Error-Aware Gaze-Based Interfaces for Robust Mobile Gaze Interaction

Michael Barz
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
michael.barz@dfki.de

Florian Daiber
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
florian.daiber@dfki.de

Daniel Sonntag
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
daniel.sonntag@dfki.de

Andreas Bulling
Max Planck Institute for Informatics
Saarland Informatics Campus, Germany
bulling@mpi-inf.mpg.de

ABSTRACT

Gaze estimation error can severely hamper usability and performance of mobile gaze-based interfaces given that the error varies constantly for different interaction positions. In this work, we explore error-aware gaze-based interfaces that estimate and adapt to gaze estimation error on-the-fly. We implement a sample error-aware user interface for gaze-based selection and different error compensation methods: a naïve approach that increases component size directly proportional to the absolute error, a recent model by Feit et al. that is based on the two-dimensional error distribution, and a novel predictive model that shifts gaze by a directional error estimate. We evaluate these models in a 12-participant user study and show that our predictive model significantly outperforms the others in terms of selection rate, particularly for small gaze targets. These results underline both the feasibility and potential of next generation error-aware gaze-based user interfaces.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Mobile computing**;

KEYWORDS

Eye Tracking; Mobile Interaction; Gaze Interaction; Error Model; Error-Aware

ACM Reference Format:
Michael Barz, Florian Daiber, Daniel Sonntag, and Andreas Bulling. 2018. Error-Aware Gaze-Based Interfaces for Robust Mobile Gaze Interaction. In *ETRA '18: 2018 Symposium on Eye Tracking Research and Applications, June 14–17, 2018, Warsaw, Poland.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3204493.3204536

ETRA '18, June 14–17, 2018, Warsaw, Poland

ACM ISBN 978-1-4503-5706-7/18/06...$15.00
https://doi.org/10.1145/3204493.3204536

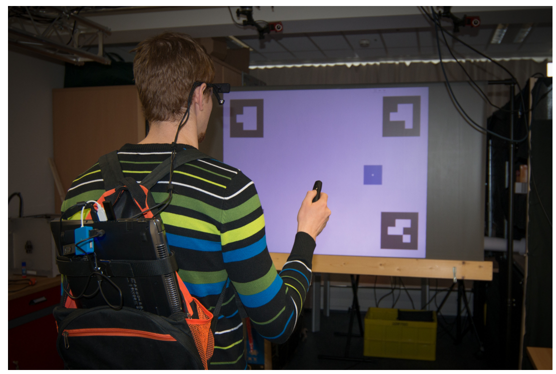


Figure 1: Study participant interacting with the proposed error-aware gaze-based interface using a mobile eye tracker.

1 INTRODUCTION

Head-mounted eye trackers are promising for mobile interaction as they provide information about the user's intentions. Human gaze conveys the user's interest [Shell et al. 2003] which is already used for interaction with one or multiple displays [Lander et al. 2015a; Stellmach et al. 2011; Turner et al. 2014]. A key problem of mobile gaze-based interaction is that the gaze estimation error, i.e., the difference between the estimated and true on-screen gaze position, can be substantial, in particular if the user moves in front of a display [Cerrolaza et al. 2012; Lander et al. 2015a; Mardanbegi and Hansen 2012]. Besides user position and orientation, also factors specific to the eye tracker and display, e.g., parameters of the calibration routine and of the display detection algorithm, can have significant impact on the gaze estimation error [Barz et al. 2016].

Some interaction techniques omit the need for accurate point of gaze (POG) estimates, e.g., by correlating raw eye movements with animated on-screen targets [Esteves et al. 2015; Vidal et al. 2013], but introduce the need for dynamic user interfaces. Other methods aim to address this problem by filtering gaze jitter [Špakov 2012], snapping gaze to on-screen objects [Špakov and Gizatdinova 2014] or by optimising interface layouts for gaze estimation error at the design stage [Feit et al. 2017]. Although such methods can improve user experience, they only alleviate the symptoms and do not embrace the inevitable gaze estimation error in the interaction design. We first aimed for that goal in [Barz et al. 2016]: we proposed a model to predict gaze estimation error for any 3D surface – specified by visual markers attached to it – in the field of view of the eye tracker's scene camera in real time. The model covers key error sources relevant for mobile gaze interaction, namely mapping of pupil positions to scene camera coordinates, marker-based display detection, and mapping from scene camera to on-screen gaze coordinates. However, the authors only described the model and neither implemented it in a real user interface nor evaluated its usefulness for mobile gaze interaction.

In this work we aim to fill this gap by presenting the first gaze-based user interface that is "aware" of the ever-changing gaze estimation error and that can adapt to the error on the fly. First, we implement a compensation method that scales targets proportional to real-time error estimates and evaluate it in a 12 participant user study. Considered estimators from our prior work include two simple models based on a fixed angular error, a predictive model and a baseline using a fixed target size. The best selection rates are achieved with the predictive model, on the cost of large target sizes: it is not clear to which extent the advantage in selection rate originates from the target size. Hence, we describe two more advanced error compensation methods and investigate their effect on selection rates and target sizes using our study corpus. The results of our analysis suggest that the selection rate can be improved using directional error estimates without increasing the average size of targets. However, this advantage vanishes with increasing target sizes. A great improvement can be achieved by training personalized directional error models: with this approach high selection rates can be achieved even with low target sizes. All in all, our architecture and error models enable a new class of gaze interaction that incorporates the gaze estimation error and is suitable for many fields of application in mobile and ubiquitous computing.

The specific contributions of our work are two-fold: First, we propose and prototypically implement the first error-aware gaze interface that scales selection targets depending on the current error estimate [Barz et al. 2016]. We evaluate different error models with this interface in a mobile interaction user study with 12 participants. Second, informed by the findings of the first study, we propose two advanced adaptation techniques: a method based on [Feit et al. 2017] that scales gaze selection targets according to the 2D error distribution in the data, and a novel method that combines scaling targets and shifting gaze by directional error estimates.

2 RELATED WORK

Our work is related to previous works on (1) mobile eye tracking, especially the detection of ambient displays, (2) studies on the gaze estimation error, and (3) gaze-based interaction.

2.1 Mobile Eye Tracking

Mobile eye trackers usually feature an eye camera recording a close-up video of the user's eye and a scene camera capturing a part of the user's field of view. Gaze is reported in terms of scene camera coordinates and accordingly needs to be mapped to the coordinate system used for interaction [Barz et al. 2016]. A key challenge for mobile gaze-based interfaces is that the tracker's pose relative to that coordinate system needs to be tracked. [Bardins et al. 2008] attach infrared LEDs to the tracker and use a stereo camera on the screen to track its 3D pose. Several works propose to attach visual markers to the displays to reconstruct the tracker pose [Breuninger et al. 2011; Hales et al. 2011; Yu and Eizenman 2004]. [Mardanbegi and Hansen 2011] directly detect displays in the eye tracker's scene camera for gaze mapping. [Lander et al. 2015a] propose GazeProjector, an approach for mobile gaze-based interaction with ambient displays. They seamlessly integrate multiple displays based on natural feature tracking to estimate the user's pose. [Kassner et al. 2014] present Pupil, an open source eye tracker that includes marker-based surface tracking, but they do not evaluate the tracking nor gaze-based interaction with the display. In this work, we rely on marker tracking similar to [Kassner et al. 2014], to map gaze from scene camera coordinates to display coordinates.

2.2 Gaze Estimation Error

The importance of the gaze estimation error has long been acknowledged, although, previous works mainly focus on error sources or post-hoc error compensation. Sources raising an error in gaze estimation are manifold and comprise, e.g., physiological differences such as the influence of pupil size [Drewes et al. 2012], experimental factors such as recording time, gaze direction as well as the experimenter's experience [Nyström et al. 2013] and errors that occur due to deformations or slipping of the head-mounted eye tracker [John et al. 2012].

Several methods are proposed that concentrate on compensating the gaze error post-hoc. Some recent works focus on measuring and compensating for a wider range of gaze estimation errors. Zhang et al. propose a correction method that can reliably reduce disparity error [Zhang and Hornof 2011] and a technique for post-hoc gaze estimation error correction for a stationary eye tracker [Zhang and Hornof 2014]. [Cerrolaza et al. 2012] show that particularly head movements perpendicular to the screen cause errors in gaze estimation. They propose a new calibration procedure and gaze estimation function to incorporate the eye-to-screen distance and compensate for the error. [Blignaut and Wium 2013] compare different models and show that the arrangement and number of calibration targets have a significant effect on gaze mapping accuracy. In another work, they investigate disparity and propose post-calibration regression to improve the accuracy of gaze recordings [Blignaut et al. 2014]. [Barz et al. 2016] present a novel approach to model and predict the gaze estimation error for mobile eye trackers. [Feit et al. 2017] share their vision of error-aware interfaces and investigate the gaze estimation error for remote tracking devices, but restricted to a constant interaction distance. They model the error as the mean gaze offset and its standard deviation (SD) for different positions on a display. They show that the SD can be used to optimize the parameters of different signal filters and the mean offset to inform gaze-based interface design. We include the findings from [Barz et al. 2016] and [Feit et al. 2017] for prototyping a gaze-based interface that is aware of the error inherent in mobile eye trackers and evaluating its effectiveness in improving the users' selection rates.

2.3 Gaze-Based Interaction

A fundamental issue when using gaze for direct input is the Midas touch problem: human gaze is always active and it has to be decided when to use the signal and when not to use it. [Jacob 1991] suggests using a dwell time or a key press to trigger an action at the current gaze location. More recently, [Stellmach and Dachselt 2012] propose a set of interactions using a mobile touch device to confirm gaze input. We decided to use a wireless presenter to trigger a selection, because it's lightweight and easy to integrate.

An alternative to overcome the Midas problem is to use gaze as context information: eye tracking passively informs interaction. [Toyama et al. 2012] propose a system for gaze-guided object recognition to detect exhibits in a museum in proximity to the user's gaze. Extensions of this work add face and text detection [Sonntag 2015; Toyama and Sonntag 2015] and increase the number of predictable classes using deep learning [Barz and Sonntag 2016]. [Lander et al. 2015b] use gaze to automatically scroll text on public displays in a multi-user setting. [Barz et al. 2017] use gaze in combination with speech to infer the location of referenced objects.

Several interaction techniques were proposed that address the problem of error inherent to eye tracking. A large body of work focused on dealing with gaze jitter, e.g., by filtering or snapping gaze to on-screen objects using gaze-to-object mapping algorithms [Špakov 2011, 2012; Špakov and Gizatdinova 2014]. [Miniotas et al. 2004] showed that target expansion was able to effectively compensate for gaze jitter in gaze-based target selection. [Monden et al. 2005] proposed a multi-modal target selection method combining gaze and mouse to compensate for jitter and similar errors. [Zhang et al. 2008] investigated means to stabilise the eye cursor and compensate for gaze jitter. They showed that their methods outperformed standard dwell-time based eye pointing.

One alternative to head pose tracking is to use interaction techniques that do not require accurate POG estimates and are therefore calibration-free. [Shell et al. 2003] introduced EyePliances, devices in the environment that were equipped with eye contact sensors to passively detect and react on the user's visual attention. Gaze gestures – sequences of consecutive relative eye movements – were introduced as a calibration-free but active interaction technique [Bulling et al. 2008]. [Zhang et al. 2013, 2014] presented SideWays and GazeHorizon, systems for calibration-free gaze interaction with ambient displays. Their approach used an off-the-shelf webcam attached to the display for face and eye detection. In a recent work, [Vidal et al. 2013] proposed smooth pursuit eye movements, i.e., the eye movements we perform when following a moving object, for gaze interaction with ambient displays. They matched users' eye movement to object movements on the screen by performing a light-weight correlation computation. Their approach represents one of the few that is both calibration-free and allows for high-fidelity interaction, but it requires a dynamic interface. The pursuits approach was also applied to mobile gaze-based interaction with smart watches [Esteves et al. 2015]. All these approaches have in common that they typically do not provide high-fidelity interactions due to the lack of accurate on-screen gaze positions.

3 DESIGN

In this work, we propose an approach for gaze-based interaction that – in contrast to existing methods – incorporates the gaze estimation error of the eye tracker in real-time. An overview of our architecture is depicted in Figure 2. The *Gaze Estimation and Input Data Acquisition* component connects a mobile head-mounted eye tracker to receive gaze information with respect to any predefined display and further inputs required for predicting the gaze estimation error. In most cases, this part will integrate with the software package that accompanies the eye tracking hardware. All parameters are sent to the *Error-Aware Interface Manager* that adjusts the interface by compensating the gaze estimation error based on a real-time estimate of an interchangeable *Error-Model Component* and presents it to the user. In the following we describe the individual components of our architecture and their interplay.

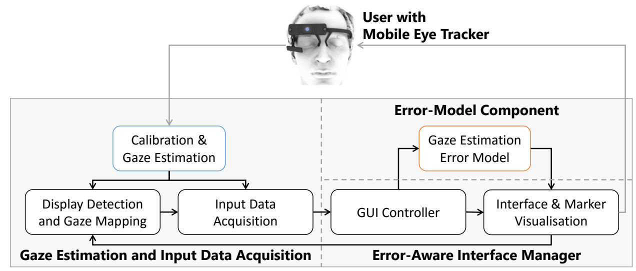


Figure 2: Architecture of the error-aware interface.

3.1 Gaze Estimation and Input Data Acquisition

This component has three major tasks. First, it connects an eye tracking device to handle calibration and to receive gaze data. Second – if not fulfilled by the eye tracker's API – it is responsible of detecting displays and mapping gaze on these displays. Third, the component acquires all necessary input parameters for the *Error-Model Component*. The marker visualisation as well as executing the error estimation is part of the *Error-Aware Interface Manager*.

3.2 Error-Aware Interface Manager

The interface manager is the core component of our framework and accomplishes several tasks. First, it handles the interface elements comprising their properties and events, similar to any user interface framework. Any module, e.g., containing the application logic, can add elements, influence its position and register for events. Second, it handles the presentation of markers that are used by the *Gaze Estimation and Input Data Acquisition* to identify the display and compute the eye tracker pose. Third, and most importantly, this component interfaces the *Error-Model Component* to gather essential data for, e.g., adjusting the size of interface elements and shifting the input signal. In this work, we implement and evaluate different compensation methods which are summarised in Figure 3.

3.3 Error-Model Component

The error-model component encapsulates all calculations and algorithms for predicting the gaze estimation error via a shared interface. It is important that the computation time for the error inference is suitable for real-time applications: it should be on a par with the sampling rate of the gaze estimation (e.g., between 30 and 200Hz for mobile devices[1]). The concrete implementation should be interchangeable for testing different approaches or updating models to newer revisions for, e.g., personalizing models. We investigate different error models that serve as input for error-compensation methods in the *Error-Aware Interface Manager*.

4 COMPENSATION BY NAÏVE SCALING

For evaluating our error-aware interaction technique with the *Naïve* target scaling approach, we conduct a user study with 12 participants. The size of the selection target is computed as two times the gaze estimation error μ, because the error estimates are absolute: they do not contain directional information (see Figure 3a). Four error models from [Barz et al. 2016] are considered for estimating the gaze estimation error and thus for scaling selection targets. This includes the pre-trained machine learning model *Predictive*, which takes into account several inputs related to gaze estimation and display detection, e.g., the positions of the pupil and the selection target and the distance and angle of the user's head to the on-screen selection target. Further, we re-implement two simple models *Best* and *Measured* that calculate the gaze estimation error based on the distance d of the user's head to the current on-screen target and a constant angular error e_c. This error is either $0.6°$ as stated by the hardware manufacturer (best-case) or $1.23°$ as measured for the actual setting. The function to approximate the gaze estimation error μ with the distance d and the error $e_c \in \{0.6°, 1.23°\}$ as input is $\mu = d \cdot tan(e_c)$. The baseline model *None* reports a constant error, which is computed once with $1.23°$ and a static distance of d_{cal} = 175 cm as input (centre of the interaction space).

4.1 User Study

The proposed error-aware interface is evaluated for mobile gaze interaction in a public display setting. The interaction with the interface consists of a gaze-based selection task on a large display. Hereby, the *Error-Aware Interface Manager* uses estimates of the *Error-Model Component* for scaling selection targets in real-time: The larger the predicted gaze estimation error, the larger the targets (see Figure 3a). We invited 12 participants (six female) aged between 20 and 53 ($M = 28.68, SD = 10.84$).

4.1.1 Conditions. We investigate the performance of our *Naïve* target scaling approach using different error models for the *Error-Model Component*. The error models introduced above correspond to our four conditions (*Best*, *None*, *Measured* and *Predictive*).

4.1.2 Tasks. For each condition, a calibration is performed at the centre of a 3×3 grid with 50×50 cm cells starting 100 cm in front of the display (approximately $175cm$ and orthogonal to the screen). Then, the selections are performed using six on-screen targets (radially arranged with one at the display centre) from all positions of the 3×3 grid, totalling 216 selections per participant. Stimuli are shown at the same positions in randomized order.

4.1.3 Design. In the user study, we consider four methods for predicting the gaze estimation error to inform our error-aware interface that is based on scaling (within-subject design). The order of conditions is counterbalanced between participants.

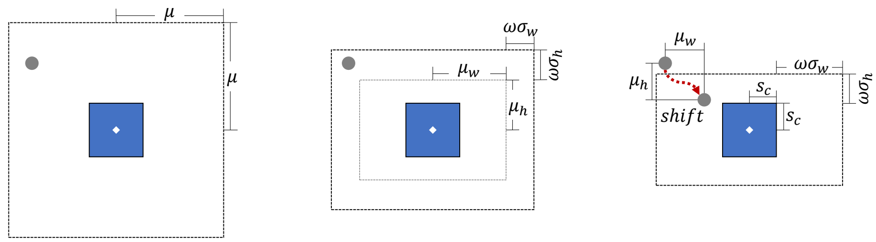
4.1.4 Procedure. First, participants are introduced to the experiment and asked to complete a general questionnaire. Afterwards each participant calibrates the eye tracker for every condition and performs all selections for the currently considered error prediction method (counterbalanced order). We instruct the participants to be as accurate as possible. The grid position is shown to the user prior to each selection. On average one run lasted $967s$.

4.1.5 Apparatus. For the user study, we develop a sample error-aware interface based on our architecture as shown in Figure 2. We use a PUPIL Pro tracker [Kassner et al. 2014] connected to a laptop inside a backpack worn by the participants (see Figure 1). The *Gaze Estimation and Input Data Acquisition* component is based on its attendant open source software for calibration and gaze estimation. We extend their tool by a marker-based display detection component to map gaze from the scene camera coordinate system to the coordinate system of any display or interactive surface in the environment. Besides, we collect all data needed for the components of our interface prototype, especially for the *Error-Model Component*. The prototype of the *Error-Aware Interface Manager* for our study supports a single button and the handling of a corresponding trigger event. Blue rectangles with a white dot at their centre are shown as stimuli on a back-projected screen ($1024 \times 768px$ with $8.88px/cm$). To select a target, participants are asked to press a button on a wireless presenter while fixating on it. This additional modality solves the Midas problem inherent to gaze-based interfaces. When a button click has been performed by the user, we check if there was a recent fixation within the area of a button and raise its trigger event accordingly. It flashes green upon a successful selection and red, otherwise. The edge size of the button is scaled based on a real-time error estimate from the *Error-Model Component* as depicted above. It is possible to change the active model on-the-fly, also during runtime. To prevent jumpy changes of an element's shape, we smooth the edge size by means of the 1-Euro-Filter [Casiez et al. 2012] with $\beta = 0.01$ and $f_{c_{min}} = 0.3$ as parameters. The 3×3 grid is marked on the floor with adhesive tape to coarsely position the participants without restricting mobility. In addition we record the 3D head pose of the user by means of marker tracking to get more fine-grained location data (continuous values for distance and angle).

4.1.6 Independent and Dependent Variables. Independent variables include the error prediction method, the user position and the on-screen target. The models correspond to the conditions as outlined above, the user position is enforced by the 3×3 grid. Additionally, we record the 3D pose to get the distance between user and target as well as the angle of the user to the display. The dependent variable is the selection rate and the size of the target area.

4.1.7 Hypotheses. We hypothesize that the *Naïve* compensation approach achieves best selection results with the *Predictive* model with respect to the selection rate (H1). This method will especially outperform the other approaches for varying distances (H2) and orientations (H3) of the participants in front of the display with reasonable target sizes (H4).

[1]based on the device specifications of Pupil-Labs eye tracking equipment: https://pupil-labs.com/pupil/



(a) ***Naïve*** **Scaling** **(b)** ***Feit et al.*** **[Feit et al. 2017]** **(c)** ***PredictiveShift***

Figure 3: We investigate different methods for adapting the interface and for compensating the gaze estimation error: (a) ***Naïve*** **scales selection targets likewise for both dimensions and is evaluated in the user study with different error models; (b) the method** ***Feit et al.*** **[Feit et al. 2017] that scales targets based on the 2D error distribution extended for mobile settings; (c) our novel method** ***PredictiveShift*** **that shifts gaze based on a directional error estimate.**

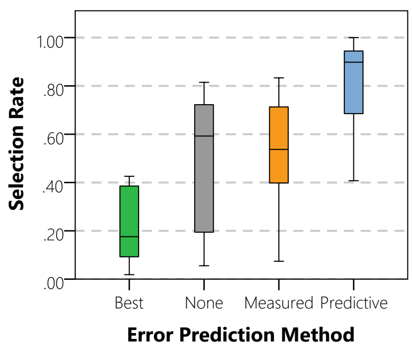


Figure 4: Mean selection rate of different error estimation models averaged over on-screen targets and grid positions.

4.2 Results

Averaged over all on-screen targets and grid positions the selection rate is 22.47% ($SD = 15.56$) for *Best*, 48.92% ($SD = 29.09$) for *None*, 53.4% ($SD = 22.53$) for *Measured* and 81.48% ($SD = 17.99$) for *Predictive* (see Figure 4). A repeated measures ANOVA ($N = 12$) shows that the differences are significant ($F(3, 9) = 56.294, p < 0.001$). All pairwise differences (bonferroni-corrected) are significant, besides the ones between *Measured* and *None*. In addition, all but 1 participant judged *Predictive* as their favourite method.

We further analyse the effect of distance and angle of the user's head to the on-screen target. We cluster the data into three groups for both variables, based on the respective histogram (visually inspected). The resulting intervals are [80, 150] cm (near), [150, 215] cm (mid) and [215, 290] cm (far) for the distances and $[-50, -12.5]^\circ$ (left), $[-12.5, 12.5]^\circ$ (centre) and $[12.5, 50]^\circ$ (right) for the angles to the selection target. For *Measured*, we observe a significant drop in selection rate when moving from the calibration position towards the display (from mid to near) of 43.18% ($F(2, 10) = 14.127, p = 0.001$). Results for *Best* also decrease by 34.92%, but not significantly ($F(2, 10) = 1.448, p = 0.28$). For *None*, we find an inverse effect, i.e., the selection rate increases by 30.39% ($F(2, 10) = 9.988, p = 0.004$). The results for the *Predictive* model reveal a similar effect as for *Measured* and *Best*, but the selection rate only decreases by 20.64% ($F(2, 10) = 17.298, p = 0.001$). We find no significant differences in selection rate considering the intervals for the angles.

Concerning the size of generated targets, we compute the angular error $e_c = 1.98^\circ$ for *Predictive* that would generate, on average, the same target sizes when using the error estimation function of *Best* and *Measured*. The inverse of this function is used with the distance and error estimation result of each selection as input. Further, we average the selection rate and the target area over all participants which maintains the variance for different user positions and targets for each error model (see Figure 5). Hereby, the performance of the *Predictive* model is measured with reduced and increased edge lengths to see whether it systematically over- or underestimates the gaze estimation error using $\{0.8, 0.9, 1.1, 1.2\}$ as factors for scaling the edge length. Concerning the *Predictive* method, the selection rate improves with a regressive slope whereas the target area grows quadratically. The mean target size for the *Predictive* model is $168.15cm^2$ ($SD = 84.67$) which is larger than $76.15cm^2$ ($SD = 41.2$) for *Measured*, $60.38cm^2$ with zero variance for *None* and $17.09cm^2$ ($SD = 9.27$) for *Best*.

4.3 Analysis

The evaluation confirms that our compensation method *Naïve* performs best with the *Predictive* model, significantly outperforming the two simple models and the baseline method for gaze error estimation (supports H1). On average, *Best* performs significantly worse and *Measured* performs as well as *None* which uses no error compensation. However, the performance for far user positions significantly increases for *Measured* whereas it decreases for *None*, same holds for *Predictive* and *None*. This inverse behaviour confirms H2. We could not find a dependency between selection rate and the angle between user and display and thus could find no evidence for H3. Our evaluation shows that using an error of 0.6° (*Best*) as reported by the manufacturer is inapplicable for mobile settings. *Measured* assumes a more realistic error, which yields better selection rates. The model based approach *Predictive* outperforms all others in terms of the selection rate, but would assume the highest error of 1.98°, if it was a distance dependent method.

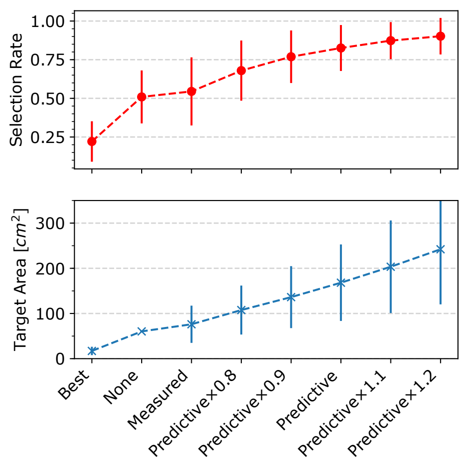


Figure 5: Mean selection rate and mean target area for error estimation models averaged over all participants. The error bars indicate the SD (±) of $n = 54$ selections per model.

The distribution of its target areas ($M = 168.15cm^2$) as shown in Figure 5 is broader compared to *Measured* and *Best* indicating a higher degree of adaptation; accordingly, the variance of *None* is zero. Systematically increasing and decreasing the estimated target size shows how the selection performance can be traded off against the target size and that *Predictive* seems to be a reasonable compromise. However, we cannot confirm H4, because it is unclear which portion of the improvement stems from choosing the target sufficiently large and to what extent from the adaptive behaviour of the *Predictive* model. Another reason might be the high variance in performance of different participants which cannot be explained by these models, i.e., the *Predictive* model might predominantly explain the system error of the eye tracking device.

Eventually, using the *Naïve* scaling approach results in undesirable selection rates or relatively high target sizes which might be up to the compensation method or the error models. For this reason, we develop a more sophisticated method that incorporates directional error information, hence, enables not only naïve scaling of interface controls, but also shifting gaze towards the user's actual point of regard. Further, we investigate user-specific error modelling for covering personal differences in tracking quality and interaction style. We evaluate our new compensation method and compare it to a recent approach by [Feit et al. 2017] in a post-hoc analysis with our recorded data which is described next.

5 ADVANCED COMPENSATION METHODS

Based on our findings from the user study, we develop two additional compensation methods: a recent model based on [Feit et al. 2017] that considers the 2D distribution of the gaze estimation error and the novel model *PredictiveShift* that shifts gaze by a directional error estimate. Both methods implement the compensation part of the *Error-Aware Interface Manager* and the *Error-Model Component*.

[Feit et al. 2017] introduce an approach for modelling the eye tracking error using a 2D Gaussian. For each on-screen position, they compute the spatial accuracy as mean offset μ and the spatial precision as SD σ of the respective gaze samples for x and y. μ and σ which define a 2D Gaussian are used for approximating the gaze error distribution. The target size for a on-screen position (width an height) is computed as $S_{w/h} = 2 \cdot (\mu_{w/h} + 2 \cdot \sigma_{w/h})$. Using the doubled SD to either side includes 95% of all samples assuming a normal distribution of the data. By definition, this approach achieves a selection rate of around 95% under the assumption that the error distribution does not differ much during interaction. However, this approach was not meant nor used for real-time error estimation in adaptive user interaction and it is limited to a static setting: only on-screen targets can vary. We extend their approach for mobile settings: we consider different positions in front of the display using a look-up table with pre-computed means and standard deviations for all combinations of grid and screen positions. Further, we add the factor ω to control the influence of the precision estimate σ which allows to investigate the effect of changing target sizes. Our adapted *Feit et al.* version computes the target size as $S_{w/h}(\omega) = 2 \cdot (\mu_{w/h} + \omega \cdot \sigma_{w/h})$ depending on the grid position where the user is standing and the location of the on-screen target (see Figure 3b).

Our novel approach *PredictiveShift* is based on modelling the directional information of the gaze estimation error which enables to shift the estimated gaze point to the actual point of regard, thus reducing the systematic error of the eye tracking setup and personal characteristics in focusing gaze targets. The error model is a multivariate ElasticNet regression model implemented in scikit-learn [Pedregosa et al. 2011] using a preceding standard scaler for standardising input features by removing the mean and scaling them to unit variance. We use the following predictor variables from our user study for estimating μ and σ (for x and y): coordinates of the on-screen target, distance to the fixated display region, horizontal angle to the display and estimated gaze position in world camera coordinates. The distance, angle and on-screen position are similar to the input of the distribution-based approach by [Feit et al. 2017]. We add the gaze estimates in world coordinates to cover potential systematic weaknesses of the tracking device, e.g., in regions close to the border of the camera's field of view. In contrast to all other methods, the spatial accuracy is modelled including its directional information which allows to shift the measured gaze towards the actual point of regard (see Figure 3c). Similar to the model above, we consider a weight factor ω for scaling the estimated spatial precision. The target size is computed as $S_{w/h}(\omega) = 2 \cdot (s_c + \omega \cdot \sigma_{w/h})$ with $2 \cdot s_c = d_{cal} \cdot \tan(1.23°)$ being the target size as computed for the baseline method *None* for the *Naïve* approach. We use this static base size as a lower bound, because we observed a significant plus in selection rates for near user positions. The participant's gaze is corrected as follows: $\textit{shift}(g_{x/y}, \mu_{w/h}) := g_{x/y} + \mu_{w/h}$.

We hypothesize that compensation with *PredictiveShift* achieves higher selection rates than *Naïve* and *Feit et al.* (H1.1), particularly enhancing the ratio of target size and selection rate (H4.1); personalised error models yield a further improvement (H4.2).

5.1 Model Training and Evaluation

For training and evaluating the compensation methods, we use the recorded interaction sessions from our user study. First, we preprocess the data extracting all relevant information and excluding outliers. The subsequent steps are based on this cleaned dataset.

Table 1: Mean selection rate and target size from the offline simulation of compensation methods from the user study.

	Best	None	Measured	Predictive
Selection Rate [%]	21.37	54.19	52.99	83.76
Target Size [cm^2]	16.33	55.24	68.62	170.36

5.1.1 Pre-Processing. The recorded dataset includes raw and meta data of all selection trials from our study. We extract the relevant parts of each selection sequence by cropping the signals starting half a second before the selection was triggered (presenter click) and stopping at that event. Outliers are removed based on the mean gaze offset to the actual target centre and on the distribution of the standard deviations of the gaze signal for the cropped intervals. We drop a selection trial if the mean offset is greater than 5 degrees [Kassner et al. 2014] and if the standard deviation is greater than the 95^{th} percentile of all standard deviations. High offsets commonly appear, when marker tracking fails or when the eye tracking headset is displaced. High variance values can occur, when the user's pupil cannot be tracked well. In total, we drop 8.80% of the data totalling in 2364 selections from 12 users. On average, the dataset contains 197 selection trials for each user ($SD = 13.93$).

5.1.2 Model Training. The error estimation of the applied compensation methods are data driven, hence, require a training phase which is described here. In general, we split the data into a train (75%) and a test-set (25%). Concerning the user-specific models for *PredictiveShift(personal)*, we split the data per user with the same ratio. For *Feit et al.*, we compute the means μ and standard deviations σ on the train-set as described above, and store them in a look-up table. For our new predictive model, we conduct a 5×5 nested crossvalidation (k-fold) on the train-set where the inner loop performs a grid search for optimising model parameters of the ElasticNet algorithm. The search space includes the degree of polynomial features $\in \{1, 2\}$, the factor alpha$\in \{10, 1, .1, .001\}$ weighting the penalty terms and the l1_ratio $\in \{0, .5, 1\}$.

5.1.3 Evaluation Procedure. We use the selection trials from the test-set for validating all compensation methods including *Naïve* with all error models, *Feit et al.*, *PredictiveShift* and its personalised version. For each selection and compensation method, we infer the gaze error using the recorded signals as input. This estimate is used for computing the target size and, in case of *PredictiveShift*, for shifting the POG (see Figure 3). The selection success and the respective target size is logged. We repeat the evaluation cycle with 20 values of ω weighting the influence of the precision estimate on the target size. The interval range and step-size is chosen, starting with $\omega = 0$ (no influence), such that selection rates of models converge to 100%. For all methods, this results in a maximum target size between 300 and 350cm^2 (see Figure 6). We consider $\omega \in \{0 \leq i < 3\}$ with step-size 0.15 for *Feit et al.* and $\omega \in \{0 \leq i < 20\}$ with step-size 1 for *PredictiveShift* and *PredictiveShift(personal)*.

5.2 Results

We average the selection rate and the target area over all on-screen targets and grid positions. Figure 6 visualises the mean selection

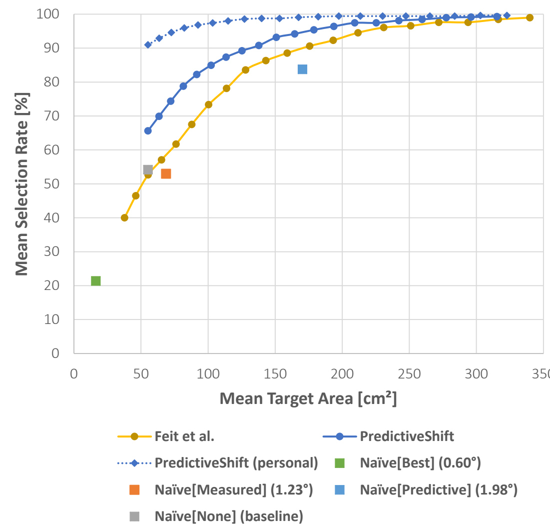


Figure 6: Mean selection rate off all methods in relation to the average target sizes. For the new compensation methods, we include the results for varying ω.

rates in relation to their mean target size for each method. This allows a direct comparison with our previous results: the four squares represent the offline simulation results of the *Naïve* scaling using different error models, confirming our previous findings (see Table 1 and Figure 4). The results of the other compensation methods are plotted as curves, due to the variable parameter ω. Tests for statistical significance are conducted using McNemar's test for selection rate and the Wilcoxon signed-rank test for target sizes.

For $\omega = 0$, the compensation method of Feit et al. [Feit et al. 2017] achieves a selection rate of 40% with an average target size of 37.74cm^2. When increasing ω by two steps to 0.3, it achieves a similar selection rate and target size than the baseline method *Naïve[None]*: 52.65% at 55.27cm^2. The selection rate further improves with greater values for ω, but the ratio to the target area decreases (i.e., the slope of the curve is regressive). Hereby, this method reaches a similar selection rate than *Naïve[Predictive]* at $\omega = 1.2$ with 83.59%, but with significantly smaller target sizes of, on average, 128cm^2 ($Z = -14.6, p < .001$). In the same way, it outperforms *Naïve[Predictive]* with a similar target size (175.7cm^2) in terms of selection rate: 90.6% at $\omega = 1.65$ ($p < .001$).

With $\omega = 0$, *PredictiveShift* generates targets of the same size than the baseline *Naïve[None]*, which is used for computing its base size s_c (see Figure 3c). However, shifting the gaze by the predicted directional gaze error increases the selection rate by 11.45% to 65.64% ($p < .001$). Similar to the *Feit et al.* method, with increasing ω, our novel method achieves higher selection rates with a regressive slope. Particularly for small targets, our method *PredictiveShift* achieves higher selection rates. *PredictiveShift(personal)* achieves the best selection rates and target sizes. For $\omega = 0$, the selection rate is 90.98%: it significantly improves by 25.34% compared to the non-personalised version and is 36.79% better than the baseline *Naïve[None]* ($p < .001$). In addition, it exceeds the selection rate of *Naïve[Predictive]*, despite significantly smaller target sizes by

67.57% ($Z = -20.45, p < .001$). Similar to the other approaches, the selection rate increases as ω grows with a regressive slope. For $\omega = 10$, the personalised compensation method achieves a selection rate of 98.01% with target sizes around $165.82cm^2$ that are comparable to *Naïve[Predictive]* ($p < .001$).

6 DISCUSSION

In this work, we show that the performance of mobile gaze interaction can be significantly improved, if the interface is aware of and can adapt to the inevitable gaze estimation error. The error compensation method *PredictiveShift* consequently achieves better selection rates than our initial *Naïve* scaling approach and excels the method based on [Feit et al. 2017] (supports H1.1). The results of our evaluation show that, given a certain selection rate, all new compensation approaches generate smaller targets (supports H4.1). Especially, the user-specific training with our shift-based compensation method, *PredictiveShift(personal)*, increases the performance measures tremendously without increasing the target size (supports H4.1 and H4.2). We conclude that personalised models better explain the variation between user, e.g., covering differences in tracking quality and in how they interact using gaze. The decreasing slope can be explained by the fact that the error might be normally distributed. Increasing the selection rate beyond a certain point comes at the cost of larger selection targets that grow quadratically in area. To put our results into context, we approximately measured the area of common controls of the Windows 10 user interface for the display setting as shown in Figure 1 with a size of 115.32×86.49 cm: Selecting taskbar icons ($19cm^2$) and small tiles in the start menu ($32cm^2$) would still be difficult; mid-sized tiles ($170cm^2$) and larger controls would achieve high selection rates. Next, we discuss use cases, advantages and short-comings of our approach.

6.1 Advantages and Use Cases

The key advantage of our error-aware interface is its application to enhance mobile gaze-based interfaces. The evaluation shows benefits when using sophisticated compensation methods with error models in selection performance and generated target sizes. To the best of our knowledge, this is the first attempt to apply such a model in real-time interaction. A direct application of our methods is the extension of the interaction framework by debugging components that visualise the gaze error just-in-time, and a simulation component that allows early assessment of gaze-based user interface prototypes. We suggest two data visualisation techniques and describe the simulation capability with our error models similar to [Feit et al. 2017], but extending it to mobile settings with varying user positions. Both visualisation methods are implemented and informally tested with regard to real-time interface analysis.

6.1.1 Uncertainty Indicator. Hereby, the gaze pointer is augmented by a transparent ellipse indicating the estimated distribution of the gaze error. Therefore, the two-dimensional spatial accuracy and spatial precision estimate of any of the presented error models is used to span an ellipse around the fixation position.

6.1.2 Heatmap Overlay. For this approach, a regular grid ($N \times M$) is laid over the display space and mapped back to the world camera space (with an inverse display mapping). The error of the tracking device is estimated for each position. A texture with the resolution $N \times M$ is generated where each pixel corresponds to one point of the grid. For each point, the colours are assigned based on the error value. Finally this texture is mapped to the display region on the world camera stream which enables an immediate feedback for a considered interface from the user's egocentric perspective.

6.1.3 Gaze-interface Design Tool. Simulations have significant potential for interaction designers for optimising interfaces, interaction techniques and visualisations, without the need to test each variant through actual human studies. Our framework enables an assessment at the design stage similar to Feit et al. [Feit et al. 2017]. Each interactive control can be compared to target sizes generated by our compensation methods for identifying potential issues. Using our method, different user positions in a mobile interaction setting with given display positions and sizes can be covered.

6.2 Limitations

Despite its novelty in terms of error-awareness, our gaze-based interface has some limitations. Currently, we restrict the number and type of interface elements to a single button that can be triggered. To showcase an error-aware gaze-based interface, this button expands in size for high-error situations and optionally shifts gaze to ease selection. Currently, there is no efficient method for collecting required training data, but the best ratio of selection rate and target size is achieved with our personalised compensation approach.

7 CONCLUSION

We introduced an error-aware gaze interaction framework. This framework enables a new class of gaze-based interfaces that are aware of the gaze estimation error. Driven by real-time error estimation this approach has the potential to outperform state-of-the-art gaze selection methods in terms of selection performance with competing target sizes. We presented a first implementation of the framework and evaluated a naïve target scaling approach with four methods to estimate the gaze error in a user study. Elaborating on our findings, we developed and compared advanced error compensation methods. The results show both the real-time capability as well as the advantages of an error-aware interface, relying on gaze shifting and personalised training with a predictive model, in terms of selection performance and target sizes.

We plan to further extend and evaluate the concept of error-aware gaze-based interaction in future work. As a first step we want to add and compare further mechanics to adapt the interface according to the gaze estimation error. One idea would be to move small objects to low error regions. This could, for example, help to automatically locate fine- and coarse-grained gaze-based interaction on near and far displays, respectively. Further, it will be interesting to explore techniques for seamless collection of training samples for training new error models or fine-tuning them for new users, e.g., as part of a calibration routine or fully automatic.

ACKNOWLEDGMENTS

This work was funded, in part, by the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University, Germany, and by the Federal Ministry of Education and Research (BMBF) under grant number 16SV7768 in the Interakt project.

REFERENCES

Stanislavs Bardins, Tony Poitschke, and Stefan Kohlbecher. 2008. Gaze-based Interaction in Various Environments. In *Proceedings of the 1st ACM Workshop on Vision Networks for Behavior Analysis (VNBA ’08)*. ACM, New York, NY, USA, 47–54. https://doi.org/10.1145/1461893.1461903

Michael Barz, Andreas Bulling, and Florian Daiber. 2016. Prediction of Gaze Estimation Error for Error-Aware Gaze-Based Interfaces. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’16)*. ACM, New York, NY, USA.

Michael Barz, Peter Poller, and Daniel Sonntag. 2017. Evaluating Remote and Head-worn Eye Trackers in Multi-modal Speech-based HRI. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Bilge Mutlu, Manfred Tscheligi, Astrid Weiss, and James E Young (Eds.). ACM, New York, NY, USA, 79–80. http://doi.acm.org/10.1145/3029798.3038367

Michael Barz and Daniel Sonntag. 2016. Gaze-guided object classification using deep neural networks for attention-based computing. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp ’16*. ACM Press, New York, New York, USA, 253–256. https://doi.org/10.1145/2968219.2971389

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*. Springer, 77–100. https://doi.org/10.1007/978-3-319-02868-2_5

Pieter Blignaut and Daniël Wium. 2013. The Effect of Mapping Function on the Accuracy of a Video-based Eye Tracker. In *Proceedings of the 2013 Conference on Eye Tracking South Africa (ETSA ’13)*. ACM, New York, NY, USA, 39–46. https://doi.org/10.1145/2509315.2509321

Jurek Breuninger, Christian Lange, and Klaus Bengler. 2011. Implementing Gaze Control for Peripheral Devices. In *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-based Interaction (PETMEI ’11)*. ACM, New York, NY, USA, 3–8. https://doi.org/10.1145/2029956.2029960

Andreas Bulling, Daniel Roggen, and Gerhard Tröster. 2008. EyeMote — Towards Context-Aware Gaming Using Eye Movements Recorded from Wearable Electrooculography. In *Proceedings of the 2Nd International Conference on Fun and Games*. Springer-Verlag, Berlin, Heidelberg, 33–45. https://doi.org/10.1007/978-3-540-88322-7_4

Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1 Euro Filter: A Simple Speed-based Low-pass Filter for Noisy Input in Interactive Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’12)*. ACM, New York, NY, USA, 2527–2530. https://doi.org/10.1145/2207676.2208639

Juan J. Cerrolaza, Arantxa Villanueva, Maria Villanueva, and Rafael Cabeza. 2012. Error Characterization and Compensation in Eye Tracking Systems. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’12)*. ACM, New York, NY, USA, 205–208. https://doi.org/10.1145/2168556.2168595

Jan Drewes, Guillaume S. Masson, and Anna Montagnini. 2012. Shifts in Reported Gaze Position Due to Changes in Pupil Size: Ground Truth and Compensation. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’12)*. ACM, New York, NY, USA, 209–212. https://doi.org/10.1145/2168556.2168596

Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST ’15*. ACM Press, New York, New York, USA, 457–466. https://doi.org/10.1145/2807442.2807499

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 - CHI ’17*. ACM Press, New York, New York, USA, 1118–1130. https://doi.org/10.1145/3025453.3025599

Jeremy Hales, David Rozado, and Diako Mardanbegi. 2011. Interacting with Objects in the Environment by Gaze and Hand Gestures. In *Proceedings of the 3rd International Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction*. 1–9.

Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 152–169. https://doi.org/10.1145/123078.128728

Samuel John, Erik Weitnauer, and Hendrik Koesling. 2012. Entropy-based Correction of Eye Tracking Data for Static Scenes. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’12)*. ACM, New York, NY, USA, 297–300. https://doi.org/10.1145/2168556.2168620

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp ’14 Adjunct)*. ACM, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

Christian Lander, Sven Gehring, Antonio Krüger, Sebastian Boring, and Andreas Bulling. 2015a. GazeProjector: Accurate Gaze Estimation and Seamless Gaze Interaction Across Multiple Displays. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST ’15)*. ACM, New York, NY, USA, 395–404. https://doi.org/10.1145/2807442.2807479

Christian Lander, Marco Speicher, Denise Paradowski, Norine Coenen, Sebastian Biewer, and Antonio Krüger. 2015b. Collaborative Newspaper: Exploring an Adaptive Scrolling Algorithm in a Multi-user Reading Scenario. In *Proceedings of the 4th International Symposium on Pervasive Displays (PerDis ’15)*. ACM, New York, NY, USA, 163–169. https://doi.org/10.1145/2757710.2757734

Diako Mardanbegi and Dan Witzner Hansen. 2011. Mobile Gaze-based Screen Interaction in 3D Environments. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications (NGCA ’11)*. ACM, New York, NY, USA, Article 2, 4 pages. https://doi.org/10.1145/1983302.1983304

Diako Mardanbegi and Dan Witzner Hansen. 2012. Parallax Error in the Monocular Head-mounted Eye Trackers. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp ’12)*. ACM, New York, NY, USA, 689–694. https://doi.org/10.1145/2370216.2370366

Darius Miniotas, Oleg Špakov, and I. Scott MacKenzie. 2004. Eye Gaze Interaction with Expanding Targets. In *CHI ’04 Extended Abstracts on Human Factors in Computing Systems (CHI EA ’04)*. ACM, New York, NY, USA, 1255–1258. https://doi.org/10.1145/985921.986037

A. Monden, K. Matsumoto, and M. Yamato. 2005. Evaluation of Gaze-Added Target Selection Methods Suitable for General GUIs. *Int. J. Comput. Appl. Technol.* 24, 1 (June 2005), 17–24. https://doi.org/10.1504/IJCAT.2005.007201

Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost van de Weijer. 2013. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods* 45, 1 (2013), 272–288. http://dx.doi.org/10.3758/s13428-012-0247-4

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

Jeffrey S. Shell, Roel Vertegaal, and Alexander W. Skaburskis. 2003. EyePliances: Attention-seeking Devices That Respond to Visual Attention. In *CHI ’03 Extended Abstracts on Human Factors in Computing Systems (CHI EA ’03)*. ACM, New York, NY, USA, 770–771. https://doi.org/10.1145/765891.765981

Daniel Sonntag. 2015. Kognit: Intelligent Cognitive Enhancement Technology by Cognitive Models and Mixed Reality for Dementia Patients. (2015). https://www.aaai.org/ocs/index.php/FSS/FSS15/paper/view/11702

Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch: Gaze-supported Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’12)*. ACM, New York, NY, USA, 2981–2990. https://doi.org/10.1145/2207676.2208709

Sophie Stellmach, Sebastian Stober, Andreas Nürnberger, and Raimund Dachselt. 2011. Designing Gaze-supported Multimodal Interactions for the Exploration of Large Image Collections. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications (NGCA ’11)*. ACM, New York, NY, USA, Article 1, 8 pages. https://doi.org/10.1145/1983302.1983303

Takumi Toyama, Thomas Kieninger, Faisal Shafait, and Andreas Dengel. 2012. Gaze Guided Object Recognition Using a Head-mounted Eye Tracker. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’12)*. ACM, New York, NY, USA, 91–98. https://doi.org/10.1145/2168556.2168570

Takumi Toyama and Daniel Sonntag. 2015. Towards episodic memory support for dementia patients by recognizing objects, faces and text in eye gaze. *KI 2015: Advances in Artificial Intelligence* 9324 (2015), 316–323. https://doi.org/10.1007/978-3-319-24489-1_29

Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-device Gaze-supported Point-to-point Content Transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’14)*. ACM, New York, NY, USA, 19–26. https://doi.org/10.1145/2578153.2578155

Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp ’13)*. ACM, New York, NY, USA, 439–448. https://doi.org/10.1145/2493432.2493477

Oleg Špakov. 2011. Comparison of Gaze-to-objects Mapping Algorithms. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications (NGCA ’11)*. ACM, New York, NY, USA, Article 6, 8 pages. https://doi.org/10.1145/1983302.1983308

Oleg Špakov. 2012. Comparison of Eye Movement Filters Used in HCI. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’12)*. ACM, New York, NY, USA, 281–284. https://doi.org/10.1145/2168556.2168616

Oleg Špakov and Yulia Gizatdinova. 2014. Real-time Hidden Gaze Point Correction. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA ’14)*. ACM, New York, NY, USA, 291–294. https://doi.org/10.1145/2578153.2578200

Lawrence H Yu and E Eizenman. 2004. A new methodology for determining point-of-gaze in head-mounted eye tracking systems. *Biomedical Engineering, IEEE Transactions on* 51, 10 (Oct 2004), 1765–1773. https://doi.org/10.1109/TBME.2004.831523

Xinyong Zhang, Xiangshi Ren, and Hongbin Zha. 2008. Improving Eye Cursor’s Stability for Eye Pointing Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’08)*. ACM, New York, NY, USA, 525–534. https://doi.org/10.1145/1357054.1357139

Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2013. SideWays: A Gaze Interface for Spontaneous Interaction with Situated Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 851–860. https://doi.org/10.1145/2470654.2470775

Yunfeng Zhang and Anthony J Hornof. 2011. Mode-of-disparities error correction of eye-tracking data. *Behavior research methods* 43, 3 (September 2011), 834–842. https://doi.org/10.3758/s13428-011-0073-0

Yunfeng Zhang and Anthony J. Hornof. 2014. Easy Post-hoc Spatial Recalibration of Eye Tracking Data. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14)*. ACM, New York, NY, USA, 95–98. https://doi.org/10.1145/2578153.2578166

Yanxia Zhang, Jörg Müller, Ming Ki Chong, Andreas Bulling, and Hans Gellersen. 2014. GazeHorizon: Enabling Passers-by to Interact with Public Displays by Gaze. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 559–563. https://doi.org/10.1145/2632048.2636071