to shifts in body position, outside of the head box of the eye tracker, which will be addressed in Experiment 3.

4 Experiment 3: Data Loss Mitigation Technique

4.1 Method

Participants. Eye tracking data was collected from 41 participants (40 male, 1 female) who were Naval and Marine Corps student pilots. They ranged in age from 22 to 29 (*M* = 24.20, *SD* = 2.03).

Equipment. All equipment was the same as described in Experiment 2.

Procedure. Data collection took place in a group setting, over four sessions. After giving consent, participants completed a brief demographic survey and then began Gazepoint's set up and calibration process, which was the same as in Experiment 1 and 2. Afterwards, participants put their headphones on and performed two brief baseline assessments, to assess accuracy and precision of the eye tracker and measure individual pupil size responses, both of which lasted only a few minutes. Next, participants completed an approximately five minute long digit-span task. This task was similar to the digit-span task run in Experiment 1, but included fewer trials. Following the digit-span task, participants completed the self-paced SCOUT training followed by the

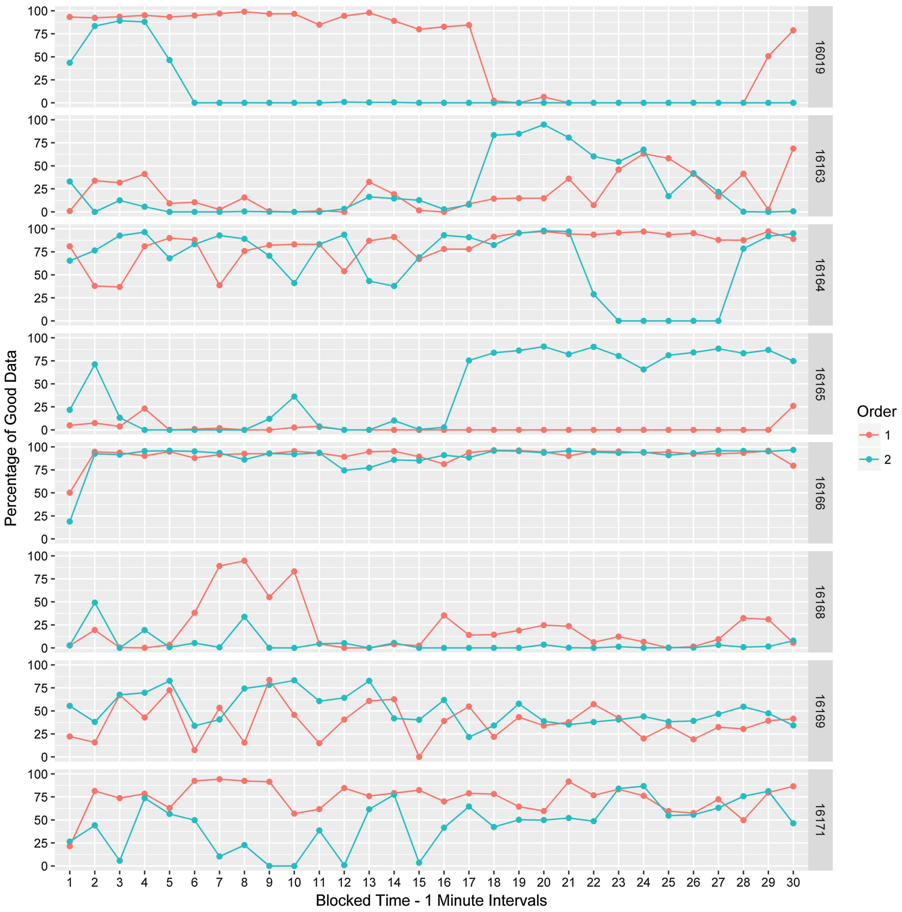


Fig. 6. Sample of eight participants' percentage of valid data, by minute, across the first and second SCOUT missions

twelve minute practice mission and one thirty-minute SCOUT mission. Afterwards they completed the baseline and digit-span tasks a second time.

Data Quality Checks. The SCOUT mission that participants experienced in Experiment 3 was different in one respect from Experiment 2: it included an additional eye tracking data quality check. These quality checks were appended to five pre-scripted workload freeze probes, which took place at approximately 6–7 min intervals throughout the SCOUT mission. Specifically, quality checks took place at the following mission clock times: 1:12, 7:25, 13:47, 20:28, 28:31. Quality checks comprised a position and accuracy assessment.

Position Assessment. Figure 7 shows the screen which participants encountered during the position assessment. Here, participants were instructed to position themselves in their chair so that the two green clusters of dots, which were being drawn on the screen in real time, fell within the bounds of the inner green rectangle. These dots

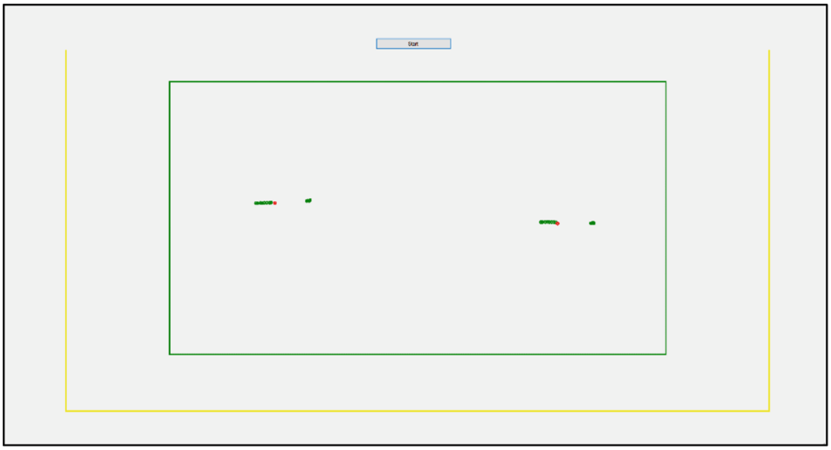


Fig. 7. Position assessment screen prior to recording, with a user's eyes positioned correctly within the green rectangle (Color figure online)

corresponded with the position of the eyes and essentially ensured that the user was positioned at an appropriate distance from the tracker, and that they were centered with respect to the tracker. Once participants' data fell within the green box, they were told to hit the start button, which would attempt to collect 300 samples, or 5 s, of continuous data, of which 75% or more had to be valid data for both eyes. The task continued until the eye tracker was able to collect the 5 s of good data.

Accuracy Assessment. After completing the position task, participants began the accuracy assessment task. Here, participants were instructed to look at the center of the target, shown in Fig. 8, and press start when they were ready for recording to begin.

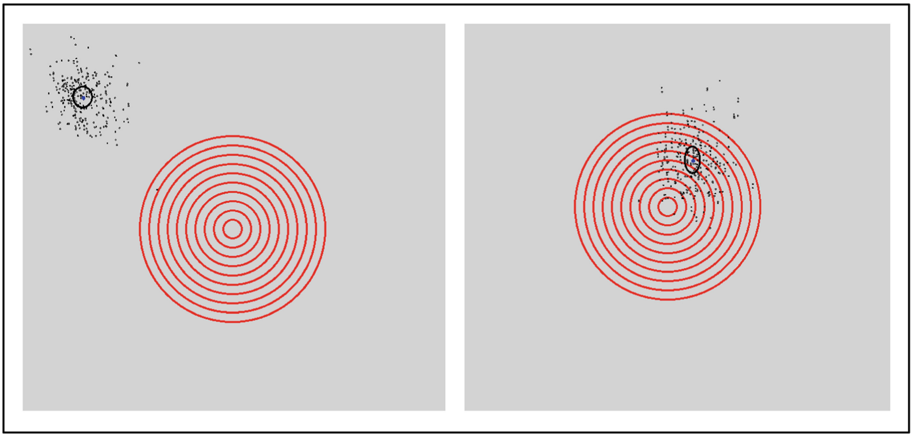


Fig. 8. Accuracy assessment screen showing two examples. Left image shows poor accuracy where the participant is told to consider recalibrating or redo the test. Right image shows acceptable accuracy and the user is told to continue. (Color figure online)

The recording collects 120 samples, or 2 s, of data which are rendered in black in real time on the red target. After data collection was complete, the average gaze location and one standard deviation of error were drawn as an ellipse on the screen. The participant was then informed either that their data accuracy was good and they may continue to SCOUT, or if accuracy was poor the user was given the option to either recalibrate the eye tracker or repeat the assessment.

4.2 Results

The data presented here will focus on the one thirty minute mission scenario, as in Experiment 2, but also includes data from both digit-span tasks as a point of reference and comparison to Experiment 1. Table 3 presents the average proportion of data loss that occurred across all participants. Note that the digit-span task did not include any

Table 3. Percentage of left, right, and overall data loss for experiment 1 and 2

Task	Task duration	Monitor size	Left pupil % data loss		Right pupil % data loss		Overall % data loss	
			Average	St. Dev	Average	St. Dev	Average	St. Dev
Digit-Span	~10 min	25″	27.10	21.20	27.60	21.10	31.90	22.40
SCOUT	30 min	25″	21.90	22.40	21.90	22.20	26.70	23.20

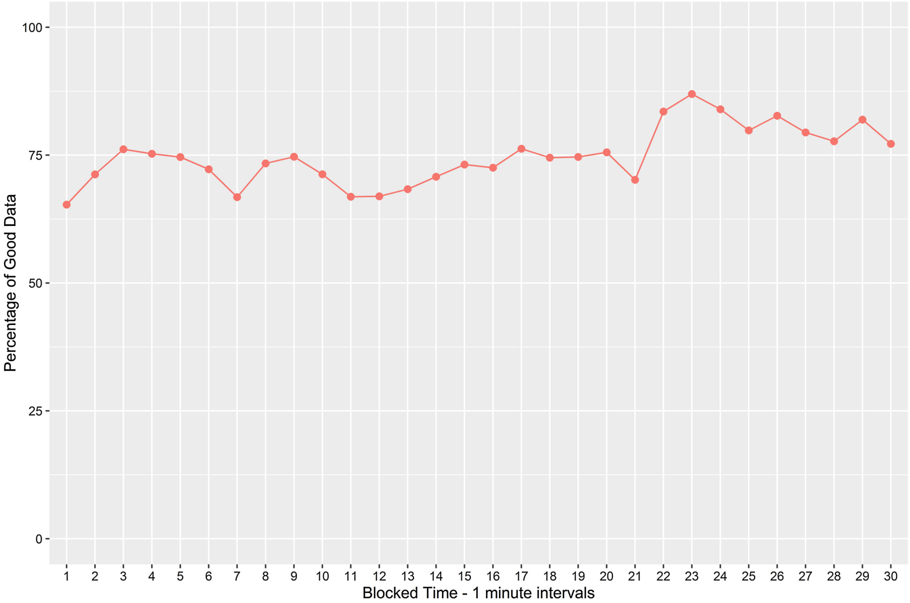


Fig. 9. Percentage of valid data each minute during the SCOUT mission

data quality checks and that the percentage of overall loss for this set of participants was within one percentage point of the group from Experiment 1. This suggests that there was nothing unique about this set of participants that could have resulted in better eye tracking data. Additionally, the increase in monitor size did not impact the data. Also note the large improvement in the percentage of data loss during the SCOUT scenario: down from approximately 58% in Experiment 2 to approximately 27% here.

Figure 9 presents the aggregate SCOUT data in one minute increments, again, to consider the impact of data loss over time. Here, the percentage of good quality data (both eyes are being tracked), is shown across the 30-min SCOUT mission.

Figure 10 shows a sample of data from eight participants, binned into one minute increments, and across the SCOUT mission scenario. One can observe a large improvement in the data quality for most participants, however, participant 170120 still experienced immense data loss. This general improvement in the data quality of participants can also be seen in the smaller standard deviation in data loss from

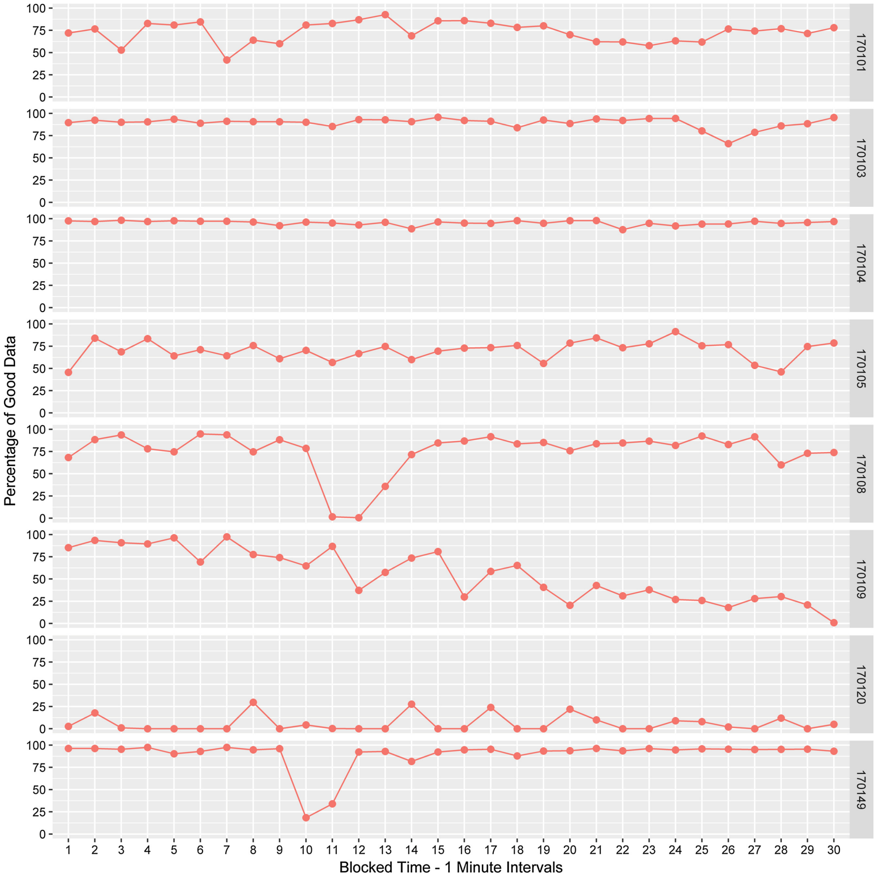


Fig. 10. Sample of data from eight participants, binned in one minute increments, across the SCOUT mission

Experiment 2 to Experiment 3 (~28.90 to 23.20). Figure 11 shows the variability in data quality across all participants for the entire SCOUT mission. Note two individuals who had close to zero percent good quality data.

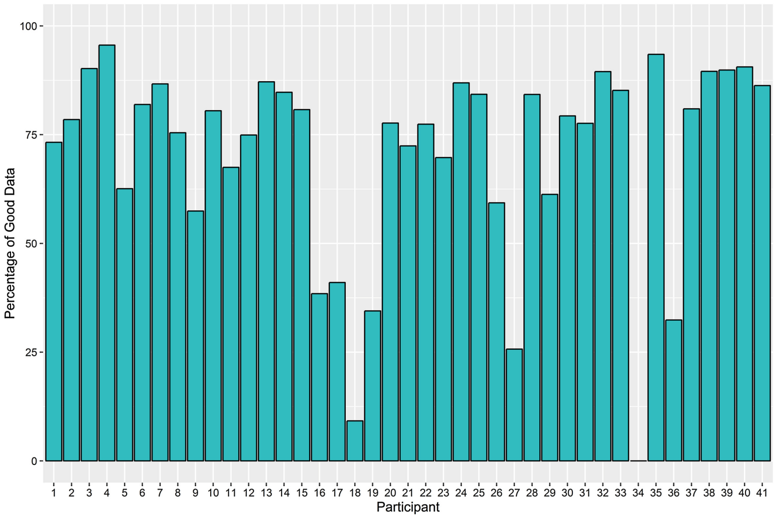


Fig. 11. Percentage of good data by participant across the SCOUT mission

5 Overall Results

For ease of comparison across the three experiments, Table 4 shows the overall percentage of data loss for each experiment, by task.

Table 4. Percentage of left, right, and overall data loss across tasks

Study	# Participants	Task	Task duration	Monitor size	Overall % data loss	
					Average	St. Dev
Experiment 1	24	OSPAN	~15 min	17″	23.20	20.40
Experiment 1	24	DOT	~6 min	17″	31.60	27.40
Experiment 1	24	Digit-Span	~14 min	17″	32.80	31.10
Experiment 2	19	SCOUT #1	30 min	25″	58.50	27.50
Experiment 2	19	SCOUT #2	30 min	25″	57.30	28.90
Experiment 3	41	Digit-Span	~10 min	25″	31.90	22.40
Experiment 3	41	SCOUT	30 min	25″	26.70	23.20

6 Discussion

Data collected from Experiments 1, 2 and 3 provided information on data loss from the Gazepoint GP3 system across a range of tasks. Tasking varied in both duration and visual complexity, requiring participants to focus primarily in the center of the screen or spread attention across the entire screen. Tasks in Experiment 1 were visually simple, while the SCOUT environment required participants to actively scan the entire display. Analysis from Experiments 1 and 2 revealed a significantly higher rate of data loss in the visually complex experiments. Furthermore, and contrary to initial assumptions, data quality did not systematically degrade over time. This finding suggests that visual complexity, rather than task duration, has a larger impact on data quality for tasks under an hour in length. The requirement to scan large areas of a screen likely perturbed participants' body positions with respect to the eye trackers, causing participants to fall outside the bounds eye tracker's head box. This finding motivated the inclusion of a data quality check in Experiment 3.

Utilizing an eye tracking data quality check at approximately 7 min intervals throughout the SCOUT scenario drastically improved the quality of data collected in Experiment 3, as compared to Experiment 2. Comparison of Figs. 5 and 9 show that the data at the beginning of the SCOUT mission in Experiment 3 was, on average, of higher quality than compared to Experiment 2. This is likely attributable to a quality check being presented during Experiment 3's practice scenario, which participants completed before the thirty minute SCOUT mission. The quality check successfully helped mitigate data loss, although it is not clear whether this was due to a greater awareness of maintaining appropriate body position, or whether it helped simply regain the appropriate position. Future studies will investigate use of a quality check which is triggered by a period of poor data quality, instead of utilizing pre-planned checks at specific time increments, even if data quality is high.

These results have widespread implications for researchers interested in utilizing eye tracking technologies for research. Although low-cost eye tracking systems are fast and easy to set up and use, the amount of data loss can be high if not carefully monitored and remediated (e.g., even simple solutions, such as using non-reclining chairs without wheels can have a large impact on data). Furthermore, the data loss was not uniformly distributed across time within participants; participants generally had lengthy periods of good data interspersed with lengthy periods of bad data. If, for example, half the data were present each minute, this might not be as problematic for some analyses, however, when several minutes of data is missing, it is highly questionable to employ techniques to deal with dropped data, such as linear interpolation. Therefore, the authors suggest researchers consider using a data quality check during experimentation. In addition, we suggest utilizing stringent cut-offs for determining inclusion of data, and consider each individual's data independently before determining whether it is appropriate to use for specific analysis purposes. Additionally, data loss figures should be presented for other researchers to assess.

In order for eye tracking to be an effective tool for research, it must be possible to employ in a truly unobtrusive manner and not inadvertently become a focal point of an experiment, which may add confounds. Future research will investigate how to further improve data quality in the least invasive manner possible. Overall, these results add to the corpus of literature showing that low-cost eye tracking has great promise for use in human subject experiments, but that data quality should also be carefully considered.

References

1. Constine, J.: Oculus acquires eye-tracking startup The Eye Tribe. Tech Crunch, December 2016. https://techcrunch.com/2016/12/28/the-eye-tribe-oculus/
2. Ooms, K., et al.: Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental set-ups. J. Eye Mov. Res. **8**, 1–24 (2015)
3. Popelka, S., et al.: EyeTribe tracker data accuracy evaluation and its interconnection with hypothesis software for cartographic purposes. Comput. Intell. Neurosci. **2016**, 9172506 (2016)
4. Dalmaijer, E.: Is the low-cost EyeTribe eye tracker any good for research? PeerJ PrePrints (2014)
5. Coyne, J., Sibley, C.: Investigating the use of two low cost eye tracking systems for detecting pupillary response to changes in mental workload. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications (2016)
6. Funke, G., et al.: Which eye tracker is right for your research? Performance evaluation of several cost variant eye trackers. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications (2016)
7. Gazepoint, Open Gaze API, in Version 2.0. http://www.gazept.com/dl/Gazepoint_API_v2.0.pdf
8. Turner, M.L., Engle, R.W.: Is working memory capacity task dependent? J. Mem. Lang. **28** (2), 127–154 (1989)
9. Ostoin, S.D.: An Assessment of the Performance-Based Measurement Battery (PBMB), the Navy's Psychomotor Supplement to the Aviation Selection Test Battery (ASTB). DTIC Document (2007)
10. Sibley, C., Coyne, J., Thomas, J.: Demonstrating the supervisory control operations user testbed (SCOUT). In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications (2016)
11. Sibley, C., Coyne, J., Avvari, G.V., Mishra, M., Pattipati, K.R.: Supporting multi-objective decision making within a supervisory control environment. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) AC 2016. LNCS (LNAI), vol. 9744, pp. 210–221. Springer, Cham (2016). doi:10.1007/978-3-319-39952-2_21