Weighted Pointer: Error-aware Gaze-based Interaction through Fallback Modalities

Ludwig Sidenmark, Mark Parent, Chi-Hao Wu, Joannes Chan, Michael Glueck, Daniel Wigdor, Tovi Grossman and Marcello Giordano

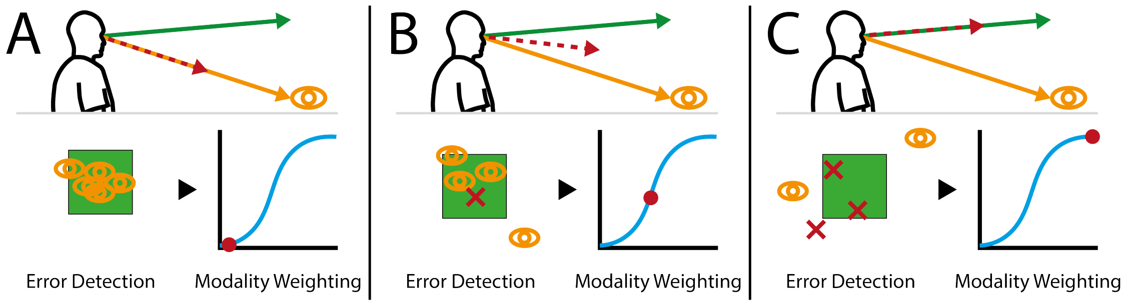

Fig. 1. The Weighted Pointer provides stable gaze-based pointing interaction irrespective of gaze signal quality via fallback modalities. The techniques detect the eye tracking error level and determines the relative weighting to assign between gaze and the fallback modality. A: The pointer direction (red) is aligned with gaze (orange) when the gaze signal is stable. B: The pointer relies on both gaze and the fallback modality (e.g., head direction, green) if the gaze signal deteriorates. C: The pointer fully relies on the fallback modality if gaze is deemed to be unusable for interaction.

Abstract—Gaze-based interaction is a fast and ergonomic type of hands-free interaction that is often used with augmented and virtual reality when pointing at targets. Such interaction, however, can be cumbersome whenever user, tracking, or environmental factors cause eye tracking errors. Recent research has suggested that fallback modalities could be leveraged to ensure stable interaction irrespective of the current level of eye tracking error. This work thus presents Weighted Pointer interaction, a collection of error-aware pointing techniques that determine whether pointing should be performed by gaze, a fallback modality, or a combination of the two, depending on the level of eye tracking error that is present. These techniques enable users to accurately point at targets when eye tracking is accurate and inaccurate. A virtual reality target selection study demonstrated that Weighted Pointer techniques were more performant and preferred over techniques that required the use of manual modality switching.

Index Terms—Eye tracking, Gaze interaction, Virtual Reality, Adaptive interfaces, Accessibility

1 Introduction

Eye tracking is becoming an increasingly prevalent input modality for use within augmented and virtual reality applications (AR/VR) when pointing at targets due to its speed and because users naturally look at objects that they are considering for manipulation. Gaze-based pointing has thus been proposed as an attractive modality for interface control [13, 37, 64, 67]. Tracking gaze alignment over a target, however, is not precise or easy. Firstly, the fovea is 1° in diameter, limiting the accuracy possible with eye trackers [38]. Additionally, as today's camera-based eye trackers require a calibration process to map eye images to gaze directions, whenever there are poor or missing calibrations, *accuracy errors* (i.e., the difference between the actual and recorded gaze directions) could be significant and shift gaze directions outside of, or to, a different target than the one intended. Furthermore, because the eyes are continuously moving, they introduce a natural level of noise within gaze signals, which can become exacerbated by tracking difficulties [27]. Such *precision errors* can impact pointing whenever jitter occurs and may also impact algorithms that rely on dispersion or velocity thresholds to detect gaze movements [58] by forcing higher threshold values [27, 28]. Finally, as eye trackers require a clear view of the eyes, user and environmental factors (e.g., lighting, glasses, lazy eyes) can cause significant issues with precision and accuracy, and even *data loss* [27], thus making interaction difficult, if not impossible [27, 28, 48]. In light of these challenges, the use of a solely gaze-based pointer within AR or VR can thus be problematic.

To address these limitations, several projects have proposed the inclusion of a secondary modality to refine inaccurate gaze directions [37, 54, 66, 69] or disambiguate one target from many in situations where there is uncertain gaze input [45, 62]. However, these techniques require users to manually detect and refine errors that are present. Alternatively, gaze-based *error-aware* systems have also proposed adjusting one's gaze position to account for accuracy errors [7] or to increase the size of a target to overcome poor signal quality [19]. Although these approaches can be useful, they all assume that there will be a stable output of gaze samples that can be used to (i) accurately adjust the signal or interface or (ii) be used in combination with a second modality to refine or disambiguate one's pointing position. However, as stable eye tracking output is not guaranteed for all users or contexts, there is a need for techniques that can provide stable output even in situations with significant error or data loss. This is especially relevant for AR where environmental conditions can change significantly.

This research explores the use of fallback modalities to address all levels of eye tracking error during target selection and to make gaze-based systems more accessible to users who struggle with eye tracking. We introduce Weighted Pointer pointing, a set of techniques that determines whether the user should rely on gaze, the fallback modality, or a combination of both while pointing based on the current level of gaze signal error. Using these techniques, a user's pointer position will be fully dictated by gaze if the gaze signal is deemed accurate and

- Ludwig Sidenmark is with Reality Labs Research and Lancaster University. E-mail: l.sidenmark@lancaster.ac.uk.
- Mark Parent, Chi-Hao Wu, Joannes Chan, Michael Glueck, Daniel Wigdor and Marcello Giordano are with Reality Labs Research. E-mail: {mrkprnt, eddywu, joanneschan, mglueck, dwigdor, giordanom}@fb.com.
- Tovi Grossman is with University of Toronto. E-mail: tovi@dgp.toronto.edu.

stable enough or, as the gaze signal deteriorates, it will increasingly rely on the fallback modality until the pointer position is completely dictated by the fallback modality. To support such an interaction, the Weighted Pointer techniques consist of three separate and replaceable components that (1) detect the current level of eye tracking error, (2) calculate the weightings between gaze and the fallback modality, and (3) leverage the weightings and output signals to enable pointing at any level of eye tracking error (Figure 1). Multiple Weighted Pointers that utilized varying fallback modalities were implemented and evaluated within a VR target selection study that manipulated the levels of eye tracking error and compared the new techniques to existing manual techniques, e.g., *Eye+Device* and *Eye+Head Pinpointing* [37].

This research thus contributes (1) error-aware gaze-based pointing techniques that support stable pointing interaction irrespective of gaze tracking quality through the adaption of fallback modalities for wearable AR and VR, and (2) the results from a user study that evaluated these techniques compared to manual refinement techniques. The study demonstrated that participants performed at equal or significantly higher performance compared to the manual techniques and that our techniques were preferred by the majority of participants. The results highlight the benefits of automating switching between modalities to increase user performance and experiences while working with multimodal interfaces.

2 Related Work

Gaze has been studied since the eighties for hands-free interaction [29,30,75]. For the design of the Weighted Pointer we are building on insights on handling eye tracking error, as well as prior work on multimodal interaction with gaze.

2.1 Handling Eye Tracking Error

Measuring and handling gaze signal error is an important aspect of eye tracking and gaze-based interaction. Within this context, *accuracy* refers to the average difference between the true and measured gaze direction. *Precision*, on the other hand, is defined as how consistent measured gaze directions are while the true gaze direction is constant and is usually measured via the root mean square (RMS) of the intersample differences [27]. Finally, *data loss*, i.e., when a tracker is unable to output a gaze direction, can be caused by a poorly aligned tracker, user factors such as eye-shape, glasses, and makeup, or by the user closing their eyes. Data loss is also commonly accompanied by tracking artifacts that cause significant spikes in a gaze signal [27]. All three of these factors can have a significantly negative impact on gaze-based interaction and may vary across tracking areas [19] or over time due to changing lighting conditions, pupil size, or slippage of a head-mounted display (HMD) in AR/VR settings [16,27,47].

Much research has proposed various gaze-based techniques to address the impact of eye tracking error on interaction. Poor accuracy can be addressed by calibration processes, where points across a tracking area are mapped to eye images through explicit [18, 20, 21, 27, 55] or implicit [18,65] procedures. Poor precision can be addressed via filtering [10, 19, 71]. Gaze-based systems have also used algorithms to map gaze positions and sequences to objects when input is uncertain [9, 41, 52, 70, 72]. There have also been numerous gaze-based interaction techniques that have been proposed to improve interaction by increasing target or cursor sizes via zooming [12,39], the use of area cursors [12], cursors that can be nudged via gaze-based buttons [56], or by the incremental disambiguation of possible targets [42]. However, these strategies do not guarantee the removal of tracker errors as demonstrated by prior work where participant data was discarded due to calibration and tracking issues [1, 4, 15, 19, 45, 48, 50, 53, 59]. Thus, this research proposes the use of fallback modalities to make gaze-based systems more robust and accessible to users.

2.2 Refinement and Disambiguation

The most common metaphor for selection in 3D environments is *Raycasting* where the user controls a ray via a controller or body part [26,51]. However, Raycasting can be challenging for small targets due to occlusion or pointing inaccuracy caused by tracking issues or user limitations (e.g. shaking hands). As such, researchers have developed a variety of target disambiguation techniques that requires additional manual steps for final selection [5,23,35,77] or that apply contextual information or heuristics for implicit disambiguation [23,24,60,68]. However, while these techniques help users perform difficult selections, many assume a stable signal to disambiguate target candidates.

For gaze interaction, research has investigated the refinement of inaccurate gaze cursors through the use of a second, more accurate, modality such as head movements [36,73], touch interaction [54,69] or hand movements [18,79,80]. Furthermore, researchers have proposed techniques that disambiguate one target out of many candidates via head [74] or hand gestures [11]. In AR and VR, gaze refinement has been investigated via head or controller movements [31, 37], or by gestural head movements that are detected from eye-head coordination insights [66]. These techniques, however, require the user to manually detect and switch modalities, thus increasing their task load, or utilize automatic switching via velocity-based thresholds, which may be prone to errors caused by poor precision or data loss. Researchers have also proposed VR disambiguation techniques where object decluttering [14] or relative eye movements that do not rely on calibration [45, 62] are used when the gazed-upon target is uncertain. However, these techniques require changes to one's environment or rely on low precision and data loss errors to accurately detect relative eye movements. The present research proposes interaction techniques that automate the decision to use a fallback modality to lower user workload, while also not relying on metrics which are sensitive to tracking errors or require changes to one's environment.

2.3 Error-aware Interaction

Interaction issues caused by poor signal quality have led to the development of *error-aware* interactive systems that account for ambiguous input caused by sensor limitations or tracking difficulties. In touch-based interaction, probabilistic approaches have been used to tackle uncertain input caused by the *fat finger problem* or gesture-recognition difficulties [43, 44, 61]. Furthermore, VR and AR researchers have leveraged the multiple available modalities (e.g. speech, hand gestures, head direction) by combining them to more robustly infer the most probable action performed by the user [32,40,49].

Error-aware gaze-based systems store the current eye tracking calibration data that contains measurements of all gaze signal errors. These records are then used to adjust interaction by, for example, correcting the gaze direction [6,7,18] or by enlarging user interface widgets that are positioned at areas where the eye tracking signal is poor [19]. Alternatively, erroneous interactions are detected through brain-computer interfaces and remedied [33]. These approaches assume a certain level of accuracy within the eye tracking signal from which one's gaze direction can be corrected or the widget can be selected. To our knowledge, there are no error-aware systems that can ensure stable interaction irrespective of gaze signal quality, even when the signal quality has completely deteriorated. Thus, this research investigates error-aware interaction with a focus on how fallback modalities can be deployed in conjunction with gaze or as a replacement for gaze when the gaze signal is too poor for standalone use.

3 Weighted Pointer

The Weighted Pointers are a set of generic ray-casting techniques designed for pointing-based interaction while using HMDs within AR and VR. The core concept of Weighted Pointer is to determine whether interaction should rely on gaze, a fallback modality, or a combination of the two depending on the gaze-signal quality. Four variations that relied on different fallback modalities were implemented, i.e., head, controller, relative controller movements, and a combination of relative controller and head movements. We used head and controller techniques as fallbacks since they are commonly available in modern VR/AR kits. Furthermore, both modalities constantly coordinate with the eyes during movement which should make hand or head pointing in conjunction with gaze feel natural to users [63, 65]. The design of the Weighted Pointer techniques was based on four goals. First, the techniques should provide stable interaction irrespective of the eye tracking error

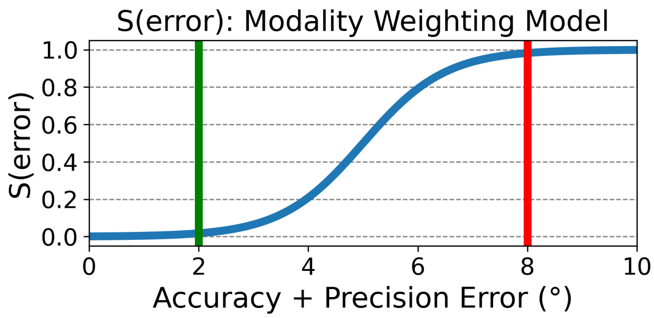

Fig. 2. Equation 1 used for modality weighting visualized. The green line represents $E_{acc} = 2°$ and the red line represents $E_{max} = 8°$. The function will weigh more heavily towards gaze if the error is below E_{acc}, otherwise it will weigh more towards the fallback modality as the error increases.

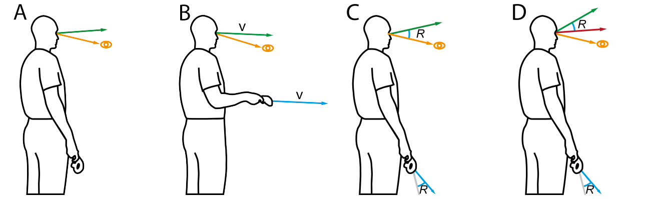

Fig. 3. The Weighted Pointer variations all use gaze (orange) as the main modality. (a) Hands-free uses the head direction as the fallback modality (green). (b) In Controller-based, the fallback direction is based on the controller direction (blue), originating from the head position. (c) In Relative, a rotational offset R is calculated from the controller direction (blue) and a calibrated neutral direction (grey), and is then applied to the gaze direction. (d) In Trimodal, the same offset is used as in (c), but is applied to the head direction (red).

level, including when the eye tracker is unusable. Second, the switch between modalities is automatic rather than manual as is common in prior techniques. Thirdly, determining whether a gaze signal is too inaccurate or noisy for interaction should depend on current interaction needs (e.g. target sizes). As such, the techniques should accommodate an application's required level of interaction granularity. Finally, as the movements performed by the user may be different when using gaze or the fallback modality, the user should be able to understand whether they are currently relying on gaze or the fallback modality to interact. Next, we describe the general role of each component that is needed to help Weighted Pointer achieve these goals.

3.1 Error Detection Model

The first component of Weighted Pointer calculates and stores the current accuracy, precision and data loss error of the gaze signal. The stored metrics are then retrieved during runtime to compute a weighting between the gaze and the fallback modality. The Error Detection Model is based on an eye tracking calibration used in prior error-aware research [7,18,19]. Previous work that adjusted gaze positions required knowledge about the errors' direction for accurate adjustments [7, 18], however, our reliance on fallback modalities lowers the required granularity as we only need to know whether gaze errors are present and their magnitude. This simplifies the required calculations. Within our implementation, the HMD's screen area is divided into an $n \times m$ grid (i.e., one cell per calibration point) to account for changes in different parts of the tracking area. Accuracy and precision measures from the latest calibration are then stored within each cell and updated each time the user performs the calibration procedure. Accuracy is measured as the average angular difference between the recorded gaze position and the calibration point position. Precision is calculated as the RMS between eye tracking samples while the user is fixating on the calibration point. During application use, the stored error information is retrieved from the cell closest to the current gaze direction. To measure data loss, the ratio of missing data versus all data points is continuously calculated using a rolling window of N_l frames. Accompanying artifacts are removed and treated as missing data. A filter removes data points with an angular gaze velocity higher than v_{max}, which is higher than what is physically possible [27].

3.2 Weighting Model

The Weighting Model retrieves the information stored by the Error Detection Model to decide whether the interaction should rely on gaze or the fallback modality. In the implementation, the Weighting Model is a logistic sigmoid function $S(error)$ that outputs a value between $0-1$ that decides the relative weighting between the gaze and fallback modality (Equation 1). The gaze weighting, W_G, is then defined as $W_G = 1 - S(error)$ and the fallback modality weighting, W_F, is defined as $W_F = S(error)$. As input for Equation 1 ($error$), the system retrieves the sum of the accuracy and precision data from the Error Detection Model's grid cell that is closest to the latest valid gaze point during each frame. As the accuracy and precision necessary for stable interaction can vary by task, context, and environment, the weighting function is defined by two adjustable parameters. E_{acc} defines the maximum eye tracking error that will not affect interaction. The interaction should rely solely on gaze for any error at or below this level. E_{max} defines the upper limit at which point the eye tracker is deemed to be usable. The interaction should fully rely on the fallback modality when the error exceeds this limit. G defines the gradient of the curve between E_{acc} and E_{max}. For example, $S = 8$ means that 97.5% of the possible values are between E_{acc} and E_{max}. The sigmoid curve is then adjusted so that the interaction relies on a combination of gaze and fallback modality for any error value between these parameters, with it relying more on the fallback modality as the error reaches E_{max} (Figure 2). A combination of gaze and fallback is used for these moderate levels of error so that users can rely on gaze as much as possible, even though gaze alone may be too inaccurate for interaction. This allows gaze to be useful in situations with error present as gaze will point in the approximate target direction. The user can then make small corrections to minimize fallback movement [66,73], or lazily point with the fallback without requiring a high level of pointing accuracy [78]. Finally, data loss is accounted for by normalizing W_G and W_F based on the ratio of valid gaze samples from the last N_w frames.

$$
\begin{aligned}
S(error) &= \frac{1}{1+e^{Factor(error)}} \\
Factor(error) &= \frac{G}{(E_{max}-E_{acc}) \cdot (error + \frac{E_{acc}+E_{max}}{2})}
\end{aligned}
\tag{1}
$$

3.3 Pointing Techniques

$$
P = \frac{1}{N_p}\left(W_G \sum_{i=1}^{N_p} P_{G(i)} + W_F \sum_{i=1}^{N_p} P_{F(i)}\right) \tag{2}
$$

The final component of Weighted Pointer, the Pointing Techniques, retrieves the latest weightings from the Weighting Model to calculate and present pointing output to the user. The latest gaze and fallback directions are added to separate rolling windows (i.e., P_G and P_F respectively) of size N_p. The rolling windows add a level of smoothing for situations with high precision error and to minimize the impact of short periods of data loss. The pointing direction (i.e., P) is then calculated as the sum of the means from each rolling window adjusted by the latest weightings (Equation 2). As such, the pointer direction is aligned with gaze when the gaze signal is stable and will rely more on the fallback as the signal deteriorates. The continuous spectrum between gaze and the fallback modality allows users to rely on gaze in situations where the signal would be unusable in a gaze-only system, while also allowing lazy pointing with the fallback modality in cases where both gaze and fallback are used for interaction. As the choice of fallback modality can have a significant impact on the technique's behaviour and one's user experience, we developed four technique variations (Figure 3).

Hands-free Weighted Pointer uses head pointing as the fallback modality (Figure 3a) to allow for fully hands-free interaction. As the head is used as a natural part of our gaze movements [64], the transition from gaze- to head-pointing should be simple for users to understand as the pointer is guaranteed to always be within the field of view, and the head has been shown to be accurate for pointing [37,57].

Controller-based Weighted Pointer uses controller pointing as the fallback modality (Figure 3b). With this technique, the pointing origin is set to the gaze origin (head pose) to avoid confusion during modality transitions. Therefore, the technique only requires a 3 degree of freedom controller, making it somewhat reminiscent of the RCE technique [2]. We assume that stable tracking of the gaze origin can be achieved via head tracking.

Relative Weighted Pointer applies controller rotations that are relative to a neutral controller direction to the gaze direction as the fallback modality (Figure 3c). This is achieved by applying the rotational difference R between the current hand pose and a defined neutral hand pose to the gaze direction. The technique is designed to minimize controller movements to avoid arm fatigue caused by hand pointing [25]. With this technique, the user can perform small adjustments by rotating the controller as they hold it in a ”lazy” resting position by their side.

Trimodal Weighted Pointer uses relative controller movements similar to the Relative Weighted Pointer technique however with this technique, the rotational difference R is applied to the head direction as the fallback modality (Figure 3d). This allows for a combination of head and controller movements to be used to adjust the pointer direction. Large adjustments can be performed with less straining head and hand movements because the refinement is distributed to both modalities. Also, in contrast to the Relative Weighted Pointer technique, the fallback modality is not based on the gaze direction and should thus be more stable during high levels of precision and data loss.

Previous refinement techniques have used different approaches to visualize the current mode of interaction to the user, such as cursor colour changes [66] or by changing cursor appearance during refinement [37]. Weighted Pointer uses cursor transparency, where the cursor's alpha value is set to W_F. Through this design, the continuous spectrum between gaze and fallback can be displayed without relying on a cursor when interaction is purely gaze-based, which has been shown to be distracting for users [8]. Pointing feedback can then be displayed on the target (e.g., highlighting). Changes in cursor transparency and weightings can then be transitioned over $t_{transition}$ seconds via linear interpolation to avoid any abrupt changes caused by updated weightings.

4 Evaluation

A VR user study was conducted to gather insights into the proposed techniques' performance and to gather user feedback at different levels of eye tracking error, and investigate the effectiveness of using error-based modality switching compared to manual modality switching. Therefore, the *Eye+Head* and *Eye+Device* Pinpointing techniques were selected as baselines, wherein the switching from gaze pointing to refinement was performed via a button press and hold [37]. We selected the baselines due to their ability to handle both accuracy and precision error, and since the modality switch is performed manually in contrast to our techniques where the switch is performed by the system.

4.1 Participants

Twenty-four participants participated in the study (35±7 age; 15 male, 9 female) and were compensated for their participation. Informed consent was obtained and protocols were approved by the Western Institutional Review Board. Eighteen participants reported occasional VR experience, 5 participants reported weekly experience and 1 participant reported daily VR experience. Six participants reported no eye tracking experience, 15 reported occasional, and 3 reported weekly experience.

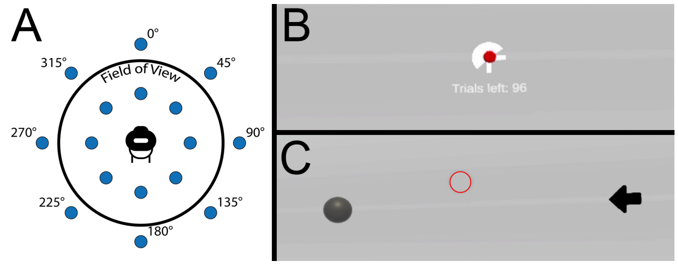

Fig. 4. Study Environment. A: Possible target positions. B: First person view of the study environment. The participant is aligning themselves against the central position. C: An arrow replacing the central position target points toward the next target to select.

4.2 Task and Procedure

Our study task was inspired by similar previous work [37,66]. During the experiment, participants were asked to select spherical targets with a diameter of 3° in visual degrees, from a 6-meter distance, at varying directions, amplitudes, and levels of accuracy and precision error. To begin a trial, participants aligned their gaze and head within 2° to a cross-shaped central target. After 1 second, a spherical target appeared in a predefined direction and amplitude, and an arrow directing participants towards the target replaced the central target. Targets were placed in one of eight compass directions to minimize bias by eye tracking error direction, and at two target amplitudes to cover target selection within and outside the initial field of view. Participants were instructed to select the target as quickly and accurately as possible. After selection, they would then return to the central starting position to start the next trial. Participants could not move on to the next trial until the object had been correctly selected or until a 10-second timeout had passed.

To investigate the effect of eye tracking precision and accuracy error, these error types were artificially injected into the eye tracker's gaze estimation. Accuracy was induced by adding a constant offset to the gaze direction in a random constant direction. Precision error was induced by adding Gaussian noise with a standard deviation that approximately resulted in an intersample RMS of a chosen value. The error conditions represented varying possible combinations, i.e., no significant error, significant accuracy error, significant precision error, and combined significant accuracy and precision error. Error magnitudes were set so that targets would be difficult to select without aid from the fallback modality (i.e., 5° for both injected accuracy and precision error). The study employed a within-subjects design, with the following independent variables and levels:

- Pointing **Technique**: **Weighted Pointers:** *Hands-free, Controller-based, Relative and Trimodal.* **Pinpointers:** *Eye+Head and Eye+Device.*
- Eye Tracking **Error**: *No Error, Accuracy error, Precision error, Both Accuracy and Precision error*
- Target **Direction**: *0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°*
- Target **Amplitude**: *30°, 60°*

At the beginning of the study, participants first signed a consent form and answered a demographic questionnaire. They then donned the head-mounted display in a standing position and performed blocks of selections where each block consisted of all trials for one pointing technique. Technique order was counterbalanced using a balanced Latin square. At the start of each block, participants calibrated the eye tracker via the standard Vive Pro Eye 5-point calibration process. Participants then performed a second calibration procedure using the GazeMetrics toolkit [3] to record eye-tracking data for the Error Detection Model. Participants were instructed to fixate on 9 calibration points, one for each grid cell in the detection model. To ensure equal conditions for all participants, the average sum of the recorded precision and accuracy

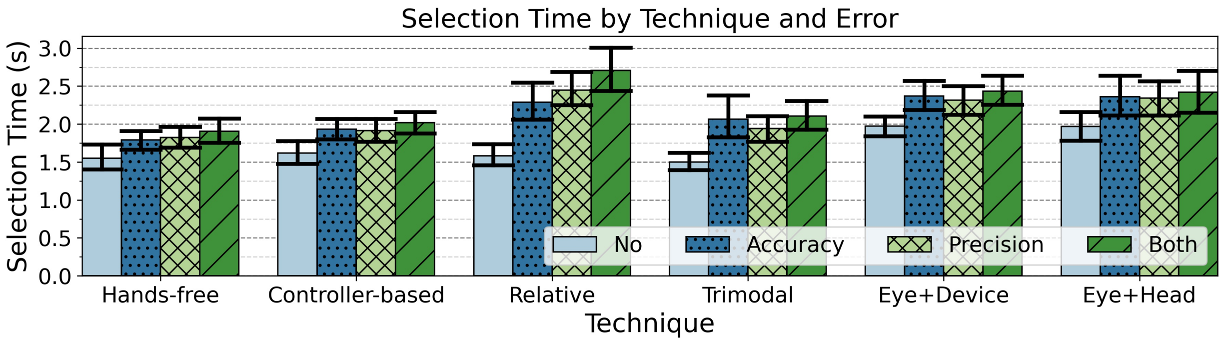

Fig. 5. Mean selection time for each technique by error condition. The error bars represent the 95% confidence intervals of the means.

errors during this process was not allowed to exceed 2°. If the error exceeded this limit, participants had to redo the process until they reached an error level below this threshold. Participants were then instructed to practice with the technique for a minimum of 15 trials before starting the experiment.

Within each technique block, trials were split into sub-blocks where participants performed all selections with a specific eye tracking error level before moving onto the next eye tracking error level. The eye tracking error level order within each technique was systematically altered to ensure even error ordering across participants. After completing a block with a technique, participants removed the head-mounted display and filled out a RAW NASA-TLX questionnaire to record the workload they experienced, and also completed a questionnaire, based on previous work [66], consisting of five 5-point Likert items about usability factors to gather user feedback. Please refer to the supplementary material for the questions in full. After the final block, participants were asked to rank the techniques and were asked open-ended questions to extract technique preferences and opinions. For each technique, each participant completed 4 sub-blocks (i.e., one per error condition) of 32 trials (i.e., 8 directions x 2 amplitudes x 2 repetitions). As such, each participant performed 6 techniques, 4 error conditions, 8 directions, 2 amplitudes, and 2 repetitions, resulting in a total of 768 trials. The study took 90 minutes to complete. In total, we collected 18,432 trials.

4.3 Apparatus

The techniques and study environment were developed in Unity version 2020.3.11f1. An HTC Vive Pro Eye head-mounted display with a resolution of 2880 x 1600 was used to record eye and head input at 90 Hz, while a single HTC Vive handheld controller was used for controller input at 90 Hz. For all of the Weighted Pointer techniques, a selection was made using a trigger press. Pilot testing was used to decide upon a common set of parameters for all Weighted Pointers (i.e., $n,m = 3$, $N_l, N_w, N_p = 20 frames$, $v_{max} = 1000°/s$, $t_{transition} = 0.2s$, $E_{acc} = 2°$, $E_{max} = 8°$. E_{acc} and E_{max} were set so that the experiment would cover different levels of gaze and fallback modality weightings). A fixed common neutral controller direction that pointed downward and forward was used for all participants when they used the Relative and Trimodal techniques so that the controller could hang comfortably next to their body (Figure 3c). The Pinpointing techniques were implemented as described by Kytö et al., wherein the cursor was only visible during the refinement mode, the switch to the refinement mode was performed via a trigger press, and a selection was made on the trigger release [37].

5 Results

The dependent variables of interest were selection time, error rate, head and controller movements, and perceived workload and usability. Unless otherwise stated, the analysis was performed with a four-way repeated measures ANOVA (α=.05) with Technique, Error, Direction and Amplitude as independent variables. When the assumption of Specificity was violated, as tested with Mauchly's test, Greenhouse-Geisser corrected values were used in the analysis. QQ-plots were used to validate the assumption of normality. Bonferroni-corrected post-hoc tests were used when applicable. Effect sizes are reported as partial eta squared (η_p^2). RAW NASA-TLX scores and the Likert-scale usability data were analyzed using Friedman tests and Bonferroni-corrected Wilcoxon signed-rank tests were used for the post-hoc analysis.

Both temporal and spatial filtering was used to remove outlier trials. For temporal filtering, trials with selection times that were more than 3 standard deviations from the grand mean were discarded. For spatial filtering, trials with a selection endpoint projected onto the task axis (i.e., the vector between the central and selection targets) that were more than 3 standard deviations from the grand mean were discarded. In total, 618 trials were discarded (3.3%). After outlier filtering, 100 of the 9216 cells had missing values for the 4-way repeated measures ANOVA. These values were replaced with the maximum value over all participants for each respective dependent variable.

5.1 Selection Time

To understand how participants accounted for eye tracking errors and the effect different types of eye tracking errors had on user performance, selection time was computed (Figure 5). Selection time was defined as the time between the start of a trial to a successful selection, irrespective of prior incorrect selections.

The results did not find a 4-way interaction, but a significant 3-way interaction was found for Technique × Error × Amplitude ($F_{8.56,196.94}$=2.55, p=.010, η_p^2=.100). Further inspection found that there were significant simple interactions for Technique × Error at the 30° ($F_{5.70,131.09}$=2.88, p=.013, η_p^2=.111) and 60° ($F_{6.47,148.76}$=6.36, p<.001, η_p^2=.217) target amplitudes. Pairwise comparisons showed that all Weighted Pointers were faster than the Pinpointers in the No Error condition, as participants could quickly point and select targets using their gaze rather than needing to use the button release to select a target (all p<.028). Also, the Hands-free technique was found to be significantly faster than the Relative, Eye+Head, and Eye+Device techniques for all other Error conditions at both amplitudes (all $p \leq .01$), while the Controller-based technique was faster than the Eye+Device and Relative techniques (all $p \leq .019$).

However, the Relative technique struggled with added precision and was significantly slower than all other Weighted Pointer techniques in the Precision and Both conditions (all $p \leq .009$). Post-hoc tests also found that for the Hands-free techniques, the Precision and Accuracy conditions were faster than the Both condition at the 30° target amplitude (both $p \leq .028$). This is presumably because participants could partially rely on gaze movements in conjunction with the head, leading to an increase in selection speed, whereas participants had to rely solely on the slower head movements in the Both condition.

For the Trimodal technique, however, the Precision condition was faster than the Accuracy and Both conditions at the 30° target amplitude ($p \leq .025$). This may be because participants had to perform fewer corrective movements and could move their head directly towards the target, similar to their natural head movements, indicating that accuracy errors may have a larger negative effect on selection time. To summarize, the results showed that the Weighted Pointers were significantly

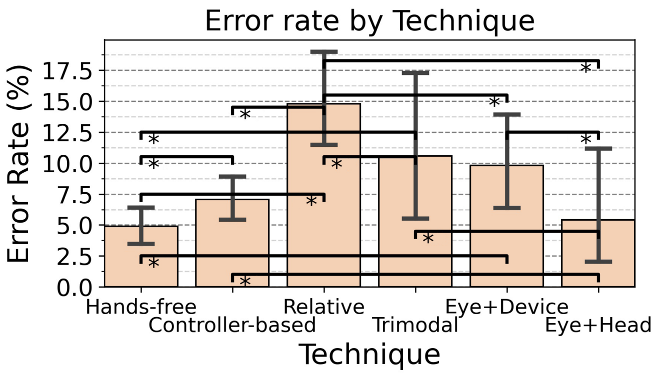

Fig. 6. The error rate for each technique. The error bars represent 95% confidence intervals of the means. Statistical significance is marked with stars ($*$=p<.01).

faster than the Pinpointers. Furthermore, the results demonstrated that users were fastest with the head as a fallback modality for gaze. Finally, the results partially justified the use of a continuous transition to the fallback modality as participants were faster with the Hands-free technique when using gaze in conjunction with a fallback modality compared to when they only used their head.

The results also found significant Technique × Amplitude simple interactions for the Accuracy ($F_{3.56,82.78}$=5.69, p=.001, η_p^2=.198), Precision ($F_{3.81,87.64}$=4.98, p=.001, η_p^2=.178) and Both ($F_{3.88,89.14}$=8.62, p<.001, η_p^2=.278) conditions, but not the No condition ($F_{5,115}$=0.29, p=.919, η_p^2=.012). The results suggest that the Relative and Trimodal techniques were affected by longer target amplitudes presumably because the fixed neutral direction led participants to compensate for the natural rotation that occurred as their bodies rotated towards targets, which affected the cursor position. This effect was less pronounced for the Trimodal technique, as participants could compensate using head movements. Furthermore, this effect was not found in the No Error condition, as participants relied only on gaze.

Finally, we found significant main effects for Technique ($F_{2.97,68.39}$=13.23, p<.001, η_p^2=.365), Error Condition ($F_{1.96,45.02}$=58.11, p<.001, η_p^2=.716), Amplitude ($F_{1,23}$=500.33, p<.001, η_p^2=.956), and Direction ($F_{4.61,105.94}$=3.08, p=.015, η_p^2=.118). The Hands-free (1.77s) and Controller-based (1.87s) techniques proved to be overall significantly faster (all $p \leq$.021) than Relative (2.27s), Eye+Device (2.30s), and Eye+Head (2.28s). Furthermore, Trimodal (1.97) proved significantly faster than Relative and Eye+Device (all $p \leq$.033). Furthermore, participants were significantly faster in the No condition than in other conditions as users could rely on accurate gaze (all p<.001). Participants were also faster in Accuracy and Precision conditions compared to the Both condition (all$p \leq$.007). Finally, post-hoc tests showed that larger amplitudes led to longer selection times (p<.001), but no significance could be found between directions.

5.2 Error Rate

The error rate, i.e., the number of trials resulting in an error divided by the total number of trials, was computed to understand the accuracy of the techniques. An error was counted whenever a participant missed the target prior to a correct selection or if a participant failed to select the target before the 10-second trial timeout. As the error rate was positively skewed and violated the repeated measures ANOVA's assumption of normality, the number of errors was used as count data and an "underdispersed" Poisson regression model was fit to the data. We included all interactions involving Technique and all main effects in the regression. The analysis found that the overall model was significant ($\chi^2(95, N$=9216)=608.65, p<.001). An investigation of the model effects revealed a significant main effect for Technique ($\chi^2(5)$=13.81, p=.019). Sequential Šidák pairwise comparisons (Figure 6) showed that participants were more accurate with techniques that used the head as the fallback modality. The Hands-free and Eye+Head techniques had significantly lower error rates than all of the other techniques. Meanwhile, the Relative technique had a significantly higher error rate than all of the other techniques. No significant differences were found between the Controller-based, Trimodal, and Eye+Device techniques.

5.3 Head and Controller Movement

The Relative technique was designed to minimize controller movement through relative moments, while the Trimodal technique was designed to minimize head and controller movement by distributing the interaction across both modalities. To verify these assumptions, we investigated the amount of head and controller movement participants exhibited for each technique. For each movement, we were mainly interested in the techniques that used the corresponding modality as the fallback modality, i.e. Hands-free, Trimodal and Eye+Head for the head, and Controller-based, Relative, Trimodal and Eye+Device for the controller. To quantify this, we computed the amount of head or controller rotation performed by the user until a successful selection was made.

5.3.1 Head Movement

The analysis found no 4-way or 3-way interactions, however, significant Technique × Error ($F_{6.33,145.73}$=6.77, p<.001, η_p^2=.227) and Technique × Direction ($F_{10.90,250.67}$=3.84, p<.001, η_p^2=.143) interactions were found for the amount of head movement made by participants. Further investigations of the Technique × Error interaction did not find any significant differences between any of the techniques during the No error condition, likely because no extra head movements were needed for selection (Figure 7a). For the other error conditions, the techniques that used the head as a fallback modality had more head movements than the techniques relying on the controller ($p \leq$.041). However, no significant differences were found between the Hands-free, Trimodal and Eye+Head techniques (i.e., all $p \leq$.140), implying that users mainly relied on the head as their fallback modality when using the Trimodal technique. However, there was a larger variance between users with the Trimodal technique (Figure 7), indicating that users deployed different amounts of head movements. When further investigating the Technique × Direction interaction, we found that participants tended to perform less head movement in the vertical directions, as has been shown in previous eye-head coordination work in VR [63].

We found significant main effects for Technique ($F_{3.16,72.76}$=20.83, p<.001, η_p^2=.475), Error Condition ($F_{3,69}$=26.64, p<.001, η_p^2=.537), Amplitude ($F_{1,23}$=5618.07, p<.001, η_p^2=.996), and Direction ($F_{3.47,79.85}$=20.90, p<.001, η_p^2=.476). Participants performed significantly more head movement with the Hands-free (48.42°), Trimodal (44.55°), and Eye+Head (49.27°) techniques that used head movement for cursor control compared to the Controller-based (38.12°), Eye+Device (40.94°) and Relative (38.27°) techniques (all $p \leq$.010). Post-hoc tests showed that users performed significantly less head movement in the No condition where users relied mainly on gaze compared to all other error conditions (all p<.001). Finally, post-hoc tests again showed that larger amplitudes led to more head movement (p<.001), and that participants performed less head movement in vertical directions compared to other directions (all $p \leq$.020).

5.3.2 Controller Movement

The analysis did not find any 4-way or 3-way interactions, however, significant 2-way interactions were found for Technique × Error ($F_{5.89,135.50}$=8.33, p<.001, η_p^2=.266), Technique × Amplitude ($F_{3.56,81.87}$=28.48, p<.001, η_p^2=.553), and Technique × Direction ($F_{7.83,180.13}$=2.66, p=.009, η_p^2=.104). For Technique × Error, post-hoc tests demonstrated that the Controller-based and Eye+Device techniques had significantly more controller movement than all other techniques during the Accuracy, Precision, and Both conditions (all $p \leq$.049; Figure 7b), showing that the Trimodal and Relative techniques required less controller movement for selection. No significant differences

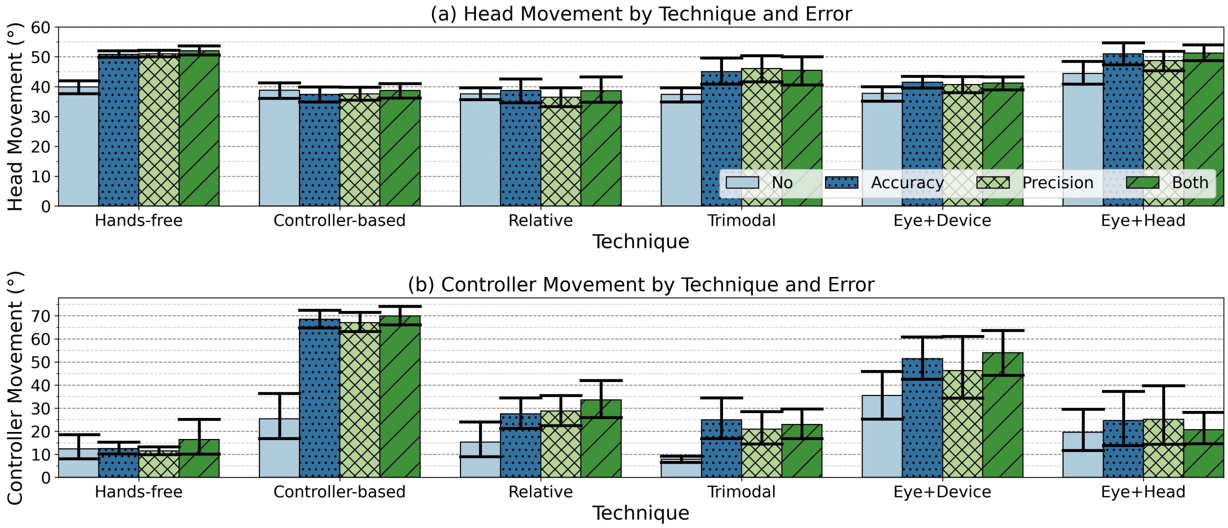

Fig. 7. Mean (a) head and (b) controller rotation for each technique by error condition. The error bars represent the 95% confidence intervals of the means.

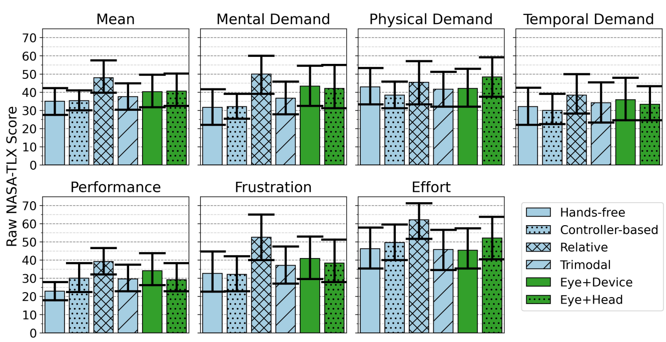

Fig. 8. The mean responses for the Raw NASA TLX questionnaire. The error bars represent the 95% confidence intervals of the means.

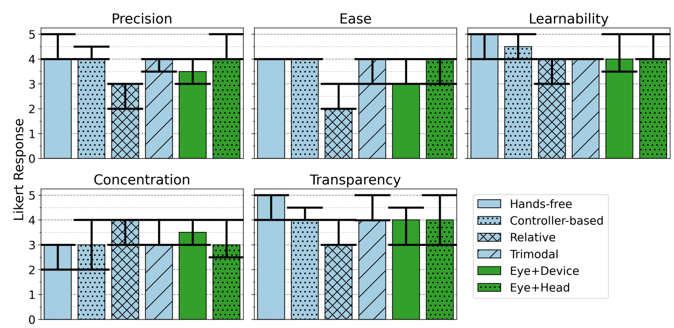

Fig. 9. The median responses for the usability questionnaire. The error bars represent the 95% confidence intervals of the medians.

were found between techniques in the No condition, likely because the users relied on gaze. Also, all techniques that used the controller for pointing had significantly lower controller movement in the No condition compared to the other Error conditions, possibly because participants relied on gaze (all $p \leq .035$). For Technique × Amplitude, the Controller-based and Controller+Device techniques had significantly more controller movement than the other techniques at both amplitudes (all $p \leq .026$), again implying that less controller movement was used for Trimodal and Relative.

Finally, we again found significant main effects for Technique ($F_{2.70,62.03}$=25.83, $p<.001$, η_p^2=.529), Error Condition ($F_{1.97,45.25}$=40.19, $p<.001$, η_p^2=.636), Amplitude ($F_{1,23}$=239.18, $p<.001$, η_p^2=.912), and Direction ($F_{3.93,90.47}$=6.80, $p<.001$, η_p^2=.228). Controller-based (57.87°) and Eye+Device (47.39°) showed significantly more controller movement than all other techniques (all $p \leq .007$). Interestingly, no significance was found between the Relative (26.78°) and Trimodal (21.38°) techniques that utilized relative controller rotations, and the Head-based (13.20°) and Eye+Head (23.21°) techniques which did not use controller movements for interaction. The No condition again showed less movement than other error conditions ($p \leq .001$), and participants performed more controller movement at larger amplitudes ($p \leq .001$). Finally, post-hoc tests showed no significance in controller movement between directions.

5.4 Workload and Usability

Friedman tests using the Raw NASA TLX workload metrics (Figure 8) revealed significant differences for Mental Demand ($\chi^2(5)$=20.48, $p<.001$), Performance ($\chi^2(5)$=21.65, p=.001), and Frustration ($\chi^2(5)$=13.15, p=.022). Post-hoc tests revealed that the Hands-free technique had significantly lower Mental Demand (p=.002) and Frustration (p=.002), and was perceived to have significantly higher Performance (p=.041) than the Relative technique.

Friedman tests on the usability ratings (Figure 9) found significant differences for perceived Precision ($\chi^2(5)$=33.89, $p<.001$), Ease ($\chi^2(5)$=36.05, $p<.001$), Learnability ($\chi^2(5)$=15.26, p=.009), and Concentration ($\chi^2(5)$=20.17, p=.001). Post-hoc tests showed that participants felt significantly less precise when using the Relative technique compared to the Controller-based, Hands-free, and Eye+Head techniques (all $p \leq .003$). For Ease, post-hoc tests showed that it was harder to perform selections when using the Relative technique than the Controller-based, Hands-free, Trimodal, and Eye+Head techniques (all $p \leq .002$). For Learnability, post-hoc tests showed that it was more difficult to learn how to use the Relative technique than the Hands-free technique (p=.039). Finally, the Relative technique required significantly higher Concentration than the Hands-free technique (p=.003).

5.5 User Preferences

The user preference results showed a trend toward the Weighted Pointers techniques (i.e, 10 participants preferred Hands-free, 5 participants preferred Controller-based, 4 participants preferred Trimodal, 2 participants preferred Eye+Head, 2 participants preferred Eye+Device, and one participant stated no preference). The Hands-free technique was the most preferred technique due to its *"ease of use"* (P10) and *"simplicity"* (P6), and participants stated that they *"had more control with the head"* (P14). However, participants also thought that the technique was *"straining"* (P6, P11) on the neck. The Controller-based technique was the second most preferred technique because participants liked the technique's *"speed"* (P5) and *"control"* (P23), but also thought it was more *"demanding"* (P4) to point with the controller.

For the Trimodal technique, participants liked that it required *"less work"* (P6), *"you could move the cursor with either your head or controller"* (P22), and they could *"use the controller to make small edits"* (P19). However, participants criticized its *"sensitivity"* (P4) because moving both the head and controller could cause large cursor movements. Participants also questioned *"why the controller needs to be in the neutral position"* (P21). Participants thought the Eye+Device technique placed a small amount of *"strain on the neck and arm"* (P11) and it was *"easy to select targets"* (P6), however, participants also thought that it *"requires a bit more concentration"* (P13) and that they *"have to intentionally look at the target before pressing the trigger"* (P6). Participants thought that the Eye+Head technique was *"easy to use"* (P10) and *"intuitive"* (P6). Participants again stated that they felt more accurate using their head, i.e., *"as someone with unsteady hands, I was more precise with this technique being isolated to just head movements"* (P11). Finally for the Relative technique, while participants liked that they *"did not have to move the hands as much"* (P6) and that they could *"adjust when gaze could not get right on the target"* (P14), participants struggled during high precision errors, stating that the pointer was *"shaky"* (P6), and *"jittery"* (P21).

Participants also commented on whether or not they felt that the Weighted Pointers' transparency-based cursor clearly communicated the current mode of interaction. Participants *"liked"* (P15) and thought it was *"intuitive"* (P10) to use target highlighting for gaze interaction, however, participants also struggled with the visualization, e.g., *"it was hard to figure out at first when I needed to use my gaze or the controller to point at the objects"* (P10). Furthermore, some participants did not understand why the modality or visualization changed, e.g., *"it felt like when the cursor showed, the controller was the appropriate option for aiming at the target, and when the cursor did not show, the headset was better at tracking my eye movements"* (P19). Furthermore, no participant mentioned the change in transparency, indicating that participants did not notice that level of granularity in the visualization.

6 Discussion

6.1 Study Results

The study results validated the Weighted Pointer's principal approach of adapting pointing modalities to eye tracking error levels. All Weighted Pointer techniques were found to be faster than the manual Pinpointers techniques in the No error condition, highlighting the performance gains that can be obtained when only using a fallback modality as needed. In addition, most Weighted Pointer techniques showed better or equal performance to the Eye+Device and Eye+Head techniques in all other Error conditions. The Hands-free technique was found to be significantly faster than Eye+Device and Eye+Head in all conditions, and had a significantly less error rate than the Eye+Device technique. Furthermore, the Controller-based technique was significantly faster than the Eye+Device technique. Finally, the majority of participants preferred a Weighted Pointer technique as their favourite.

One of the main benefits of the Weighted Pointer techniques comes from their use of gaze when gaze data is accurate and precise enough for interaction, thereby allowing users to make faster selections without any extra steps. However, just as with other work [66], the results suggested that participants were less accurate with gaze compared to the fallback modality, as they made errors as a result of premature selections whilst under- or overshooting targets. Combining the Weighted Pointers with target acquisition techniques such as the Bubble cursor [22] could improve this. Furthermore, the techniques do not require explicit input to switch to a fallback modality, resulting in seamless transitions and users needing to perform fewer interaction steps. In addition, although the Weighted Pointers required that participants used a button click to confirm their selections in our study, the automatic switching between modalities would allow them to be combined with other confirmation techniques such as dwelling to support hands-free interaction.

The study results also implied that the head is more suitable as a fallback modality for gaze than a controller. The Hands-free and Eye+Head techniques were among the quickest and most accurate techniques. A possible explanation for this could be that the motions performed during head-pointing were similar to the eye-head coordinated motions performed during gaze-pointing, and that the head does not suffer from factors such as hand tremors. Furthermore, an interesting result from the study was that participants were quicker with the Hands-free technique in the Accuracy and Precision conditions, where gaze was used in conjunction with the head, than in the Both condition, where users only relied on the head. These results point towards the benefits of having a continuous transition to a fallback modality, as gaze can still be used when it is too noisy for stand-alone use, it is helpful when moving the cursor quickly to the target, and it lowers the required pointing accuracy for the fallback modality.

For the controller techniques, with the exception of the No error condition, there were little performance differences between the other Error conditions. These results suggest that the techniques may handle accuracy and precision errors, as well as their combination, while still providing opportunities for stable interaction. However, the results also point towards improvements that can be made to reduce the amount of head or controller movement needed to refine inaccurate gaze signals. While using the weighted mean to select the current pointing position proved effective in dealing with gaze error, future research should investigate other weighting functions that could be used to combine gaze and fallback modalities, and which may be more suitable than the controller as a fallback modality.

Both the Relative and Trimodal techniques demonstrated how relative movements can be used to minimize the controller movements needed for selection. However, the selection time and error rate results also highlighted weaknesses with the implementation of these techniques. For example, the controller would naturally follow the body movements needed to reach distant targets, causing an offset from the neutral direction which affected the cursor position. Other devices such as the Myo armband, which can detect wrist flexion and extension, could be used as an alternative input device to avoid this issue. Furthermore, as highlighted by performance and usability metrics, the Relative technique proved difficult to use when there was additional precision error as the fallback modality was partially gaze-based, thereby highlighting that all types of eye tracking error have to be considered when choosing a fallback modality. More research is needed to understand the influence of each type of eye tracking error on selection performance, as has been done in collaborative settings [17, 48]. Furthermore, these results indicate that additional interaction parameters can be adapted to eye tracking error. For example, expanding N_p when precision error is detected could smoothen the additional noise.

6.2 Limitations

The study results provided little evidence that participants had issues adapting to new error levels. However, we only considered one level of accuracy and precision error, and did not include data loss for practical reasons. Furthermore, we did not formally investigate whether a sudden transition in error affects performance, so this should be a future direction of research. A key aspect of adapting to new error levels is visualizing the current mode of interaction. In our work, we used cursor transparency and target highlighting to visualize pointing and selection feedback. However, purely visual feedback mechanisms may be problematic in certain situations, such as when switching to a controller that is pointing outside of a user's field of vision. An interesting area of research could be to investigate different types of feedback, such as haptic or auditory, to indicate the current mode of interaction to the user.

Our study used synthetic noise to simulate eye tracking accuracy and precision. Using the techniques in the wild may result in different technique behaviours and performances. In addition, we did not formally test data loss and related noise artifacts in our study. Although data loss was present in the study in the form of user blinks, long and frequent periods of data loss may impact the results. The use of synthetic noise is mainly a practical aspect as it may be difficult who struggle with the sensor while also controlling study conditions. Future research could introduce artificial eye tracking noise that is more in line with in the wild eye tracking noise and encourage a standardized way to evaluate the impact of noise as various synthetic noise functions have been employed in previous work [28,46,48]. Such developments would be beneficial when comparing different approaches to handling eye tracking noise and evaluating existing gaze-based techniques' ability to cope with errors.

Furthermore, the study only investigated target selection in an abstract environment without any distractor targets. Further research needs to be done to evaluate the techniques in more natural settings. Occlusion is interesting to investigate as targets are closely placed together, and occlusion lessens the effective target width making interaction more difficult, especially with noisy data [66, 77]. The impact of data noise on interaction has to be further investigated within applications and natural environments beyond controlled lab environments.

A current weakness of our implementation of the Weighted Pointer, similar to other error-aware techniques, is that they are dependent on a user calibration step to generate the data needed to model eye tracking errors. Ideally, the Error Detection Model should seamlessly detect and update itself without repeated user calibrations. Several projects have investigated this approach for 2D and 3D interfaces [18, 65] but these proposed solutions have yet to replace explicit calibration methods. Compared to gaze-only, error-aware interfaces, using fallback modalities has the advantage of not needing to know the direction of the current eye tracking error and instead only needing to know of its existence. This loosens the requirements for the Error Detection Model and could make it easier to implement and deploy seamless re-calibration techniques to continuously update the current error.

6.3 Future Work

The individual parts of the Weighted Pointer (i.e., Error Detection Model, Weighting Model, Pointing Techniques) are replaceable, making it possible to swap out components to change pointing behaviours. This opens up an exciting design space, where different components could be swapped and compared to provide the best user experience given a certain context. Although the present work assumed that the fallback modality was stable, this may not always be the case. In future work, it would be useful to investigate how Weighted Pointer techniques could be adjusted so that the Error Detection Model and weightings could be dependent on the errors from both modalities.

Furthermore, adapting interaction to the current signal noise is not limited to gaze-based interaction and eye tracking errors but may also be useful for other sensors that have tracking issues or for modalities that may have participant-specific factors such as hand tremors. Gaze can easily be replaced as the main modality within the proposed techniques and different modalities could be used in combination. Furthermore, this concept could be further expanded to include switching modalities based on other factors such as social acceptance [34] and one's available range of motion [76]. For example, a hand-based interaction system could switch to gaze as a fallback modality in crowded environments where one's range of motion may be limited and in-air gestures would be socially uncomfortable. This opens up many research avenues to explore to investigate ways of switching modalities to ensure comfortable and efficient interaction irrespective of user context.

Lastly, although the results from the study were obtained in VR, we do not expect this to limit the applicability of the Weighted Pointers or the findings. The Weighted Pointer may be especially useful in AR settings that have to adapt to a multitude of environmental and contextual changes which may severely impact eye tracking data quality. Providing a smooth and stable user experience within the wider population is a key requirement for gaze-based interaction in AR and a significant future area of research. Being able to adaptively swap to different modalities could be a way to achieve this and opens up exciting research opportunities within AR interaction.

7 Conclusion

This research presented Weighted Pointer techniques that utilized fallback modalities alongside error-aware, gaze-based pointing interaction techniques to provide stable interaction in the event of eye tracking errors brought about by accuracy, precision and data loss issues. An evaluation of the techniques found that they were more performant and preferred than the use of manual techniques to switch between modalities, however, the choice of which backup modality was used had a significant impact on user performance. The notion of using backup modalities when sensor noise is present extends beyond traditional gaze-based pointing techniques, and thereby presents new opportunities for adaptive interfaces that do not hamper user experiences.

Acknowledgments

We'd like to thank Casey Brown, Michael Frederick, Philipp Popp for their help with conducting the study.

References

[1] S. Ahn, J. Son, S. Lee, and G. Lee. Verge-it: Gaze interaction for a binocular head-worn display using modulated disparity vergence eye movement. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, p. 1–7. ACM, 2020. doi: 10.1145/3334480.3382908

[2] F. Argelaguet and C. Andujar. Efficient 3d pointing selection in cluttered virtual environments. *IEEE Computer Graphics and Applications*, 29(6):34–43, 2009. doi: 10.1109/MCG.2009.117

[3] I. B. Adhanom, S. C. Lee, E. Folmer, and P. MacNeilage. Gazemetrics: An open-source tool for measuring the data quality of hmd-based eye trackers. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Short Papers. ACM, 2020. doi: 10.1145/3379156.3391374

[4] P. Bækgaard, J. P. Hansen, K. Minakata, and I. S. MacKenzie. A fitts' law study of pupil dilations in a head-mounted display. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19. ACM, 2019. doi: 10.1145/3314111.3319831

[5] M. Baloup, T. Pietrzak, and G. Casiez. Raycursor: A 3d pointing facilitation technique based on raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12. ACM, 2019. doi: 10.1145/3290605.3300331

[6] M. Barz, F. Daiber, and A. Bulling. Prediction of gaze estimation error for error-aware gaze-based interfaces. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, p. 275–278. ACM, 2016. doi: 10.1145/2857491.2857493

[7] M. Barz, F. Daiber, D. Sonntag, and A. Bulling. Error-aware gaze-based interfaces for robust mobile gaze interaction. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18. ACM, 2018. doi: 10.1145/3204493.3204536

[8] R. Bates and H. O. Istance. Why are eye mice unpopular? a detailed comparison of head and eye controlled assistive technology pointing devices. *Universal Access in the Information Society*, 2(3):280–290, Oct 2003. doi: 10.1007/s10209-003-0053-y

[9] M. Bernhard, E. Stavrakis, M. Hecher, and M. Wimmer. Gaze-to-object mapping during visual search in 3d virtual environments. *ACM Trans. Appl. Percept.*, 11(3), Aug. 2014. doi: 10.1145/2644812

[10] G. Casiez, N. Roussel, and D. Vogel. 1 € 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, p. 2527–2530. ACM, 2012. doi: 10.1145/2207676.2208639

[11] I. Chatterjee, R. Xiao, and C. Harrison. Gaze+gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, p. 131–138. ACM, 2015. doi: 10.1145/2818346.2820752

[12] M. Choi, D. Sakamoto, and T. Ono. Bubble gaze cursor + bubble gaze lens: Applying area cursor technique to eye-gaze interface. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Full Papers. ACM, 2020. doi: 10.1145/3379155.3391322

[13] B. David-John, C. Peacock, T. Zhang, T. S. Murdison, H. Benko, and T. R. Jonker. Towards gaze-based prediction of the intent to interact in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '21 Short Papers. ACM, 2021. doi: 10.1145/3448018.3458008

[14] S. Deng, J. Chang, S.-M. Hu, and J. J. Zhang. Gaze modulated disambiguation technique for gesture control in 3d virtual objects selection. In

2017 3rd IEEE International Conference on Cybernetics (CYBCONF), pp. 1–8. IEEE, 2017. doi: 10.1109/CYBConf.2017.7985779

[15] H. Drewes, M. Khamis, and F. Alt. Smooth pursuit target speeds and trajectories. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, p. 139–146. ACM, 2018. doi: 10.1145/3282894.3282913

[16] J. Drewes, G. S. Masson, and A. Montagnini. 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, p. 209–212. ACM, 2012. doi: 10.1145/2168556.2168596

[17] A. Erickson, N. Norouzi, K. Kim, J. J. LaViola, G. Bruder, and G. F. Welch. Effects of depth information on visual target identification task performance in shared gaze environments. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1934–1944, 2020. doi: 10.1109/TVCG.2020.2973054

[18] R. Fares, S. Fang, and O. Komogortsev. Can we beat the mouse with magic? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 1387–1390. ACM, 2013. doi: 10.1145/2470654.2466183

[19] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris. 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, p. 1118–1130. ACM, 2017. doi: 10.1145/3025453.3025599

[20] D. R. Flatla, C. Gutwin, L. E. Nacke, S. Bateman, and R. L. Mandryk. Calibration games: Making calibration tasks enjoyable by adding motivating game elements. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, p. 403–412. ACM, 2011. doi: 10.1145/2047196.2047248

[21] A. R. Gomez and H. Gellersen. Smooth-i: Smart re-calibration using smooth pursuit eye movements. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, ETRA '18. ACM, 2018. doi: 10.1145/3204493.3204585

[22] T. Grossman and R. Balakrishnan. The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, p. 281–290. ACM, 2005. doi: 10.1145/1054972.1055012

[23] T. Grossman and R. Balakrishnan. The design and evaluation of selection techniques for 3d volumetric displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, p. 3–12. ACM, 2006. doi: 10.1145/1166253.1166257

[24] G. d. Haan, M. Koutek, and F. H. Post. IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In E. Kjems and R. Blach, eds., *Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2005. doi: 10.2312/EGVE/IPT_EGVE2005/201-209

[25] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 1063–1072. ACM, 2014. doi: 10.1145/2556288.2557130

[26] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. A survey of design issues in spatial input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, UIST '94, p. 213–222. ACM, 1994. doi: 10.1145/192426.192501

[27] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, J. Halszka, and J. van de Weijer. *Eye Tracking : A Comprehensive Guide to Methods and Measures*. Oxford University Press, 2011.

[28] K. Holmqvist, M. Nyström, and F. Mulvey. Eye tracker data quality: What it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, p. 45–52. ACM, 2012. doi: 10.1145/2168556.2168563

[29] T. E. Hutchinson, K. P. White, W. N. Martin, K. C. Reichert, and L. A. Frey. Human-computer interaction using eye-gaze input. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1527–1534, Nov. 1989. doi: 10.1109/21.44068

[30] R. J. K. Jacob. 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*, CHI '90, pp. 11–18. ACM, 1990. doi: 10.1145/97243.97246

[31] S. Jalaliniya, D. Mardanbegi, and T. Pederson. Magic pointing for eyewear computers. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ISWC '15, p. 155–158. ACM, 2015. doi: 10.1145/2802083.2802094

[32] E. Kaiser, A. Olwal, D. McGee, H. Benko, A. Corradini, X. Li, P. Cohen, and S. Feiner. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th International Conference on Multimodal Interfaces*, ICMI '03, p. 12–19. ACM, 2003. doi: 10.1145/958432.958438

[33] F. P. Kalaganis, E. Chatzilari, S. Nikolopoulos, I. Kompatsiaris, and N. A. Laskaris. An error-aware gaze-based keyboard by means of a hybrid bci system. *Scientific Reports*, 8(1):13176, Sep 2018. doi: 10.1038/s41598-018-31425-2

[34] M. Koelle, S. Ananthanarayan, and S. Boll. *Social Acceptability in HCI: A Survey of Methods, Measures, and Design Strategies*, p. 1–19. ACM, 2020.

[35] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and accurate 3d selection by progressive refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 67–74, 2011. doi: 10.1109/3DUI.2011.5759219

[36] A. Kurauchi, W. Feng, C. Morimoto, and M. Betke. Hmagic: Head movement and gaze input cascaded pointing. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '15. ACM, 2015. doi: 10.1145/2769493.2769550

[37] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise head- and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–14. ACM, 2018. doi: 10.1145/3173574.3173655

[38] M. F. Land, N. Mennie, and J. Rusted. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311–1328, 1999. PMID: 10755142. doi: 10.1068/p2935

[39] C. Lankford. Effective eye-gaze input into windows. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, p. 23–27. ACM, 2000. doi: 10.1145/355017.355021

[40] M. Lee, M. Billinghurst, W. Baek, R. Green, and W. Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Reality*, 17(4):293–305, Nov 2013. doi: 10.1007/s10055-013-0230-0

[41] Z. Li, M. Zhao, Y. Wang, S. Rashidian, F. Baig, R. Liu, W. Liu, M. Beaudouin-Lafon, B. Ellison, F. Wang, I. Ramakrishnan, and X. Bi. Bayesgaze: A bayesian approach to eye-gaze based target selection. In *Proceedings of Graphics Interface 2021*, GI 2021, pp. 231 – 240. Canadian Information Processing Society, 2021. doi: 10.20380/GI2021.35

[42] C. Lutteroth, M. Penkar, and G. Weber. Gaze vs. mouse: A fast and accurate gaze-only click alternative. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, p. 385–394. ACM, 2015. doi: 10.1145/2807442.2807461

[43] J. Mankoff. Providing integrated toolkit-level support for ambiguity in recognition-based interfaces. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '00, p. 77–78. ACM, 2000. doi: 10.1145/633292.633339

[44] J. Mankoff, S. E. Hudson, and G. D. Abowd. Interaction techniques for ambiguity resolution in recognition-based interfaces. In *ACM SIGGRAPH 2007 Courses*, SIGGRAPH '07, p. 11–es. ACM, 2007. doi: 10.1145/1281500.1281522

[45] D. Mardanbegi, T. Langlotz, and H. Gellersen. Resolving target ambiguity in 3d gaze interaction through vor depth estimation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12. ACM, 2019. doi: 10.1145/3290605.3300842

[46] M. H. Mughrabi, A. K. Mutasim, W. Stuerzlinger, and A. U. Batmaz. My eyes hurt: Effects of jitter in 3d gaze tracking. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 310–315, 2022. doi: 10.1109/VRW55335.2022.00070

[47] D. C. Niehorster, T. Santini, R. S. Hessels, I. T. C. Hooge, E. Kasneci, and M. Nyström. The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, 52(3):1140–1160, Jun 2020. doi: 10.3758/s13428-019-01307-0

[48] N. Norouzi, A. Erickson, K. Kim, R. Schubert, J. LaViola, G. Bruder, and G. Welch. Effects of shared gaze parameters on visual target identification task performance in augmented reality. In *Symposium on Spatial User Interaction*, SUI '19. ACM, 2019. doi: 10.1145/3357251.3357587

[49] A. Olwal, H. Benko, and S. Feiner. Senseshapes: using statistical geometry for object selection in a multimodal augmented reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pp. 300–301, 2003. doi: 10.1109/ISMAR.2003.1240730

[50] J. L. Orquin and K. Holmqvist. Threats to the validity of eye-movement

research in psychology. *Behavior Research Methods*, 50(4):1645–1656, Aug 2018. doi: 10.3758/s13428-017-0998-z

[51] Y. S. Pai, T. Dingler, and K. Kunze. Assessing hands-free interactions for vr using eye gaze and electromyography. *Virtual Reality*, 23(2):119–131, Jun 2019. doi: 10.1007/s10055-018-0371-2

[52] C. Palmer and B. Sharif. Towards automating fixation correction for source code. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, p. 65–68. ACM, 2016. doi: 10.1145/2857491.2857544

[53] N. Pathmanathan, M. Becher, N. Rodrigues, G. Reina, T. Ertl, D. Weiskopf, and M. Sedlmair. Eye vs. head: Comparing gaze methods for interaction in augmented reality. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Short Papers. ACM, 2020. doi: 10.1145/3379156.3391829

[54] K. Pfeuffer and H. Gellersen. Gaze and touch interaction on tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, p. 301–311. ACM, 2016. doi: 10.1145/2984511.2984514

[55] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, p. 261–270. ACM, 2013. doi: 10.1145/2501988.2501998

[56] M. Porta, A. Ravarelli, and G. Spagnoli. Cecursor, a contextual eye cursor for general pointing in windows environments. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, p. 331–337. ACM, 2010. doi: 10.1145/1743666.1743741

[57] Y. Y. Qian and R. J. Teather. The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 91–98. ACM, 2017. doi: 10.1145/3131277.3132182

[58] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, p. 71–78. ACM, 2000. doi: 10.1145/355017.355028

[59] T. Santini, D. C. Niehorster, and E. Kasneci. Get a grip: Slippage-robust and glint-free gaze estimation for real-time pervasive head-mounted eye tracking. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19. ACM, 2019. doi: 10.1145/3314111.3319835

[60] G. Schmidt, Y. Baillot, D. Brown, E. Tomlin, and J. Swan. Toward disambiguating multiple selections for frustum-based pointing. In *3D User Interfaces (3DUI'06)*, pp. 87–94, 2006. doi: 10.1109/VR.2006.133

[61] J. Schwarz, S. Hudson, J. Mankoff, and A. D. Wilson. A framework for robust and flexible handling of inputs with uncertainty. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, p. 47–56. ACM, 2010. doi: 10.1145/1866029.1866039

[62] L. Sidenmark, C. Clarke, X. Zhang, J. Phu, and H. Gellersen. Outline pursuits: Gaze-assisted selection of occluded objects in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. ACM, 2020. doi: 10.1145/3313831.3376438

[63] L. Sidenmark and H. Gellersen. Eye, head and torso coordination during gaze shifts in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 27(1), Dec. 2019. doi: 10.1145/3361218

[64] L. Sidenmark and H. Gellersen. Eye&head: Synergetic eye and head movement for gaze pointing and selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 1161–1174. ACM, 2019. doi: 10.1145/3332165.3347921

[65] L. Sidenmark and A. Lundström. Gaze behaviour on interacted objects during hand interaction in virtual reality for eye tracking calibration. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19. ACM, 2019. doi: 10.1145/3314111.3319815

[66] L. Sidenmark, D. Mardanbegi, A. R. Gomez, C. Clarke, and H. Gellersen. Bimodalgaze: Seamlessly refined pointing with gaze and filtered gestural head movement. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Full Papers. ACM, 2020. doi: 10.1145/3379155.3391312

[67] L. Sidenmark, D. Potts, B. Bapisch, and H. Gellersen. Radi-eye: Hands-free radial interfaces for 3d interaction using gaze-activated head-crossing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. ACM, 2021. doi: 10.1145/3411764.3445697

[68] F. Steinicke, T. Ropinski, and K. Hinrichs. *Object Selection in Virtual Environments Using an Improved Virtual Pointer Metaphor*, pp. 320–326. Springer, 2006. doi: 10.1007/1-4020-4179-9_46

[69] S. Stellmach and R. Dachselt. Look & touch: Gaze-supported target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2981–2990. ACM, 2012. doi: 10.1145/2207676.2208709

[70] O. Špakov. Comparison of gaze-to-objects mapping algorithms. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*, NGCA '11. ACM, 2011. doi: 10.1145/1983302.1983308

[71] O. Špakov. Comparison of eye movement filters used in hci. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, p. 281–284. ACM, 2012. doi: 10.1145/2168556.2168616

[72] O. Špakov and Y. Gizatdinova. Real-time hidden gaze point correction. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, p. 291–294. ACM, 2014. doi: 10.1145/2578153.2578200

[73] O. Špakov, P. Isokoski, and P. Majaranta. Look and lean: Accurate head-assisted eye pointing. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, p. 35–42. ACM, 2014. doi: 10.1145/2578153.2578157

[74] O. Špakov and P. Majaranta. Enhanced gaze interaction using simple head gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, p. 705–710. ACM, 2012. doi: 10.1145/2370216.2370369

[75] C. Ware and H. H. Mikaelian. An evaluation of an eye tracker as a device for computer input2. *SIGCHI Bulletin*, 17(SI):183–188, May 1986. doi: 10.1145/30851.275627

[76] J. R. Williamson, M. McGill, and K. Outram. *PlaneVR: Social Acceptability of Virtual Reality for Aeroplane Passengers*, p. 1–14. ACM, 2019.

[77] D. Yu, Q. Zhou, J. Newn, T. Dingler, E. Velloso, and J. Goncalves. Fully-occluded target selection in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3402–3413, 2020. doi: 10.1109/TVCG.2020.3023606

[78] R. C. Zeleznik, A. S. Forsberg, and J. P. Schulze. Look-that-there: Exploiting gaze in virtual reality interactions. Technical report, Brown University, 2005.

[79] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, p. 246–253. ACM, 1999. doi: 10.1145/302979.303053

[80] X. Zhang, X. Ren, and H. Zha. Improving eye cursor's stability for eye pointing tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 525–534. ACM, 2008. doi: 10.1145/1357054.1357139