SynchronEyes: A Novel, Paired Data Set of Eye Movements Recorded Simultaneously with Remote and Wearable Eye-Tracking Devices

Samantha Aziz
Dillon J Lohr
Oleg Komogortsev
sda69@txstate.edu
djl70@txstate.edu
ok11@txstate.edu
Texas State University
San Marcos, Texas, USA

ABSTRACT

Comparing the performance of new eye-tracking devices against an established benchmark is vital for identifying differences in the way eye movements are reported by each device. This paper introduces a new paired data set comprised of eye movement recordings captured simultaneously with both the EyeLink 1000—considered the "gold standard" in eye-tracking research studies—and the recently released AdHawk MindLink eye tracker. Our work presents a methodology for simultaneous data collection and a comparison of the resulting eye-tracking signal quality achieved by each device.

CCS CONCEPTS

• **Hardware** → **Hardware reliability screening**; • **Human-centered computing** → **Ubiquitous and mobile computing design and evaluation methods**.

KEYWORDS

eye tracking, paired data, data set, data quality, mindlink

1 INTRODUCTION

Eye movement signals captured by an eye-tracking device are influenced by a large number of factors, including device-specific gaze estimation artifacts [Holmqvist and Blignaut 2020], a user's physical attributes (e.g., age, eye shape, iris color [Holmqvist 2017]), a user's mental state (e.g., fatigue [Zargari Marandi et al. 2018]),

a user's health history (e.g., schizophrenia [Holzman et al. 1973], brain injury [Samadani et al. 2015]), and miscellaneous environmental noise.

Due to the sheer quantity of these influential variables, recording the same person twice in quick succession with the same eye-tracking device as he/she performs the same eye-tracking task can result in noticeably different characteristics in the captured eye-tracking signals (e.g., overlapping saccades caused by the onset of eye fatigue [Bahill and Stark 1975] may be present in the second recording but not the first). When analyzing eye-tracking signal quality and comparing different eye-tracking devices, it can be difficult to disentangle these sources of variation in the eye-tracking signals. Not accounting for these factors may complicate the comparison of eye-tracking signal quality measures computed on different data sets with different populations.

Most of the above factors can be controlled for by capturing eye-tracking signals with two devices simultaneously. In theory, any differences between the captured eye-tracking signals could be solely attributed to differences in each device's gaze estimation pipeline. An eye movement data set collected concurrently on two devices–described herein as a *paired data set*–can therefore be used to benchmark the performance of new eye trackers relative to a well-established counterpart. This approach offers practical advantages over using artificial eyes, which bear limited resemblance to human eye movement [Reingold 2014]. Paired eye movement data sets can support hypothesized disparities in the way different eye trackers report the same eye movement ([Ehinger et al. 2019; Holmqvist et al. 2020; McCamy et al. 2015]) or demonstrate the sufficiency of eye-tracking devices for specific research uses ([McCamy et al. 2015; Titz et al. 2018]).

There are other potential applications for a paired data set aside from more rigorous eye-tracking signal quality comparisons. Deep learning techniques have recently started being applied to eye-tracking signals for various purposes including user authentication [Makowski et al. 2021] and basic eye movement type classification [Zemblys et al. 2019]. A major concern with such approaches is the problem of *domain adaptation*, or how well they can generalize to eye-tracking signals collected with different eye-tracking devices than were used to train a given model. Having access to a paired data set may help when assessing the generalizability of a model to

different eye-tracking devices by controlling for most sources of variability between the domains.

In this work, we introduce SynchronEyes, a novel, publicly available, paired eye-tracking data set collected simultaneously with the EyeLink 1000 (SR Research, Ottawa, Ontario, CA) at 1000 Hz and the AdHawk MindLink (AdHawk Microsystems, Waterloo, Ontario, CA) at 500 Hz. Using our paired data set, we present preliminary evaluations of eye-tracking signal quality for both devices to support the data set's usability. Unlike other paired eye-tracking data sets (e.g., [Ehinger et al. 2019; Titz et al. 2018]), we constructed SynchronEyes to enable paired data investigations beyond signal quality analysis by capturing eye movement during a variety of tasks. Moreover, we are the first to assess the eye-tracking signal quality of the AdHawk MindLink in isolation. SynchronEyes is available for download at https://doi.org/10.18738/T8/YI1ISX.

2 METHODOLOGY

2.1 Hardware

2.1.1 EyeLink 1000. The EyeLink 1000 (SR Research, Ottawa, Ontario, CA), hereon referred to as the "EyeLink," is considered the "gold standard" for eye-tracking research. It uses video oculography (VOG) to estimate gaze via pupil-corneal reflection (P-CR) tracking. The EyeLink can be employed in several different configurations and can track either monocularly or binocularly. To enable synchronous data collection with a second eye-tracking device, we used the tower-mounted setup and captured monocular (left eye) data at a sampling rate of 1000 Hz. Digital filtering of the eye-tracking signal was disabled via the EyeLink software.

2.1.2 AdHawk MindLink. The AdHawk MindLink (AdHawk Microsystems, Waterloo, Ontario, CA), hereon referred to as the "MindLink," is a camera-free eye tracker embedded within an eyeglass frame. Rather than using VOG to estimate gaze, it uses a microelectromechanical system (MEMS) device to beam infrared (IR) light into the eye and track the resulting corneal reflections [Sarkar. 2020]. The MindLink can track the eyes binocularly at various sampling rates. The highest available sampling rate of 500 Hz was used in the present study.

2.1.3 Interference Between Devices. Both devices employed in this study rely on tracking corneal reflections using IR light. While the MindLink is designed to operate in the presence of multiple light sources, the EyeLink assumes the presence of exactly one IR source. Tracking the same eye with both devices simultaneously produces an unusually noisy signal in the EyeLink data, as illustrated in Fig. 1. This noise represents excessive interference from the MindLink, as the EyeLink's gaze estimation pipeline assumes the presence of a single IR source. Conversely, the additional IR light from the EyeLink did not appear to introduce additional noise in the MindLink's signal trace.

While we would have preferred to record the same eye with both eye-tracking devices, we instead recorded each eye with a different device to preserve eye-tracking signal quality while collecting paired data. This experimental setup does not have a significantly detrimental effect to the overall usability of the dataset, as eye movement patterns are strongly correlated between individual eyes for people with normal binocular perception [Reynaud and Hess
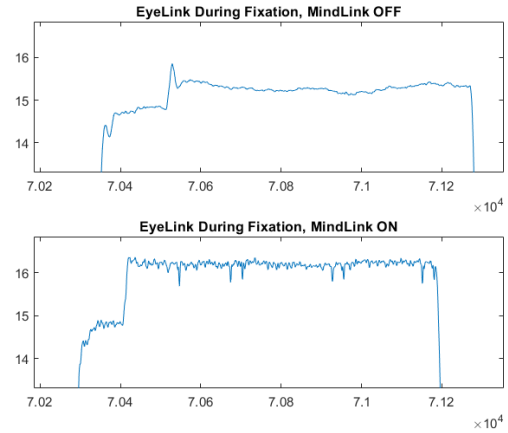

Figure 1: Top: An example of the additional noise in the Eye-Link gaze signal during simultaneous operation of the Mind-Link.
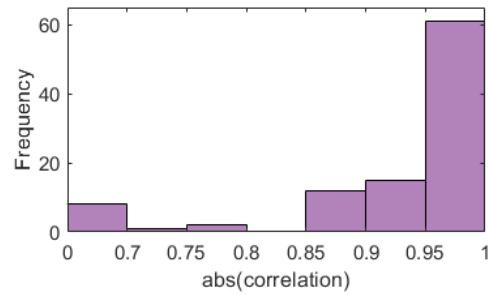

Figure 2: The distribution of Pearson correlations between the gaze positions of the left and right eyes in the MindLink across all binocular SynchronEyes recordings used in this investigation (n = 100).

2018]. We demonstrate that our data set exhibits this phenomenon by calculating the Pearson correlation coefficient of gaze positions between the left and right eyes in our binocular MindLink data. As shown in Fig. 2, most binocular recordings in our data set are indeed strongly correlated between the eyes. In a paired data setting, we assume that the data captured simultaneously by the MindLink and the EyeLink on separate eyes would similarly be strongly correlated.

To minimize interference from the MindLink on the EyeLink, we crafted an IR light blocker (illustrated in Fig. 3) using a combination of aluminum foil to block the IR light, a paper napkin to prevent sensor damage, and masking tape to hold it together. When recording data with both devices simultaneously, this IR light blocker was taped over the left eye sensor of the MindLink, napkin-side down. The MindLink is consequently unable to track the left eye while the IR light blocker is in place. This augmentation did not impact the user's ability to view the stimulus.

Figure 3: An illustration of the IR light blockers used to minimize inter-device interference.

2.2 Participants

We recruited 20 college-aged participants (10 female, 10 male, mean age 23.5, age range 21–41) from the student population of Texas State University. An additional 5 participants were excluded from this study due to excessively noisy data in one or more devices. All participants gave their informed, written consent to participate in the study and were each reimbursed $40 upon their completion of the experiment. The study was approved by Texas State University Instructional Review Board (IRB).

10 participants required vision correction; 7 glasses-wearing participants removed their glasses, while 3 contact-wearing participants wore contact lenses while participating. All participants were able to see the stimuli presented on the screen without additional vision correction. Participants were seated 55 cm away from a 22-inch Panasonic monitor with a pixel resolution of 1680x1050 (474 mm x 297 mm). Participants' heads were stabilized by a forehead and chin rest at all times to minimize head movement.

2.3 Data Collection

Eye movement data was collected over 2 recording sessions separated by a 5-minute break, totaling around 60 minutes. In Session 1, binocular gaze data was captured by the MindLink while the EyeLink was powered off. In Session 2, monocular gaze data from the left and right eyes were captured by the EyeLink and the MindLink, respectively. We blocked IR light from the MindLink's left eye sensor using custom IR light blockers (Fig. 3) to minimize interference between the devices. An overview of the experimental setup is illustrated in Fig. 4.

Each session consists of a battery of 5 eye-tracking tasks designed to elicit a variety of eye movements. The tasks are identical in form to the same tasks from the GazeBase [Griffith et al. 2021] data set. A concise description of each task is listed below in the order they were presented during a session. For a more comprehensive description of the tasks and their parameters, see [Griffith et al. 2021].

(1) Random Saccades (RAN): A 1° target jumps to random positions across the visual field over 100 seconds.
(2) Reading (TEX): An excerpt of Lewis Carroll's *Hunting of the Snark* is displayed for 60 seconds.
(3) Fixation (FXS): A stationary 1° target remains in the center of the visual field for 15 seconds.
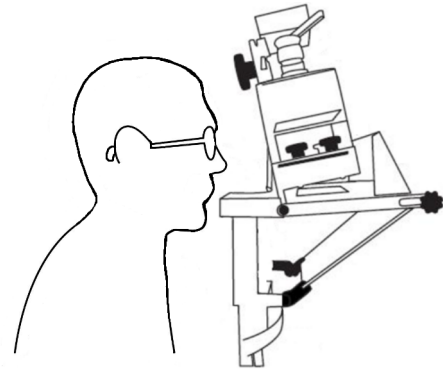
	MindLink	EyeLink
Session 1	Both Eyes	OFF
Session 2	Right Eye	Left Eye

Figure 4: An overview of the experimental setup. (Top) Illustration of a participant seated at the forehead and chin rests while wearing the MindLink. Note that participants' heads were resting on the EyeLink's chin- and forehead- rest. The EyeLink is mounted above the forehead rest. EyeLink 1000 image courtesy of SR Research. (Bottom) Summary of which eyes were tracked by which device for each recording session.

(4) Video (VID): A 60-second excerpt of a trailer for *The Hobbit: The Desolation of Smaug* is presented. In Session 1, the first 60 seconds of the trailer are shown. In Session 2, the next 60 seconds of the trailer are shown.
(5) Horizontal Saccades (HSS): A 1° target alternates between the left and right sides of the visual field over a period of 100 seconds. Each jump of the target spans 30° of the visual angle.

At the start of Session 1, the best-fitting nose piece was chosen for the MindLink from the four provided by the manufacturer to ensure that the device's sensors were positioned appropriately relative to the subject's eyes. The MindLink was then auto-tuned and calibrated using the AdHawk Hub software (version 5.7). In the interest of time, tuning and calibration was not repeated between tasks unless the participant adjusted the MindLink during the session.

At the start of Session 2, the EyeLink was turned on, and the IR light blocker was applied over the left eye sensor of the MindLink. Next, the MindLink was calibrated using the AdHawk Hub software. Then, the EyeLink was calibrated using the EyeLink software. In the interest of time, calibration was not repeated for either device between tasks unless the participant adjusted the MindLink or made large head movements during the session.

2.4 Temporal Synchronization Across Devices

Both the MindLink and the EyeLink report timestamps relative to device startup time. In order to synchronize timestamps across devices, we performed automatic, post-hoc temporal synchronization of the paired Session 2 data. After converting the raw timestamps of

each device into milliseconds elapsed since the first reported times-tamp, the gaze signals were automatically aligned by temporally shifting the MindLink signal backward until the mean squared error between the signals was minimized. Alignment quality was then manually verified–no post-alignment corrections were necessary.

3 SIGNAL QUALITY ANALYSIS

To evaluate the efficacy of our setup, we estimated the average spatial accuracy and spatial precision for 3 disjoint subsets of the data set: Session 1 MindLink data, Session 2 MindLink data, and Session 2 EyeLink data. The latter two subsets comprise the paired portion of the data set.

The signal quality investigation described herein uses data from the RAN task wherein participants fixated on a bulls-eye target that appeared in random positions across a 30° x 18° field of view. The target remained fixed at each location for 1 second before jumping to the next location. The minimum distance between successive target locations was 2°. The trajectory of the target was randomized between participants and kept consistent across sessions.

3.1 Data Processing

We first converted the raw gaze data into degrees of visual angle. We transformed raw gaze samples from the EyeLink into degrees of visual angle using the following MATLAB functions:

$$x_{dva-el} = \text{atan2d}\left(x_{px} - \frac{w_{px}}{2} - \frac{w_{px}}{w_{mm}}\hat{x}_{mm}, \frac{w_{px}}{w_{mm}}d_{mm}\right) \quad (1)$$

$$y_{dva-el} = \text{atan2d}\left(y_{px} + \frac{h_{px}}{2} - \frac{h_{px}}{h_{mm}}\hat{y}_{mm}, \frac{h_{px}}{h_{mm}}d_{mm}\right) \quad (2)$$

$\hat{x}_{mm}$ and $\hat{y}_{mm}$ are the horizontal and vertical offsets, respectively, of the primary viewing position from the center of the display. In our case, $\hat{x}_{mm} = 0$ mm and $\hat{y}_{mm} = 36$ mm. x_{px} and y_{px} denote the gaze positions reported by the EyeLink in pixels, w_{px} and h_{px} denote the width and height of the stimulus display monitor in pixels, and w_{mm}, h_{mm}, and d_{mm} respectively denote the width, height, and distance of the monitor in millimeters.

The MindLink reports gaze as a three-dimensional vector $v = (v_x, v_y, v_z)$. We transformed raw gaze samples into degrees of visual angle using the following MATLAB functions:

$$x_{dva-ml} = \text{atan2d}\left(v_x, \sqrt{v_y^2 + v_z^2}\right) \quad (3)$$

$$y_{dva-ml} = \text{atan2d}(v_y, -v_z) \quad (4)$$

The MindLink provides separate data streams for the left, right, and cyclopean eye. For the MindLink, we use data from the right eye for both recording sessions. Because our signal quality analysis assumes that target locations are shown relative to the center of the primary viewing position, we shift the MindLink target positions right by 31.5 mm to center the target positions to the right eye (following Fig. 4 in [Lohr et al. 2019]). Because the MindLink is a wearable device, its horizontal and vertical offset from the center of the display differs by subject. To center the MindLink's gaze signal relative to the display, we calculate the average horizontal and vertical offset between the gaze and the target during the FXS task, and shift the RAN data based on this offset.
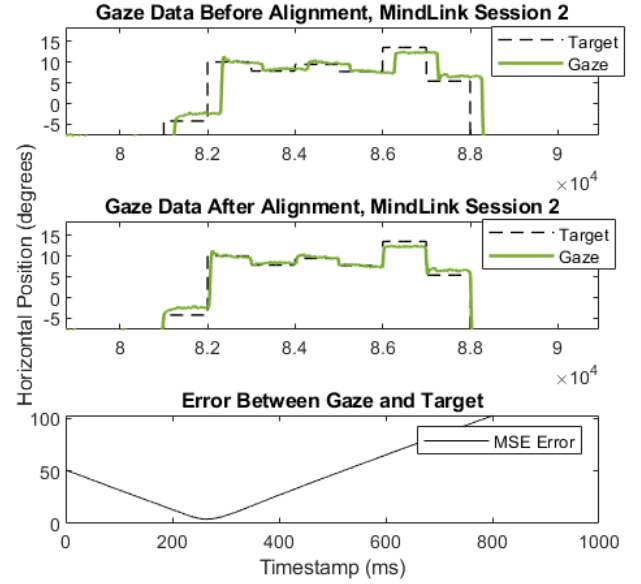

Figure 5: A depiction of the temporal alignment between the captured gaze signal and target position signal, including an illustration of mean squared error as the gaze signal is shifted backward in time. In this case, the gaze signal was shifted 264 ms backward.

We also temporally align all gaze signals to the target position, effectively removing saccade latency (Fig. 5). The optimal alignment is determined by temporally shifting the gaze signal backward until the mean squared error between the gaze and the target signals is minimized.
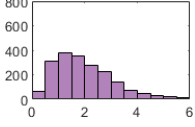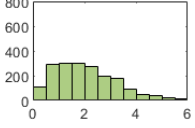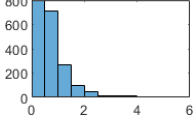
Eye-tracking signal quality metrics should be measured during periods of stable fixation. Fixations were coarsely identified in our signals following an approach proposed by [Lohr et al. 2019]. We first identified timestamps that mark changes in target position, then discarded the first 400 milliseconds that followed to eliminate periods of gaze instability. The following 500 milliseconds were used for eye-tracking signal quality analysis.

3.2 Results

Eye movement signal quality metrics were calculated for each device at their native sampling rates. All metrics are reported separately for the horizontal, vertical, and combined radial dimensions. Spatial accuracy is calculated as the mean angular distance in degrees between the target position and the reported gaze position, as described in [Holmqvist et al. 2012]. We also calculated the sample-to-sample root mean square (RMS) spatial precision described by [Holmqvist et al. 2012]. We report each measure as the median value across all subjects in the data set.

Table 1 shows the spatial accuracy and spatial precision calculated on each subset of data and the distribution of spatial accuracy across all fixations identified in each subset. The distribution for spatial precision was largely unchanged between each subset of the data set and therefore omitted.

Table 1: Eye-tracking signal quality measures for each subset of our data set. The spatial accuracy distributions are shown for the combined (C) dimension.

Device	Dim*	Accuracy (°)	Precision (°)	Acc. Distribution (°)
S1 MindLink	H	1.15	0.16	
	V	1.15	0.19	
	C	1.97	0.17	
S2 MindLink	H	1.05	0.15	
	V	1.30	0.19	
	C	2.09	0.18	
S2 EyeLink	H	0.39	0.07	
	V	0.43	0.07	
	C	0.63	0.09	

* Respectively denotes Horizontal, Vertical, or Combined dimensions.

4 DISCUSSION

Based on the quantitative results presented in Table 1, the MindLink's eye-tracking signal quality level remains consistent across recording sessions. This consistency across sessions is especially notable since the MindLink had one of its eye-tracking sensors blocked during Session 2. The MindLink consistently exhibits worse spatial accuracy and spatial precision than the EyeLink. The difference in spatial precision between devices indicates that the fixation periods identified for signal quality analysis appear to be less stable in the MindLink than in the EyeLink.

We use the GazeBase [Griffith et al. 2021] data set as a baseline for our Session 2 EyeLink eye-tracking signal quality measures. Our measured spatial accuracy is slightly worse than the spatial accuracy reported for GazeBase (approximately 0.5°), but our data nonetheless matches the expected level of signal quality that can be achieved with the device. It is important to reiterate that we disabled digital filtering of our EyeLink data in the EyeLink software, whereas GazeBase did not.

Qualitatively, Fig. 6 shows that the MindLink's eye-tracking data from Session 2 does not appear to be significantly noisier than Session 1, indicating that it experiences minimal interference from the EyeLink. The MindLink data are slightly noisier than the EyeLink data, but fixations and large saccades are still clearly distinguishable in the MindLink. There are a number of notable differences in the way eye movements manifest in the signals captured by each device. Particularly, smaller saccades that are clearly visible in the EyeLink data are often more difficult to distinguish from the surrounding fixations in the MindLink data. This is particularly prevalent in low-amplitude saccades (e.g., the corrective saccades shown in Fig. 6). Additionally, low-amplitude saccades are not always reported as instantaneous movement in the MindLink, but instead appear as smooth transitions between gaze positions. As an aside, Fig. 6 also
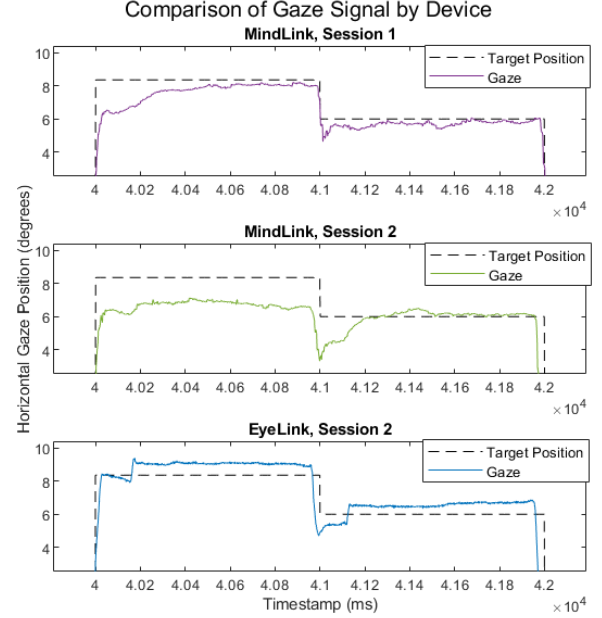

Figure 6: A representative sample of eye-tracking signal from each subset of our data set. Note that the corrective saccades captured by the EyeLink (blue) near timestamps $4.02{\times}10^4$ and $4.10{\times}10^4$ are more difficult to discern as saccades in the Session 2 MindLink data (green), i.e., they appear to be distorted.

showcases the success of our post-hoc temporal synchronization protocol described in Section 2.4.

4.1 Limitations

Although recording each eye separately produces high-quality signals from each device, it fails to completely eliminate variation in the paired gaze data based on differences between individual eyes. We have successfully demonstrated that gaze patterns are highly correlated between individual eyes in our population of healthy young adults, which provides strong evidence that the impact on the usability of our data set is minimal. However, the actual observed impact of our experimental design on future use cases remains unclear. Additionally, while our chosen measure of correlation is a simple and interpretable way to illustrate this correlation, it may fail to capture the complex inter-ocular dynamics that enable binocular vision.

5 CONCLUSION

This paper presents an eye-tracking data set collected simultaneously with two high-quality eye-tracking devices, including the novel MindLink. We describe inter-device interference caused by interactions between each device's IR emitter, and provide a cursory measures of spatial accuracy and precision to demonstrate that our efforts to reduce interference between devices were effective in preserving the eye-tracking signal quality between recording sessions. Future work in this domain will provide more comprehensive analyses of SynchronEyes's eye-tracking signal quality, including the

MindLink's ability to capture small eye movements, by leveraging the full battery of tasks in the data set.

ACKNOWLEDGMENTS

REFERENCES

A.Terry Bahill and Lawrence Stark. 1975. Overlapping saccades and glissades are produced by fatigue in the saccadic eye movement system. *Experimental Neurology* 48, 1 (1975), 95–106. https://doi.org/10.1016/0014-4886(75)90225-3

Benedikt V. Ehinger, Katharina Groß, Inga Ibs, and Peter König. 2019. A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. *PeerJ* 7 (09 Jul 2019), e7086–e7086. https://doi.org/10.7717/peerj.7086 31328028[pmid].

Henry Griffith, Dillon Lohr, Evgeny Abdulin, and Oleg Komogortsev. 2021. GazeBase, a large-scale, multi-stimulus, longitudinal eye movement dataset. *Scientific Data* 8 (07 2021). https://doi.org/10.1038/s41597-021-00959-y

Kenneth Holmqvist. 2017. Common predictors of accuracy, precision and data loss in 12 eye-trackers. https://doi.org/10.13140/RG.2.2.16805.22246

Kenneth Holmqvist and Pieter Blignaut. 2020. Small eye movements cannot be reliably measured by video-based P-CR eye-trackers. *Behavior research methods* 52, 5 (Oct 2020), 2098–2121. https://doi.org/10.3758/s13428-020-01363-x 32206998[pmid].

Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. 2012. Eye tracker data quality: What it is and how to measure it. *Eye Tracking Research and Applications Symposium (ETRA)* (28 Mar 2012). https://doi.org/10.1145/2168556.2168563

Kenneth Holmqvist, Saga Lee Orbom, Michael Miller, Albert Kashchenevsky, Mark M. Shovman, and Mark W. Greenlee. 2020. Validation of a Prototype Hybrid Eye-Tracker against the DPI and the Tobii Spectrum. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers).* Association for Computing Machinery, New York, NY, USA, Article 6, 9 pages. https://doi.org/10.1145/3379155.3391330

Philip S. Holzman, Leonard R. Proctor, and Dominic W. Hughes. 1973. Eye-Tracking Patterns in Schizophrenia. *Science* 181, 4095 (1973), 179–181. https://doi.org/10.1126/science.181.4095.179 arXiv:https://www.science.org/doi/pdf/10.1126/science.181.4095.179

Dillon J. Lohr, Lee Friedman, and Oleg V. Komogortsev. 2019. Evaluating the Data Quality of Eye Tracking Signals from a Virtual Reality System: Case Study using SMI's Eye-Tracking HTC Vive. arXiv:1912.02083 [cs.HC]

Silvia Makowski, Paul Prasse, David R. Reich, Daniel Krakowczyk, Lena A. Jäger, and Tobias Scheffer. 2021. DeepEyedentificationLive: Oculomotoric Biometric Identification and Presentation-Attack Detection Using Deep Neural Networks. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3, 4 (2021), 506–518. https://doi.org/10.1109/TBIOM.2021.3116875

Michael B. McCamy, Jorge Otero-Millan, R. John Leigh, Susan A. King, Rosalyn M. Schneider, Stephen L. Macknik, and Susana Martinez-Conde. 2015. Simultaneous Recordings of Human Microsaccades and Drifts with a Contemporary Video Eye Tracker and the Search Coil Technique. *PLOS ONE* 10, 6 (06 2015), 1–20. https://doi.org/10.1371/journal.pone.0128428

Eyal M. Reingold. 2014. Eye tracking research and technology: Towards objective measurement of data quality. *Visual Cognition* 22, 3-4 (2014), 635–652. https://doi.org/10.1080/13506285.2013.876481 arXiv:https://doi.org/10.1080/13506285.2013.876481 PMID: 24771998.

Alexandre Reynaud and Robert F. Hess. 2018. Interocular correlation sensitivity and its relationship with stereopsis. *Journal of vision* 18, 1 (01 Jan 2018), 11–11. https://doi.org/10.1167/18.1.11 29362804[pmid].

Uzma Samadani, Robert Ritlop, Marleen Reyes, Elena Nehrbass, Meng Li, Elizabeth Lamm, Julia Schneider, David Shimunov, Maria Sava, Radek Kolecki, et al. 2015. Eye tracking detects disconjugate eye movements associated with structural traumatic brain injury and concussion. *Journal of neurotrauma* 32, 8 (2015), 548–556.

Niladri Sarkar. 2020. System and Method for Resonant Eye-Tracking. https://patents.google.com/patent/US20180210547A1/en US20180210547A1.

Johannes Titz, Agnes Scholz, and Peter Sedlmeier. 2018. Comparing eye trackers by correlating their eye-metric data. *Behavior Research Methods* 50, 5 (01 Oct 2018), 1853–1863. https://doi.org/10.3758/s13428-017-0954-y

Ramtin Zargari Marandi, Pascal Madeleine, Øyvind Omland, Nicolas Vuillerme, and Afshin Samani. 2018. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Scientific Reports* 8, 1 (03 Sep 2018), 13148. https://doi.org/10.1038/s41598-018-31577-1

Raimondas Zemblys, Diederick C. Niehorster, and Kenneth Holmqvist. 2019. gazeNet: End-to-end eye-movement event detection with deep neural networks. *Behavior Research Methods* 51, 2 (01 Apr 2019), 840–864. https://doi.org/10.3758/s13428-018-1133-5