A Psychophysics-inspired Model of Gaze Selection Performance

Immo Schuetz*
schuetz.immo@gmail.com
Facebook Reality Labs
Redmond, Washington

T. Scott Murdison
smurdison@fb.com
Facebook Reality Labs
Redmond, Washington

Marina Zannoli
marinazannoli@fb.com
Facebook Reality Labs
Redmond, Washington

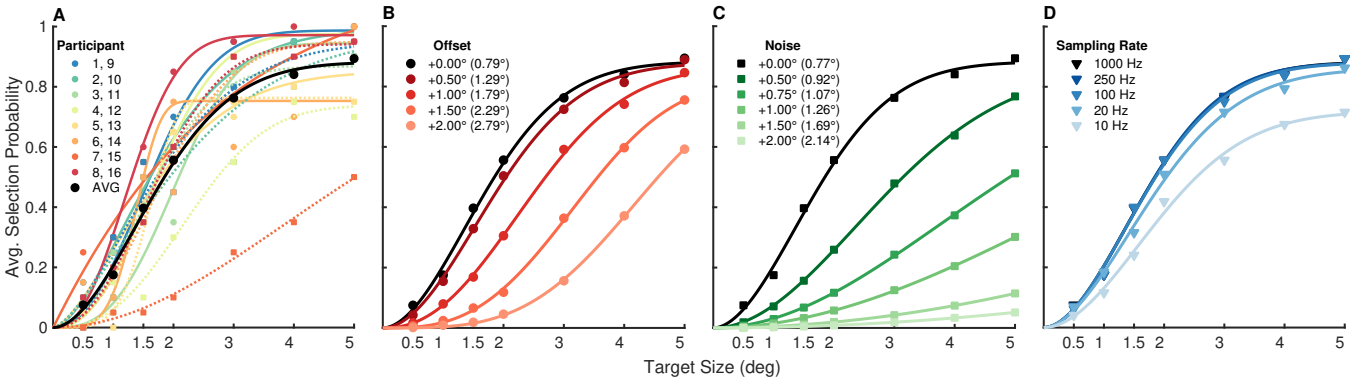

Figure 1: Average selection performance by target size (probability of successful selection). *A:* individual performance and across-participant average, *B:* added spatial offset, *C:* added Gaussian noise, *D:* down-sampled signal. Performance determined at endpoint of first saccade per trial only. Numbers in parentheses indicate final accuracy and precision after manipulation.

ABSTRACT

Eye gaze promises to be a fast and intuitive way of interacting with technology. Importantly, the performance of a gaze selection paradigm depends on the eye tracker used: Higher tracking accuracy allows for selection of smaller targets, and higher precision and sampling rate allow for faster and more robust interaction. Here we present a novel approach to predict the minimal eye tracker specifications required for gaze-based selection. We quantified selection performance for targets of different sizes while recording high-fidelity gaze data. Selection performance across target sizes was well modeled by a sigmoid similar to a psychometric function. We then simulated lower tracker fidelity by adding noise, a constant spatial bias, or temporal sub-sampling of the recorded data while re-fitting the model each time. Our approach can inform design by predicting performance for a given interface element and tracker fidelity or the minimal element size for a specific performance level.

CCS CONCEPTS

• **Human-centered computing** → **User models**; *User studies*; *Pointing devices.*

KEYWORDS

eye tracking, gaze interaction, selection, accuracy, precision, sampling rate

ACM Reference Format:
Immo Schuetz, T. Scott Murdison, and Marina Zannoli. 2020. A Psychophysics-inspired Model of Gaze Selection Performance. In *Symposium on Eye Tracking Research and Applications (ETRA '20 Short Papers), June 2–5, 2020, Stuttgart, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3379156.3391336

1 INTRODUCTION & RELATED WORK

Eye tracking for Human-Computer interaction, explored since the late 1980s [Jacob 1990; Ware and Mikaelian 1987], could enable fast, intuitive and hands-free input [Majaranta and Bulling 2014]. Gaze-driven assistive technologies for persons with disabilities are widely available today [Majaranta et al. 2011]. More recently, increased availability of eye trackers in wearable devices is inspiring novel interaction methods [Blattgerste et al. 2018; Bulling and Gellersen 2010; Majaranta and Bulling 2014; Tanriverdi and Jacob 2000].

However, some challenges remain to be solved to enable widespread and robust gaze interaction. Besides correctly inferring when a selection is to be triggered (the "Midas Touch" problem, [Hansen et al. 2003; Jacob 1990]), a major challenge in gaze-driven interactions is related to the spatial and temporal accuracy and precision of the eye tracking system used [Feit et al. 2017; Holmqvist et al. 2012; McConkie 1981; Reingold 2014]. Not all studies have historically reported data quality measures [Reingold 2014], and signal quality achieved in the real world is often lower than the theoretical performance stated by the device manufacturer [Blignaut et al. 2014; Hansen and Ji 2010].

*Current affiliation: Justus Liebig University, Giessen, Germany

ETRA '20 Short Papers, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7134-6/20/06…$15.00
https://doi.org/10.1145/3379156.3391336

Previous work has modeled performance at low tracker fidelity by artificially degrading recorded eye tracking data. For example, [Holmqvist et al. 2012] found a strong influence of a constant offset and Gaussian noise on common eye tracking measures such as fixation durations and dwell times. [Graupner et al. 2008] added noise to the eye tracking signal in a gaze selection task and observed longer reaction times and reduced hit rates. [Orquin and Holmqvist 2018] calculated expected performance ("capture rate") across target sizes and tracker quality using a purely simulation-based approach. Most work in gaze selection models tracker error only and assumes optimal performance on the user's part. However, selection performance can also be defined as a combination of *human behavior* (influenced by e.g. oculomotor noise [Harris and Wolpert 2006; van Beers 2007; Van Opstal and Van Gisbergen 1989] and corrective saccades [Becker 1972; Becker and Fuchs 1969; Prablanc et al. 1978; Schuetz et al. 2019; Wu et al. 2010]), and *tracker error*, often modeled as normally distributed variation around the true gaze point [Reingold 2014]. Here, we present a behavior-centered approach to quantifying gaze selection performance for a given level of eye tracker accuracy, precision, and sampling rate. The method is inspired by psychophysics, specifically the psychometric function [Wichmann and Hill 2001], which is typically used to describe human performance (e.g., the probability of detecting a stimulus) as a function of some physical stimulus property (e.g., the brightness of the above stimulus). We apply the psychometric function in a novel context to model gaze selection performance relative to physical target size. Using recorded human gaze behavior as a baseline, we then simulate lower-fidelity eye tracking data by adding spatial offset, noise, and down-sampling and predict the resulting performance function.

2 METHODS

2.1 Participants

Sixteen volunteers participated in the user study (10 male, 6 female; mean age 27.8 ± 5.7 years). All participants had normal or corrected-to-normal vision and no known history of visual or oculomotor deficits. Participants were recruited via social media, email, and word of mouth, provided written informed consent and could end the experiment at any time. Digital gift cards were provided as a gratuity.

2.2 User Study

Participants were instructed to select visual targets using their gaze, by performing goal-directed eye movements from a starting location to each target as it appeared on the screen. Each trial began with a fixation cross (type ABC in [Thaler et al. 2013]) presented pseudo-randomly at one of nine possible starting locations (centrally or ± 10° horizontally and/or vertically). After the participant fixated the cross and pressed a key, a target circle appeared at a 5° or 10° distance from the starting location following a randomized delay of 250-1000 ms. Eye movements were then recorded for a period of 2000 ms after target onset. As the goal of the task was to collect gaze behavior and not emulate an interactive system, no feedback about successful selection was given. Targets were presented at a random angle to avoid adaptation to a specific target location [Van Opstal and Van Gisbergen 1989] and comprised

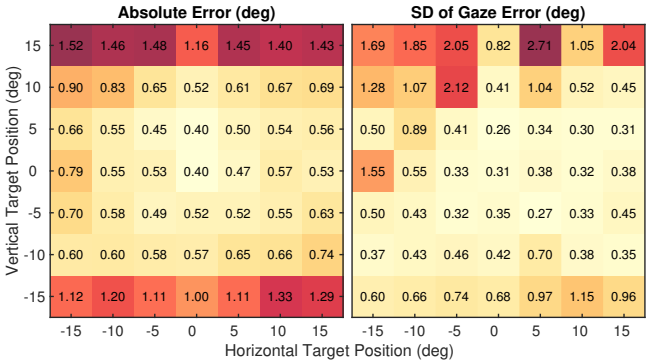

Figure 2: Baseline eye tracking accuracy (absolute error) and precision (SD of gaze error) across a visual field of ±15° re: screen center, recorded at the start of each experimental block.

seven different diameters (0.5, 1, 1.5, 2, 3, 4, and 5°). Each combination of target size and saccade amplitude was repeated 10 times from a pseudo-random starting location, leading to 140 trials per participant. Trials were presented in five blocks of 28 trials, and participants had the option to take a break after each block.

Stimuli were presented on a back-projection screen using a projector. Participants sat in front of the screen at a distance of 1.8 m with their head stabilized by a chin and forehead rest. An Eyelink 1000plus eye tracker was used to record eye movement data. At the start of each block, the eye tracker was calibrated using the built-in 13-point calibration and validation procedure. Additionally, we recorded custom validation data at the beginning of each block by asking participants to fixate each target in a 7 × 7 grid of fixation crosses in pseudo-random order (spanning ± 15 visual degrees horizontally and vertically) for 250 ms before advancing to the next target. This data was used to quantify the baseline gaze accuracy and precision of our eye tracking setup. For further details on the experimental setup and paradigm, please see [Schuetz et al. 2019].

2.3 Gaze Data Analysis

Eye movement analyses were performed using MATLAB. Saccadic eye movements in each trial were detected using the manufacturer-provided algorithm at default thresholds for velocity and acceleration (30°/s and 8000°/s^2). Saccades with durations below 20 ms or above 100 ms were discarded as artifacts. No additional filters were applied to gaze data for the purpose of this study.

To compensate for individual saccadic reaction times, we report all eye movement times relative to the offset of the first saccade after the target appeared on each trial. Gaze positions were extracted for the saccadic offset time and a range of *sampling windows* between 1 and 300 ms (1-300 samples at 1 kHz) after this time. To determine selection success on a given trial, median horizontal and vertical gaze positions within the sampling window were calculated and their Euclidean distance to the target center was compared to the target radius. A trial was successful if this gaze error was smaller than or equal to the target radius. Individual success rates were aggregated into selection performance (probability of successful selection) per target size and participant, and further averaged across participants (Fig. 1A).

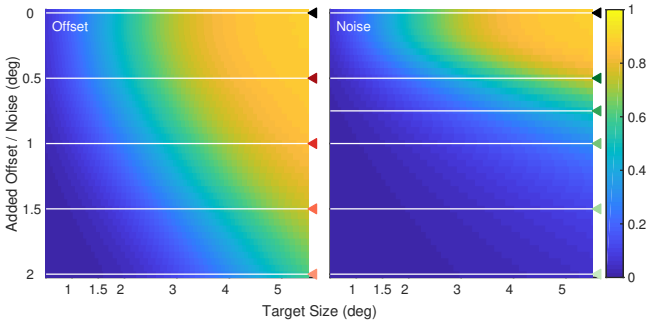

Figure 3: Average selection performance by target size across all simulated steps for offset (accuracy) and noise (precision). Horizontal lines and markers correspond to the specific curves shown in Fig. 1B and C.

2.4 Signal Quality Simulation and Model Fit

To estimate gaze selection performance for eye tracking fidelity that is lower than for our recorded gaze data, we extracted gaze time series for every trial, degraded the quality of the recorded time series as described below, then determined successful or unsuccessful selection for different time frames within the recorded trial. Signal quality was manipulated along three dimensions: First, the time series was down-sampled by retaining every *n*́th frame of data for all possible integer factors of 1000 Hz (1000 - 10 Hz). Second, a constant offset of 0 - 2° in 0.05° intervals was applied to the gaze position to simulate reduced accuracy. To ensure that simulated offset was not directionally biased, gaze data was offset by the same distance in each of the 8 cardinal directions. Finally, 63 noise samples drawn from a Gaussian distribution with a Standard Deviation (SD) of between 0 and 2° were generated for each sample of the down-sampled and offset time series to simulate reduced precision. For each combination of offset, noise, and sampling rate, the resulting 504 (8 directions × 63 noise samples) permutations of each trial's gaze data were then analyzed as described in Section 2.3 and averaged to determine the proportion of successful selection trials.

To model selection performance, we extended a method first described in [Schuetz et al. 2019], akin to the fitting of a Psychometric function (PF; [Wichmann and Hill 2001]). Selection probability across target size was fit with a Weibull cumulative distribution function (CDF). In addition to the Weibull parameters α and β, a PF commonly models a base error rate (also "lapse rate" λ; here: rate of unsuccessful selections) as well as a "guess rate" γ corresponding to chance performance. As we expect selection probability to trend to zero as targets get smaller, we fixed the guess rate parameter at zero and only modeled the base error rate. The following function was thus fit using MATLAB's implementation of the Nelder-Mead simplex search algorithm:

$$P_{sel}(s;\alpha,\beta,\lambda) = (1-\lambda)1 - e^{-(\frac{s}{\alpha})^{\beta}} \quad (1)$$

where s refers to target size. The choice of Weibull over other common PFs (such as a Gaussian CDF, [Wichmann and Hill 2001]) was further motivated by the fact that the Weibull asymptotes to zero performance as target size approaches zero. The *threshold* α indicates the target size for a performance of 0.816, while β controls the *shape* of the function [May and Solomon 2013].

3 RESULTS

3.1 Baseline Eye Tracking Fidelity

As a first indicator of the performance of our eye tracking setup, the built-in Eyelink validation routine provides accuracy values averaged across the 13 standard calibration targets. Reported accuracies across all blocks (16 participants × 5 blocks = 80 validations) ranged from 0.25 - 1.30° for the left eye (mean: 0.50°, median: 0.47°) and 0.15 - 1.83° (mean: 0.50°, median: 0.45°) for the right eye, in line with typical values reported in the device manual (0.25-0.50°).

Additionally, we analyzed data from a custom validation routine (see Methods). Fig. 2 displays average accuracy (gaze error; Euclidean distance between median gaze position and target) and precision (SD of gaze error) for each custom target. Mean absolute gaze errors ranged from 0.40 - 1.52° (mean: 0.79°, median: 0.65°), and SDs from 0.26 - 2.71° (mean: 0.77°, median: 0.50°). The largest error and variability were reported for the extreme vertical target positions (+15° and -15°). Analyzed separately (not shown in Fig. 2), average horizontal and vertical errors were close to zero, suggesting that no global spatial bias was present across the screen plane (means: horizontal, <0.001°, vertical, 0.04°; range: horizontal, -0.27-0.31°, vertical, -0.28-0.55°).

3.2 Modeling Selection Performance

Despite variation in individual participant performance (e.g., base error rate 0.0-0.5, mean 0.11; see also Fig. 1A), individual Weibull CDFs fit to the average selection performance across target sizes explained the selection performance curves well (R^2 between .870 and .997). Figure 1 (B-D) shows how changes from baseline average selection performance (black line and markers) are captured by the Weibull model when signal quality is manipulated along each direction independently (numbers in parentheses indicate final accuracy and precision of the dataset after manipulation, assuming independent Gaussian noise sources and the baseline values described above). With increasing *offset*, the threshold (Fig. 1B: $\alpha_{0.0}$ = 2.04, $\alpha_{0.5}$ = 2.23, $\alpha_{1.0}$ = 2.88, $\alpha_{1.5}$ = 3.63, $\alpha_{2.0}$ = 4.58), shape ($\beta_{0.0}$ = 1.87, $\beta_{0.5}$ = 1.92, $\beta_{1.0}$ = 2.26,$\beta_{1.5}$ = 2.92, $\beta_{2.0}$ = 3.51), and base error rate ($\lambda_{0.0}$ = 0.12, $\lambda_{0.5}$ = 0.12, $\lambda_{1.0}$ = 0.14, $\lambda_{1.5}$ = 0.18, $\lambda_{2.0}$ = 0.20) all increased monotonically, indicating larger necessary target sizes for similar performance and more missed selections. Increasing *noise* raised the threshold and thus minimal target size up to an added noise level of about 1° ($\alpha_{0.0}$ = 2.04 , $\alpha_{0.5}$ = 3.39, $\alpha_{1.0}$ = 8.72), with minor variation in the other parameters ($\beta_{0.0}$ = 1.87, $\beta_{0.5}$ = 1.96, $\beta_{1.0}$ = 1.95; $\lambda_{0.0}$ = 0.12, $\lambda_{0.5}$ = 0.14, $\lambda_{1.0}$ = -0.05). Past this value, α and λ showed stark variation despite following a similar overall trend ($\alpha_{1.5}$ = 41.76, $\alpha_{2.0}$ = 27.75; $\beta_{1.5}$ = 1.95, $\beta_{2.0}$ = 1.98; $\lambda_{1.5}$ = -6.14, $\lambda_{2.0}$ = -0.56), which could indicate that the tested range of target sizes is not sufficient to accurately fit beyond this level of noise (see also Fig.1C). Finally, *sampling rate* had no impact on all parameters except at rates $< 25Hz$ (α_{1kHz} = 2.04, α_{10Hz} = 2.31; β_{1kHz} = 1.87, β_{10Hz} = 1.91; λ_{1kHz} = 0.12, λ_{10Hz} = 0.28) and is therefore not considered further in the current analysis. Fig. 3 plots a surface of the full ranges of added offset and noise, showing that the model smoothly

captures changes along both dimensions (horizontal cross-sections: functions in Fig. 1B and C).

3.3 Predicting Selection Performance

Using the above model, we can estimate expected selection performance for a target of given size. Fig. 4 illustrates this for the largest target included in our set (5°). Individual panels in Fig. 4 demonstrate how averaging over longer time windows reduces the effect of noise. The last panel (300 ms) also shows a large increase in performance compared to shorter windows, likely due to the capture of secondary ("corrective") saccades within this time frame [Schuetz et al. 2019; Wu et al. 2010].

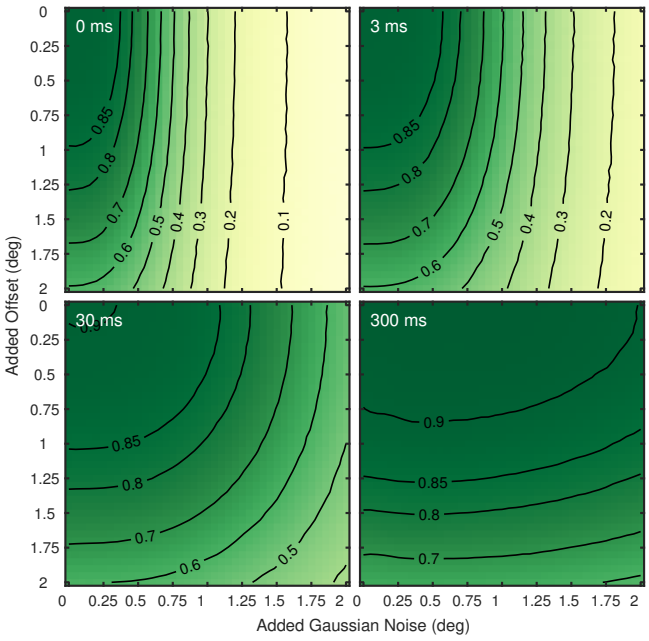

Figure 4: Predicted average selection performance (isolines) for a 5° target, as a function of added offset and noise. Panels show different window sizes for averaging after the offset of the first saccade (data sampled at 1kHz).

3.4 Predicting Minimal Target Size

Besides quantifying performance for a given target size, the model also allows to choose a desired level of selection performance and estimate the minimal target size to achieve this level of performance with a specific combination of eye tracker parameters. Fig. 5 plots predicted minimal target sizes for a performance level of 80% correct selections across accuracy and precision. White areas in the plot indicate regions in signal quality space where the the Weibull curve asymptotes at a lower value and never reaches 80%, or could not be fit due to limited range of measured target sizes (cf. Section 3.2).

4 DISCUSSION

We present a psychophysics-inspired model of combined human and eye tracker performance in a gaze selection task, which can be

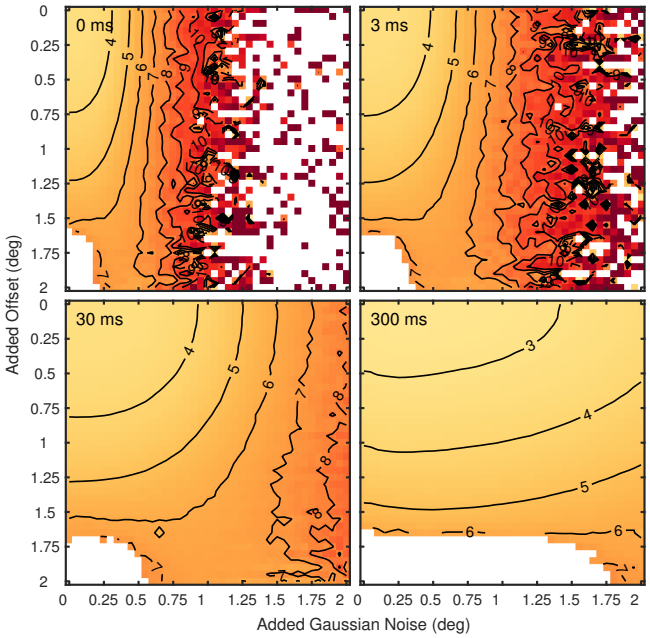

Figure 5: Predicted minimal target size (°) to achieve 80% selection performance. Panels as in Fig. 4. White areas indicate where specified selection performance cannot be achieved.

applied to the design of a gaze interaction scenario in multiple ways. First, the model can be fit to individual or averaged data to quantify selection performance for a given system. Second, gaze selection performance can be predicted at reduced tracking fidelity, such as for a subset of participants with low calibration accuracy or where environmental factors (e.g., headset slippage, vibration) impact eye tracking. Third, we can directly estimate design suggestions such as minimal target size when building a gaze interaction task for a given eye tracker. Finally, this approach could enable fast and efficient characterization of gaze interaction using adaptive psychophysics methods ([Treutwein 1995; Watson and Pelli 1983]).

The model as presented here includes some limitations: It is only valid in the screen plane and cannot predict performance for 3D targets or gaze depth from eye vergence (e.g., [Duchowski et al. 2014; Elmadjian et al. 2018]). The task only included a single target and thus does not yet model disambiguation and spacing between multiple targets. The number of participants should be increased in future studies to better capture human variability. Moreover, the current model only implements signal degradation and cannot predict performance for eye tracker fidelity *better* than the recorded data. Future work could predict PF parameters directly from signal quality measures to allow extrapolation.

Although the results provided here are likely applicable to similar equipment (Eyelink 1000 plus, desktop mode) as a conservative baseline, they might not directly generalize to other hardware due to different tracker-specific noise behavior. Future studies should measure selection performance with a lower end system to validate the model, and determine how common hardware varies in bias and noise. Finally, while our method integrates human behavior and eye tracker fidelity in a specific performance curve, it does not yet allow

separation of these sources of variability. To fully independently model the contributions of both sources to gaze selection (or other tasks), further work could simulate optimal human behavior in a given task, then apply offset and noise distributions derived from a given eye tracker before predicting performance.

5 ACKNOWLEDGMENTS

We thank Mark Opaliski for help with data collection.

REFERENCES

W Becker. 1972. The control of eye movements in the saccadic system. *Bibliotheca ophthalmologica: supplementa ad ophthalmologica* 82 (1972), 233–243.

W Becker and AF Fuchs. 1969. Further properties of the human saccadic system: eye movements and correction saccades with and without visual fixation points. *Vision research* 9, 10 (1969), 1247–1258.

Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*. ACM, 1.

Pieter Blignaut, Kenneth Holmqvist, Marcus Nyström, and Richard Dewhurst. 2014. Improving the accuracy of video-based eye tracking in real time through post-calibration regression. In *Current Trends in Eye Tracking Research*. Springer, 77–100.

Andreas Bulling and Hans Gellersen. 2010. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing* 9, 4 (2010), 8–12.

Andrew T Duchowski, Donald H House, Jordan Gestring, Robert Congdon, Lech Świrski, Neil A Dodgson, Krzysztof Krejtz, and Izabela Krejtz. 2014. Comparing estimated gaze depth in virtual and physical environments. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 103–110.

Carlos Elmadjian, Pushkar Shukla, Antonio Diaz Tula, and Carlos H Morimoto. 2018. 3D gaze estimation in the scene volume with a head-mounted eye tracker. In *Proceedings of the Workshop on Communication by Gaze Interaction*. ACM, 3.

Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1118–1130.

Sven-Thomas Graupner, Michael Heubner, Sebastian Pannasch, and Boris M Velichkovsky. 2008. Evaluating requirements for gaze-based interaction in a see-through head mounted display. In *Proceedings of the 2008 symposium on Eye tracking research & applications*. ACM, 91–94.

Dan Witzner Hansen and Qiang Ji. 2010. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE transactions on pattern analysis and machine intelligence* 32, 3 (2010), 478–500.

John Paulin Hansen, Anders Sewerin Johansen, Dan Witzner Hansen, Kenji Itoh, and Satoru Mashino. 2003. Command without a click: Dwell time typing by mouse and gaze selections. In *Proceedings of Human-Computer Interaction–INTERACT*. 121–128.

Christopher M Harris and Daniel M Wolpert. 2006. The main sequence of saccades optimizes speed-accuracy trade-off. *Biological cybernetics* 95, 1 (2006), 21–29.

Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. 2012. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications*. ACM, 45–52.

Robert JK Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 11–18.

Paivi Majaranta, Hirotaka Aoki, Mick Donegan, Dan Witzner Hansen, and John Paulin Hansen. 2011. Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies. (2011).

Päivi Majaranta and Andreas Bulling. 2014. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*. Springer, 39–65.

Keith A. May and Joshua A. Solomon. 2013. Four Theorems on the Psychometric Function. *PLOS ONE* 8, 10 (10 2013), 1–34. https://doi.org/10.1371/journal.pone.0074815

George W McConkie. 1981. Evaluating and reporting data quality in eye movement research. *Behavior Research Methods & Instrumentation* 13, 2 (1981), 97–106.

Jacob L Orquin and Kenneth Holmqvist. 2018. Threats to the validity of eye-movement research in psychology. *Behavior research methods* 50, 4 (2018), 1645–1656.

C Prablanc, D Masse, and JF Echallier. 1978. Error-correcting mechanisms in large saccades. *Vision research* 18, 5 (1978), 557–560.

Eyal M Reingold. 2014. Eye tracking research and technology: Towards objective measurement of data quality. *Visual cognition* 22, 3-4 (2014), 635–652.

Immo Schuetz, T. Scott Murdison, Kevin J. MacKenzie, and Marina Zannoli. 2019. An Explanation of Fitts' Law-like Performance in Gaze-Based Selection Tasks Using a Psychophysics Approach. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 535, 13 pages. https://doi.org/10.1145/3290605.3300765

Vildan Tanriverdi and Robert JK Jacob. 2000. Interacting with eye movements in virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 265–272.

Lore Thaler, Alexander C Schütz, Melvyn A Goodale, and Karl R Gegenfurtner. 2013. What is the best fixation target? The effect of target shape on stability of fixational eye movements. *Vision Research* 76 (2013), 31–42.

Bernhard Treutwein. 1995. Adaptive psychophysical procedures. *Vision research* 35, 17 (1995), 2503–2522.

Robert J van Beers. 2007. The sources of variability in saccadic eye movements. *Journal of Neuroscience* 27, 33 (2007), 8757–8770.

AJ Van Opstal and JAM Van Gisbergen. 1989. Scatter in the metrics of saccades and properties of the collicular motor map. *Vision research* 29, 9 (1989), 1183–1196.

Colin Ware and Harutune H Mikaelian. 1987. An evaluation of an eye tracker as a device for computer input. In *Acm sigchi bulletin*, Vol. 17. ACM, 183–188.

Andrew B Watson and Denis G Pelli. 1983. QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics* 33, 2 (1983), 113–120.

Felix A Wichmann and N Jeremy Hill. 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. *Attention, Perception, & Psychophysics* 63, 8 (2001), 1293–1313.

Chia-Chien Wu, Oh-Sang Kwon, and Eileen Kowler. 2010. Fitts's Law and speed/accuracy trade-offs during sequences of saccades: Implications for strategies of saccadic planning. *Vision Research* 50, 21 (2010), 2142–2157.