Online-EYE: Multimodal Implicit Eye Tracking Calibration for XR

Baosheng James Hou
Google, USA
Lancaster University, UK

Lucy Abramyan
Google, USA

Prasanthi Gurumurthy
Google, USA

Haley Adams
Google, USA

Ivana Tosic Rodgers
Google, USA

Eric J Gonzalez
Google, USA

Khushman Patel
Google, USA

Andrea Colaço
Google, USA

Ken Pfeuffer
Aarhus University, Denmark

Hans Gellersen
Lancaster University, UK
Aarhus University, Denmark

Karan Ahuja
Google, USA
Northwestern University, USA

Mar Gonzalez-Franco
Google, USA

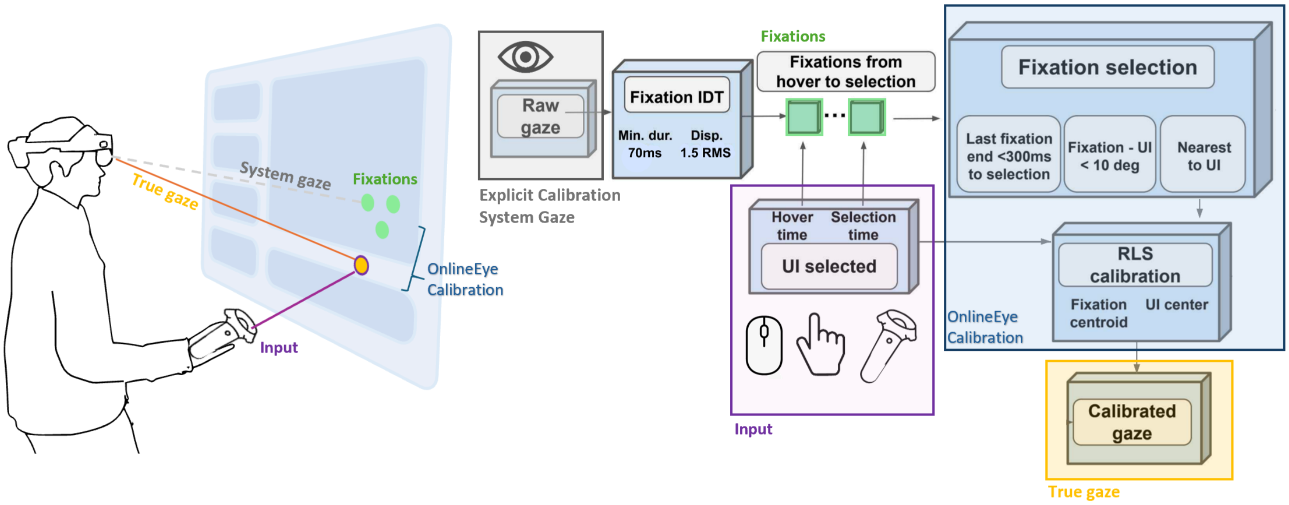


Figure 1: Implicit eye tracking calibration algorithm proposed for Online-EYE using multimodal input UI selection. In this work, we contribute an evaluation of the approach with controllers in VR, but the principle remains the same and can be extended to bare hands, mouse, and other input modalities for interaction, in any XR device with known UI positions.

Abstract

Unlike other inputs for extended reality (XR) that work out of the box, eye tracking typically requires custom calibration per user or session. We present a multimodal inputs approach for implicit calibration of eye tracker in VR, leveraging UI interaction for continuous, background calibration. Our method analyzes gaze data alongside controller interaction with UI elements, and employing ML techniques it continuously refines the calibration matrix without interrupting users from their current tasks. Potentially eliminating the need for explicit calibration. We demonstrate the accuracy and effectiveness of this implicit approach across various tasks and real time applications achieving comparable eye tracking accuracy to native, explicit calibration. While our evaluation focuses on VR and controller-based interactions, we anticipate the broader applicability of this approach to various XR devices and input modalities.

CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713461

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; **Virtual reality**; **Interaction techniques**; **Mixed / augmented reality**; **Empirical studies in HCI**; **Gestural input**.

Keywords

Gaze estimation, implicit calibration, eye tracking

ACM Reference Format:
Baosheng James Hou, Lucy Abramyan, Prasanthi Gurumurthy, Haley Adams, Ivana Tosic Rodgers, Eric J Gonzalez, Khushman Patel, Andrea Colaço, Ken Pfeuffer, Hans Gellersen, Karan Ahuja, and Mar Gonzalez-Franco. 2025. Online-EYE: Multimodal Implicit Eye Tracking Calibration for XR.

In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3706598.3713461

1 INTRODUCTION

As extended reality (XR) devices increasingly adopt eye gaze as a default input modality alongside hands for gaze + pinch [45], but also as an attention mechanism to drive AI agents [5] the need for efficient eye tracking calibration becomes crucial.

Currently the predominant approach to solve eye tracking calibration is to ask every participant to do a native explicit calibration when they start the system. This requires the user to fixate on appearing targets across the head mounted display (HMD). Upon completion the user is ready to start using their HMDs. And despite techniques that propose using alternatives such as pursuit: e.g. moving targets, for calibration [47]; explicit calibration is still a preferred approach.

While it is clear that input experiences require personalizing of sensing to enable high accuracy on our interactions, whether they be hand tracking, or eye tracking, the approach on both has been very different. For hand tracking generally a large database of many hands is collected and labeled, to later be used to train a computer vision model. With the hope of a generic model that will detect any and all hands out-of-the box. However, when it comes to eye tracking somehow we still need to do a native calibration, to estimate the user-specific offset between the optical axis of the eye and the visual axis of the point of gaze.

In this paper we propose an approach that will make eye gaze calibration feel more out-of-the box. We first make the assumption that there can be such thing as a Generic Eye Tracking Model, in a similar way to generic Head Related Transfer Functions (HRTFs) for spatial audio [2], eyes, like ears, are generally located within a range on human faces, and human eyes behave within certain patterns of normality. Hence, HMDs could come pre-calibrated with a generic model [40, 49, 72]. Of course, that would never perform as well as a native calibration personalized by user. But that's where we believe a much simpler multimodal implicit eye tracking calibration can fill the gap.

We present Online-EYE as a technique that leverages other input modalities inside VR to upgrade a generic calibration into performing like a traditional explicit calibration. We can do that because it is natural for the human eye to look at a target before clicking. Hence if we are asking a person that is using their controllers or direct touch to manipulate the user interface (UI) we can use the actual UI elements that are being selected as pseudo-calibration targets. Whether it is a keyboard, a slider, a button or a space invader, we can use that target for implicit eye tracking calibration. And once the eye tracker is calibrated, we can enable gaze and hands [35] or any other gaze interaction [37].

To make sure this form of calibration is robust enough, and can be usable in real scenarios we embark on a series of studies and tests to make sure that our assumptions are correct:

- Hypothesis 1: users are looking at the targets when they click
- Hypothesis 2: we can use those clicks on the UI to do implicit eye tracking calibration
- Hypothesis 3: implicit spoof calibration is robust and can perform like traditional explicit calibration
- Hypothesis 4: it is natural and easy for users to transition between pointing and gaze driven selections

As our eye behaviours and hand-eye coordination change depending on the tasks, and even depending on our visual hemifield [38], it is unclear to what extent all these hypotheses can prove true. We explore all this in a series of studies and find that implicit calibration brings a good trade-off between accuracy of eye tracking and simplicity for the users.

In sum, Online-EYE contributes a multimodal calibration method based on UI interaction. Building on prior work of eye-hand coordination in UI interactions, we first validate our hypothesis across different target sizes and densities of UI elements. We explore how our implicit calibration compares to traditional explicit calibration. We show Online-EYE approach is feasible and fast, − it starts calibrating straight away without waiting for a large number of training data− requires only 8-9 clicks to achieve enough accuracy to enable gaze-and-pinch. Making it a contender and alternative eye tracker calibration method for XR enabling a transition directly from initial controller, or potentially direct touch, to use of gaze-driven selections. We contribute an evaluation of the idea using controllers in VR, but we expect results to be generalizable to any manual pointing device, including controller, direct touch, and mouse, and any type of XR where the position of the UI elements are known.

2 RELATED WORK

Our contribution is based on a number of studies that partially support our hypothesis. But also that collectively contribute to the ongoing development of eye-tracking calibration techniques, each offering unique insights and potential avenues for future research and improvement.

2.1 Hypothesis 1: People Look Where They Point

Despite we also dive into testing this hypothesis to characterize the limits of this assumption, we aren't the first ones to realize of this fact, and there is mounting prior work that both supports and builds on this aspect.

Our vision serves as both a source of information and a guide for our actions [38]. People tend to look ahead to where they will grasp or place objects, to plan and execute movements [14, 27, 30]. It has been demonstrated that gaze can be used to predict target selection [9, 70]. Gaze direction naturally aligns with intended selection points, with strong connection between gaze and cursor or direct touch positions [8, 24, 32, 71].

Eye movement has long been proposed as a fast and convenient method for human-computer interaction [4, 26] precisely because of the visuo-motor guided coordination of our actions [38]. Perhaps the most obvious proof of this is the seamless gaze + pinch interaction we have seen emerge in commercial VR/AR [46]. Eye-hand coordination has enabled natural interaction techniques for selection [67], menu interaction[36], and 3D object manipulation [35].

Overall, these prior works already supports the validation of our hypothesis 1. In our paper we further explore the limits of this assumption in our particular context of UI calibration in VR via experimentation.

2.2 Hypothesis 2 and 3: Explicit and Implicit Calibration

Despite accuracy limitations, eye tracking has the potential to be a more efficient and comfortable input method for AR/VR interactions, faster than head or controller, less fatiguing, and more likely to adopt [13]. Traditional calibration methods require users to explicitly fixate on a set of calibration points, the gaze estimates are then fitted onto the ground truth locations [17]. However, this type of calibration over-relies on the users following correctly the instructions, and actually looking at the targets [33]. If they don't complete the calibration correctly the system will fail [49]. Furthermore, calibration may drift change due to headset slippage. While some eye trackers are robust against slippage, others may exhibit 0.8-3.1 degrees of error increase due to participant behaviour such as speech, facial expressions, and movement [41]. Therefore, methods for obtaining robust calibration more independent of users performance on an explicit task have long been researched. Some approaches have tried including data from multiple session stages [52] or to temporally interpolate gaze locations [3]. A challenge to overcome these issues this way is that explicit calibration interrupts users from their tasks, and sometimes works only offline when calibration data from different temporal stages are utilized. All in all, showing that explicit calibration, despite their prevalence, are still a barrier to eye tracking.

The only alternative then is to make calibration better via implicit techniques. In this context, Implicit calibration techniques have gained attention for their potential to automate the process without bringing users out of their regular device use. Huang et al. [25] introduced PACE, an auto-calibrating eye tracking leveraging user interactions, fixation dynamics, that uses a random forest classification to validate gaze data points for calibration. Pi et al. [48] proposed a task-embedded online calibration method to enhance robustness to head motion, utilizing a transformation matrix and a weight term to optimize calibration. They do so by comparing the similarity between current data to data history. Papoutsaki et al. [42] developed WebGazer, a scalable webcam eye-tracking system that employs online self-calibration using clicks, fixations, and cursor movements. Fares et al. [12] also explored cursor control using eye movements and local calibration points. However, all these projects were ran on regular PCs, and while they might translate to Head Mounted Displays (HMD) the nature of the interactions in VR are very different, as for example, there is no general use of mouse for input for VR.

What is clear is that since natural interactions with UIs, particularly in immersive AR/VR environments, involve complex eye-hand coordination, implicit calibration can become quite complex too. As such, there are many elements that can influence this type of calibration. Sidenmark and Lundström's research demonstrated that the probability of fixation on an object is influenced by factors such as object movement, task precision, and feedback [56]. The eye-hand coordination poses a challenge to identify the optimal timing to enable UI calibration [45], the trade off between the time required to collect many data points, and the accuracy of the calibration is also an consideration.

To reduce the need to optimize timing between eye and interaction, other implicit calibration methods leverage visual saliency [19, 20, 34, 43, 60, 73], however that has proven to be very inter-subject variable and still very passive, an alternative of a bit more robust passive visual attention calibration can focus on smooth-pursuit eye movement with moving targets [16, 47, 50, 64], or on and task specific natural gaze patterns [1, 68].

Other work has proposed use of gaze in ways that does not require calibration. Vidal et al.'s Pursuits method tracks the user's eye movements in relation to the movement of objects on the interface, robustly detection attention based on motion correlation without need for accurate gaze estimation [65]. Esteves et al. applied the principle for input at a glance on smartwatches [11], and Sidenmark to infer attention to objects that move in the depth plane [55]. These works are related in the spirit of facilitating gaze interaction out of the box, but they are limited in that they rely on animation of UI elements.

2.3 Hypothesis 4: Gaze and Hand Transitions Naturally

Eye and hand naturally work together, with intuitive division of labour of gaze pointing and hand confirmation [36, 39, 59]. The two modalities can be combined seamlessly for interacting with devices, complementing direct-touch [44], manipulating XR objects [35, 46], selecting regions [54] and disambiguating objects in 3D space [7, 66]. Hand input complements gaze in multimodal cascaded interactions that harnesses the speed of gaze and precision of hand, manually transitioning between the two modalities [29, 74]. Additionally, the coexistence of controller and gaze pointers can allow users to seamlessly toggle between the two input modalities based on their preferences [63]. Based on all the prior work supporting natural gaze and hand interaction, we build online-EYE to both leverage our implicit calibration on this by using the ray of the controller, (or potentially the hand ray), as well as enable an easy back and forward from hand ray to gaze + pinch type of input for interaction once implicit calibration is achieved.

3 HYPOTHESIS 1: PEOPLE LOOK AT WHERE THEY CLICK

To evaluate this hypothesis we prepared a first study (Study 1) that presented participants (n=9, 4 females) with different targets inside a VR headset, we then asked participants to hit the targets with the VR controller pointer. We recorded where people were looking at, the target location and the controller pointer hit location. This enabled us to find the offsets on hit and gaze on this task for targets of different size, distance and density. Figure 3 shows the average performance of all participants when presented to the variety of targets.

As it would be expected, smaller targets may require more precision, and targets that are more densely positioned might also influence alignment. Furthermore, eye trackers are typically calibrated at a fixed distance, often around 2 meters. In contrast, direct touch interactions occur much closer, typically at arm's length (around

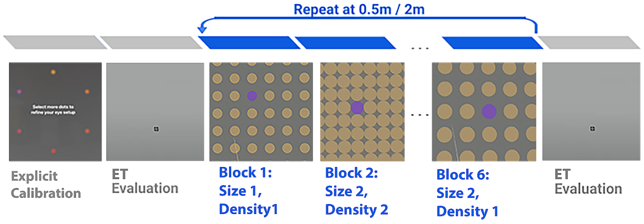

Repeat at 0.5m / 2m
Explicit Calibration
ET Evaluation
Block 1: Size 1, Density1
Block 2: Size 2, Density 2
Block 6: Size 2, Density 1
ET Evaluation

Figure 2: Procedure of Study 1. Participants perform the device native calibration before experiment and evaluated the eye tracking accuracy with the head-locked target fixation task. Then, they complete the counter-balanced target selection task at 0.5m and 2m. To conclude, the fixation evaluation task is performed to confirm eye tracker drift during the study.

0.5 meters). Therefore we also compare how render distance affects selection precision and gaze-target alignment.

3.1 Procedure

Figure 2 illustrates the procedure for Study 1. Participants put on the Meta Quest Pro headset and were instructed to adjust the fit and interpupillary distance. Eye movements were tracked using the headset's built-in eye tracker, and participants performed the native device calibration. Then we completed an Eye Tracker Accuracy evaluation with a head-locked target task. In this task, participants were instructed to look at appearing targets as closely as they could. 25 head-locked targets were shown to the participants, spanning a 5x5 grid of ±15 degs and appeared one at a time for 2s each. Target design follows Thaler et al. [62], 1.5° in diameter with a cross hair and an inner dot of 0.2°. Participants performed the fixation task at 0.5m and 2m render distance. We perform the same accuracy evaluation with head-locked fixation task before and after the study, to see if the eye tracker has shifted. We measured eye tracking accuracy as the mean visual angle difference between gaze samples and target positions [21], over the 1-1.8s period after target onset, in accordance with the approach of previous work [22, 53, 69]. We employed a Meta Quest Pro headset for the experiment. Its eye tracking accuracy has been evaluated to be 1.65° (SD:0.17) where head is free to move, and 2.16° (SD: 0.69) when head is restrained, supporting it as a capable tool for studying visual attention in VR [69], providing reliable tracking capabilities for our experiments.

Three sizes were tested: 2.3 degrees, 4 degrees, and 6 degrees (approximating small, medium, and large UI elements). Two densities were tested: 0 degrees separation (touching) and 3 degrees separation (measured between the closest points on circumference). An example is shown in figure 2.

Participants were instructed to use a VR controller to select targets in the virtual environment. The field of view was filled with a grid of circular objects. The target was marked in purple, while others were yellow. When the controller's ray collided with a target, visual feedback was provided. When the participant correctly selected a target by pulling the trigger button, a new target was marked purple. Target positions were arranged in a 3x3 grid spanning approximately ±15 degrees with a random sequence. The process continued until all targets were selected with 5 repetitions.

Each participant completed 2 render distances × 3 sizes × 2 densities × 9 targets × 5 repetitions = 540 trials. The conditions were counterbalanced.

3.2 Results

The overall mean eye tracking accuracy across all participants was 1.77° (SD:0.74),

We found that the offset (euclidean distance) between where people look and the target center increased with target size, but not with target density or render distance.

For gaze-target offset, a significant main effect of target size was found (F(2, 48)=6.33, p=0.004). Tukey's post-hoc test revealed that the 6° target size resulted in significantly larger gaze-target offsets compared to the 2.3° target size (p=0.003). No other pairwise comparisons were significant. The main effects of density and the interaction effect between size and density were not significant (p>0.05).

Similarly, a significant main effect of target size was observed for gaze-hit position offsets (F(2, 48)=7.55, p=0.001). Tukey's post-hoc test again showed that the 6° target size led to significantly larger offsets compared to the 2.3° target size (p=0.001). The main effects of density and the interaction effect between size and density were not significant (p>0.05).

No significant difference was found between gaze-target and gaze-hit position offsets at any of the target sizes, between the 0.5m and 2m render distances, or between the 0 and 3 degree target densities ($p > 0.05$). These results are shown in Appendix Figure 11 and 12.

Overall these results point in the direction that people look at where they are clicking. The ET accuracy remained the same in our pre-post experimental evaluation. In the subsequent experiment, we further test when H1 in naturalistic settings, testing against typing, scrolling and drag-and-drop beyond just buttons.

3.3 Fixation Analysis

For accurate calibration of gaze based on UI selection data, two key criteria need to be met: stable gaze (fixation) and proper timing. First, the system must exclude saccades and only use gaze that is in fixation.

We removed system invalid gaze samples and employed the IDT algorithm to detect fixations [51], with a minimum duration threshold of 70 milliseconds and a dispersion threshold of 1.5× the inter-sample root mean square (RMS) of measured gaze data (Figure 4).

Second, precise timing is crucial to avoid inaccuracies. Due to potential "late-trigger errors" [28], the selection timestamp might not coincide with the moment of gaze fixation on the target. The user's gaze may have moved away before the selection occurred [25, 56]. To address this, the system traces back from the selection event to identify the first time the cursor hovered on the target. Then all valid fixations from that first hover to selection are measured against the target center, the one with the shortest distance is chosen for calibration. We define the last valid fixation for UI selection as one that either contains the selection timestamp itself

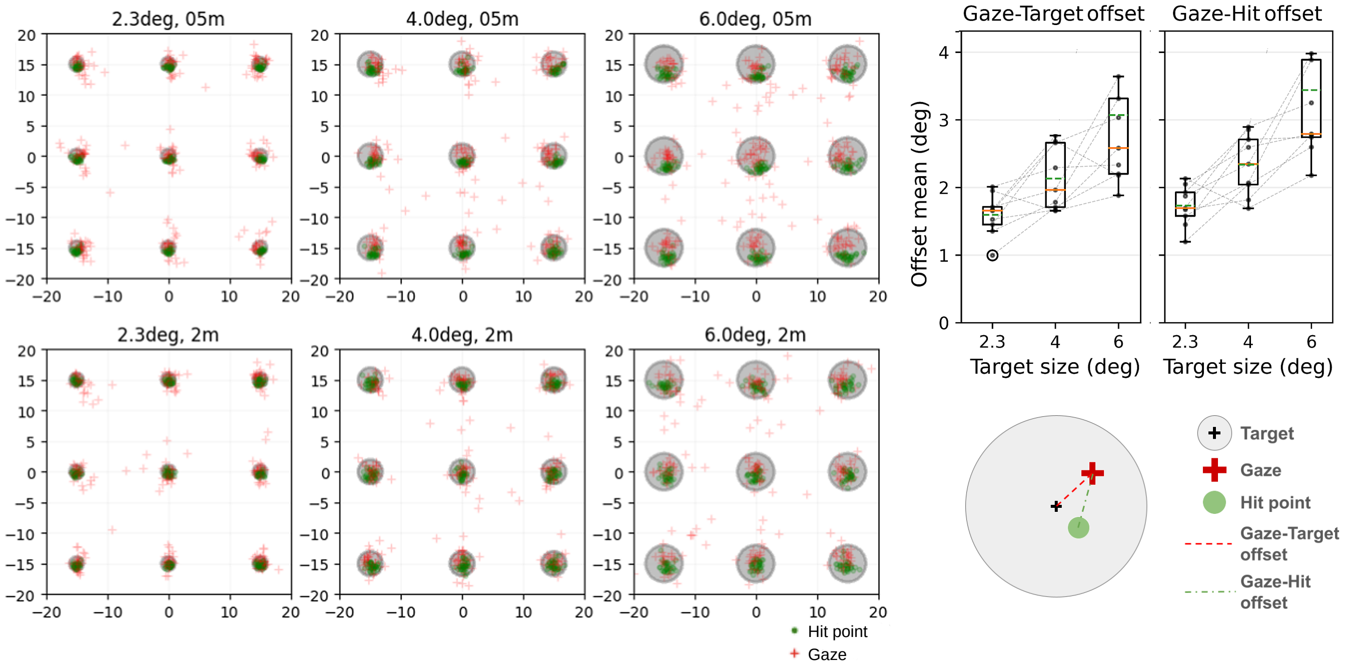


Figure 3: To test H1, we run Study 1, that formalized where people look at when asked to click on different targets. Each data point is alpha blended with 0.2. Gaze-target and gaze-hit position offsets increased significantly as target sizes increases. Gaze-target and gaze-hit offsets measure the angular distance between gaze position and the target center or hit point, respectively. These metrics help determine the optimal ground truth position for implicit gaze calibration. We found no significant difference between the two. Data is collapsed across render distances in analysis.

or ends no earlier than 300 milliseconds prior to the selection event. Furthermore, the centroid of the fixation is < 10 degrees from the selected UI. This ensures we capture the user's true intention when selecting the element, even if a slight delay occurred between fixation and selection. The IDT thresholds and fixation error thresholds were chosen empirically during pilot testing. The 300ms timing is also influenced by prior work [25, 42].

4 HYPOTHESIS 2: WE CAN USE A REGULAR UI TO PERFORM IMPLICIT EYE TRACKING CALIBRATION

We designed a second study to explore how good an implicit calibration based on UI targets could be if HMDs came with a Generic Model of Eye Tracking. To simulate the scenario of headset with a factory default gaze position estimate, each participant used the previous participant's explicit calibration as a simulated factory default Generic Model. When using this device we would run an implicit calibration to refine this gaze model estimate to fit the individual user as they interact with various UIs, potentially eliminating the need for manual explicit calibration.

4.1 Implicit Calibration

Using any target we can calculate a new form of spoof calibration that works implicitly as a user is selecting different UI elements.

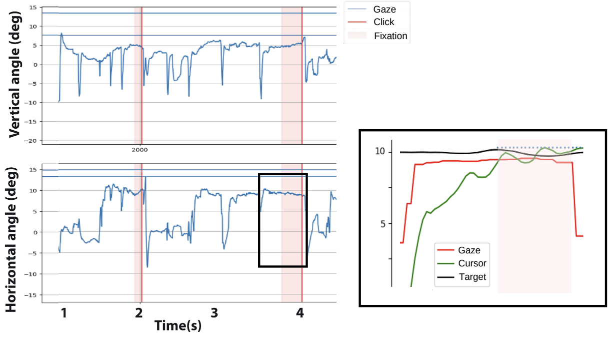


Figure 4: Representative example of gaze analysis showing areas where fixation is detected and a zoomed in plot with the distance to the target and the cursor.

Once we detect eye fixations we find best matching gaze-UI element pairs that will later be used to train a machine learning algorithm to calibrate the implicit Eye Tracking.

The calibrated gaze positions, denoted by $x_{calibrated}$ and $y_{calibrated}$, represent the refined horizontal and vertical components of the user's gaze, respectively. These are obtained through a linear transformation of the measured gaze components, x (horizontal) and y (vertical), according to the following equations:

$$x_{calibrated} = A^x x + B^x y + C^x,$$
$$y_{calibrated} = A^y x + B^y y + C^y,$$

where A, B, and C (with subscripts for horizontal and vertical) are weight coefficients that need to be determined during the calibration process.

To achieve this calibration, we employ a design matrix, denoted by $\mathbf{X}$, which incorporates the measured gaze data. Each row of $\mathbf{X}$ represents a single gaze observation in the form $[x, y, 1]$. The ground truth for the calibration process is represented by a vector, $\mathbf{Y}$, in the form $[x_{groundTruth}, y_{groundTruth}]$, containing the corresponding UI positions that the gaze points to.

Our objective is to establish a mapping between the measured gaze data (in $\mathbf{X}$) and the actual UI positions (in $\mathbf{Y}$). This is achieved by minimizing

$$||\mathbf{Xw} - \mathbf{Y}||_2$$

where the weight matrix, $\mathbf{w}$, contains the A, B, C coefficients. Using a recursive least squares (RLS) algorithm we can update the weights in a computationally efficient approach. Iteratively we update the weight matrix, $\mathbf{w}$, with each new gaze observation, effectively adapting the calibration as more data becomes available [18][1]. We employ two RLS processes, one for each of the horizontal and vertical gaze directions, so we are fitting for six weight coefficients: A^x, B^x, C^x to calibrate gaze in the horizontal direction, and A^y, B^y, C^y to calibrate gaze in the vertical direction.

4.2 Participants

22 participants took part in the study. One was removed due to bad data quality. 11 were females. 14 were between 18-25 years, 3 were between 25-35, 3 were between 35-45, and 1 was 45-55 years. 11 had good and uncorrected vision, 4 wore glasses, 6 wore contact lenses. 7 participants never used a VR device before, 13 used one occasionally and 1 used one weekly. 19 participants have never used an eye tracking device before, 2 used one occasionally. 2 participants never played video games, 11 played occasionally, 4 played weekly and 4 played daily. Participants rated their comfortableness with using a VR device as 3.57 (SD:1.05), on the scale of 1 to 5, 5 being very comfortable.

4.3 Procedure

Fig 5 illustrates the procedure for Study 2. Participants were instructed to put on the Meta Quest Pro headset and adjust the fit and interpupillary distance. Gaze positions were tracked using the headset's built-in eye tracker. Participants began the experiment without undergoing any calibration process, the eye tracker outputs gaze position using the previous participant's calibration. Gaze position and controller states were recorded throughout the tasks. The ID and position of an UI element was recorded when the controller ray collided with it. Trigger pulls and UI selections were tracked.

Participants performed different UI tasks, at the end of each UI tasks, we conducted the fixation task to evaluate the eye tracker accuracy.

We explore the following UIs and tasks:

- Buttons: Participants select radio buttons to complete a multi-choice questionnaire. Buttons are 2° in diameter and the interactive UI area is approximately 5° to 16° horizontally and 0.5° to 22° vertically. At least 40 clicks are collected per participant.
- Slider: Participants adjusted sliders to modify the appearance of an avatar. Each slider handle is 1.5° in diameter. The interactive area is approximately −7° to 1° horizontally and 12° to −29° vertically. At least 30 clicks are collected per participant.
- Drag-and-drop: Participants solved puzzles by moving puzzle pieces to their correct locations. Each puzzle piece is 5° wide. The interactive area is approximately −22° to 21° horizontally and 7° to −13° vertically. At least clicks are collected per participant. At least 27 clicks are collected per participant.
- Keyboard: Participants typed a sentence while being able to observe the characters appear in the input field. Each key is 3° wide and 4.5° tall. The interactive area is approximately −15° to 21° horizontally and 0.87° to −17° vertically. The sentence is "this experiment is really interesting". Participants are allowed to use "Delete" button freely. At least 33 clicks are collected per participant.
- Password: Participants typed the same sentence, but with asterisks masking the characters, simulating password entry. Keyboard layout is the same. Participants are allowed to use "Delete" button freely.

We also explore if UIs rendered at 0.5m (potential direct touch applications) and at 2m, affect behaviour and calibration accuracy differently. 11 participants completed the study at 0.5m render distance, and 11 participants completed the study at 2m render distance. The order of the UI tasks was counterbalanced.

At the end of the experiment, participants completed the native explicit device calibration process, and repeated the head-locked fixation ET evaluation task again. This allows us to measure the explicitly-calibrated eye tracking accuracy for each participant, and compare against an implicit online calibration approach.

4.4 Results

To see if clicks on UI elements can really be used to improve ET calibration like an explicit calibration we evaluated eye tracking accuracy under three distinct calibrations: our General Model proxy, Implicit –using our spoofed targets–, and Explicit calibration.

In general we find (Figure 6) implicit calibration is more accurate than General calibration. The grand mean of eye tracking error for Implicit calibration is 1.94° (SD:0.73°), significantly lower than the grand mean of General calibration 3.08° (SD:1.66°) (t-test=-2.80, p=0.008, df=40). However the grand mean of Explicit calibration is 1.19° (SD:0.56°), also significantly lower than the grand mean of Implicit calibration (Fig 6a). Which would seem to indicate that Implicit calibration would still over-perform an spoofing method.

[1] We adapted the code by Craig Kleski, available at https://github.com/craig-m-k/Recursive-least-squares

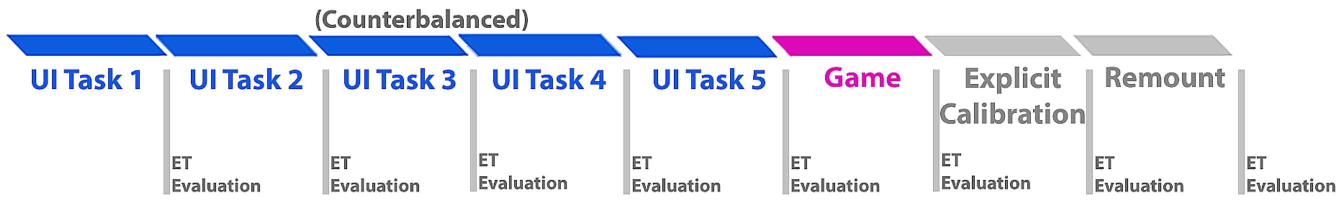


Figure 5: Procedure of Study 2. Following counterbalanced UI tasks, participants performs head-locked fixation eye tracking evaluation task. Real-time performance of Online-EYE is explored in the game application. After the completion of all UI tasks, native device calibration is performed and evaluated, and finally the evaluation of eye tracking accuracy after remounting.

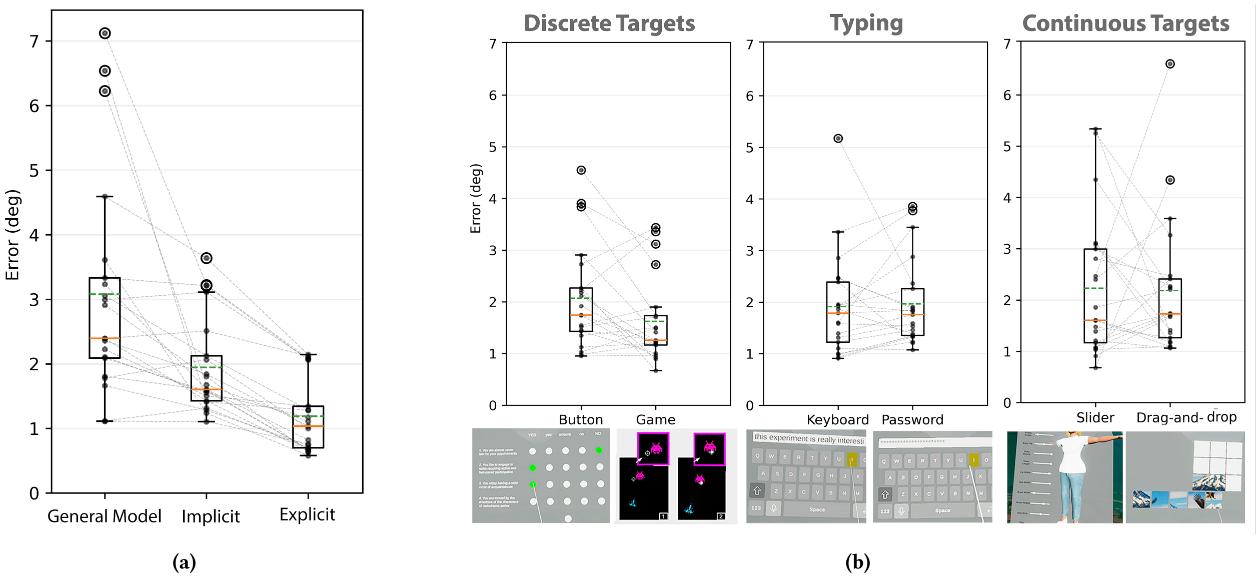


Figure 6: 6a: ET accuracy using a proxy of General Model, using Implicit calibration and Explicit calibration. 6b: Implicit calibration accuracy using as different UI elements as targets. Trials with errors more than 2 standard deviations from the mean error were excluded as outliers condition-wise. In total, 439 trials (3.66%) were removed.

However, we further explored different types of UIs and tasks to better understand how human behaviour with UI elements would influence implicit calibration results (Figure 6). While some UI elements require users to fixate precisely on a target (e.g., selecting a radio button), others, such as sliders, might demand less coordination of gaze and selection, or perhaps as in typing, rely on faster shifts in gaze attention.

To test our hypothesis we assessed the calibration using different UI elements as targets: discrete (buttons and game), continuous (slider and drag-and-drop), and typing (keyboard and password). Results of the calibration accuracy achieved can be found in Figure 6b. We found no significant differences in performance between UI elements ($p > 0.05$). We did however see a trend towards higher accuracy using discrete targets. We believe the power of our sample was perhaps too small to see real effects between UI elements.

5 HYPOTHESIS 3: IMPLICIT SPOOF CALIBRATION IS ROBUST AND CAN PERFORM LIKE TRADITIONAL EXPLICIT CALIBRATION

We wanted to further understand the capabilities of Implicit vs Explicit calibration so we delved further into our data. It is well established that users rarely look at the periphery unless they are on a head-fixed ET task [6], like the one we use on our evaluation. Which means that our training data for Implicit calibration did not cover eccentric areas of the Field of View that were later evaluated. To better understand the impact of this concentration of targets around the center of vision of each participant FoV we ran a follow up analysis.

We constructed the area of normal interaction as the convex hull of the gaze positions collected during UI tasks (Figure 7, left)These convex hulls served as reference regions for grouping gaze positions into categories - inside or outside the interaction region. In the head-locked fixation task, we collected evaluation gaze positions, and compared them to participants' respective convex hulls. We classified gaze positions as "inside" if they fell within the hull and "outside" if they fell outside [2].

We found no significant differences between Implicit and Explicit calibration methods for "inside" positions (t-test=1.91, p=0.06, df=40), with mean errors 1.61° (SD:1.06°) and 1.10° (SD:0.57°), respectively. However, as expected, for "outside" gaze positions, Explicit calibration demonstrated higher accuracy compared to Implicit calibration (t-test=4.14, p=0.00017, df=40), with mean errors 1.22° (SD:0.57°) and 2.07° (SD:0.71°), respectively (Figure 7).

5.1 Distance and Timing Dependence

Another issue to make sure implicit calibration is robust is if it works at different distances. Normally explicit calibrations are performed at one predefined distance, which might differ from the distances used in the user interface, usually presented at 0.5 to 1 meter (Figure 7). Therefore, to further asses robustness of this technique we investigate implicit calibration at two distances: 0.5m and 2m. Independent t-test revealed no significant difference in implicit calibrated eye tracking errors between these two distances (t-test=-1.58, p=0.12, df=40).

5.2 Remount

We measured also another axis of robustness on eye tracking by exploring how well explicit calibration persists after a participant takes off the device and then puts it back on, aka remounts. We found no significant differences between eye tracking accuracy pre- and post- remount (t-test=1.12, p=0.27, df=36).

6 ONLINE-EYE IMPLEMENTATION

We further implement Online-EYE to bring all this implicit calibration into real applications. We use the same Recursive Least Square implementation we described priory but run it live as the users interact with the UI (Figure 1).

For that, we attach the RLS process to a forgetting factor ($0 \leq \lambda \leq 1$) to control the balance between how much importance is placed on the current observation versus the memory of previous observations. A smaller λ places more importance on the current observation. The forgetting factor allows the calibration to prioritize recent gaze data while still incorporating valuable information from past observations.

The recursive least squares (RLS) algorithm leverages the Sherman-Morrison lemma to achieve efficient weight updates. Given the design matrix, $\mathbf{X}$, constructed from previous observations, and a new observation, $\mathbf{g}$, with its corresponding ground truth, t, the following update steps are performed:

- given $\mathbf{A} = \lambda^{-1}(\mathbf{X}^T\mathbf{X})^{-1}$, compute,
- $\mathbf{z} = \mathbf{A}\mathbf{g}$, and,
- $\alpha = (1 + (\mathbf{g}^T\mathbf{z}))^{-1}$.
- Update the weight matrix: $\mathbf{w} \leftarrow \mathbf{w} - \alpha\mathbf{g}^T(\mathbf{w} + t\mathbf{z})\mathbf{z} + t\mathbf{z}$.
- Update A matrix for next iteration: $\mathbf{A} \leftarrow \mathbf{A} - \alpha\mathbf{z}\mathbf{z}^T$.

Online-EYE calibration begins after establishing four valid gaze-UI pairs. This process repeats every three selections. Each data point is processed ten times by the algorithm using a decreasing forget factor. The forget factor is initially set to 0.95, then linearly decreased towards a final minimum of 0.45 over ten fitting iterations. Pilot testing showed no gain in accuracy by reducing the forget factor further or with more iterations. This iterative process continuously refines the weight matrix based on new gaze and UI pairs, ensuring the calibration adapts to participants' individual characteristics and any drift and slippage overtime.

6.1 Applications

To assess the effectiveness of the Online-EYE in a real-world scenario, we explored two applications with the same participants. In both we aim to perform an Implicit calibration online, as an effective alternative to traditional explicit calibration.

In both applications, Online-EYE begins once four UI elements have been selected. RLS weights is updated each time the system has collected three controller clicks. The mean gaze-target offset of the three-click-batch is calculated and used to access the quality of calibrated gaze, dynamically switch on/off gaze pointing according to an accuracy threshold. The calibration matrix is then updated using these three clicks' data, according to Section 4.1.

The recursive least squares algorithm ran in a Python backend, which communicated with the Unity game environment via the Lab Streaming Layer (LSL) [15], ensuring real-time calibration updates. Gaze fixation is detected for the 10s period prior to each click. We recorded gaze direction, controller position and ray direction, head rotation and position, and interaction events throughout the game.

6.1.1 *Application 1: Space Invaders.*

We present participants with a space invaders game (Figure 8), where they need to select appearing targets at random, unknown locations. Initially, participants use the controller to point at the asteroids. When implicit calibration reaches a good accuracy threshold (1.5° degrees of error), the game transitions to Online-EYE. If the accuracy was to drop, they pointer would be reattached to the controller. However, participants in this application did not have control over the type of pointing they were using, and in many cases remained unaware that the pointer had transition to gaze.

This application explores both the edges of using a minimal number of targets, as well as the possibility to seamless transition back and forward to gaze driven pointing and controller pointing.

Participants used a controller to direct the reticle of their spaceship's gun. Asteroids and aliens, each measuring 4.5° in diameter, emerged at random locations within +-15 degrees. The objective was to move the reticle onto these targets and destroy them by clicking the trigger button. The selectable area for the asteroids and aliens was based on their circular diameter of 4.5 degrees, even though their actual shapes were not perfectly round.

Throughout the game, consisting of 50 targets, gaze data and target positions were continuously collected while the online calibration ran in the background.

Initially, participants used the controller to aim. When the online calibration achieved $< 1.5°$ error, the reticle transitioned to gaze control, signified by a white dot appearing within the reticle. In order to survey how intuitive it is to notice and understand the system has transitioned to gaze pointing, participants were not

[2] The convex hulls for one example participant are shown in the appendix, Figure 13

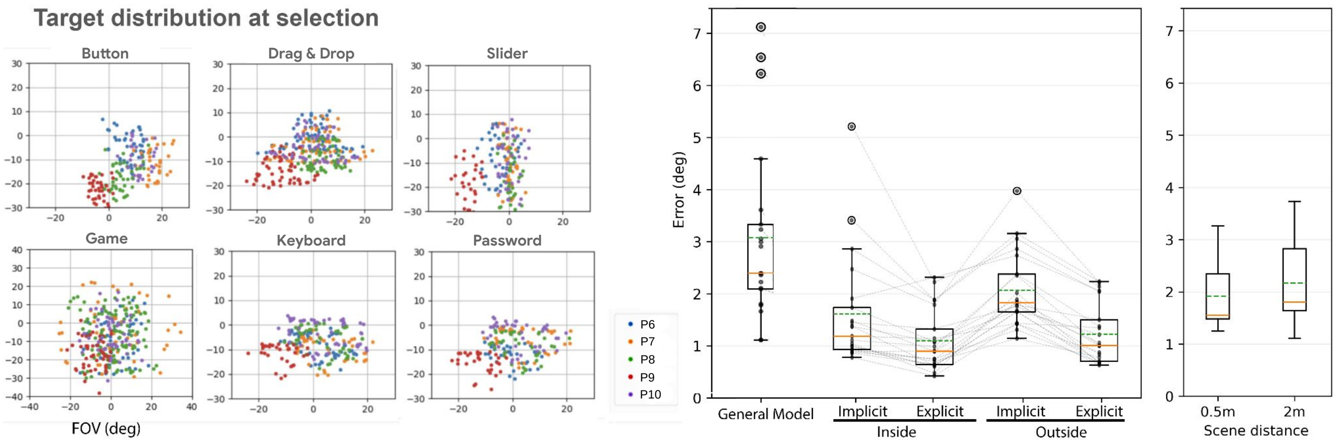


Figure 7: (left) Distribution of locations where the users selected targets in the different UI tasks. (center) Further analysis shows comparable accuracy between Implicit and Explicit calibration inside interaction region. But outside the convex hull, Explicit calibration is more accurate. (right) Results are independent of the UI distance used for Implicit calibration.

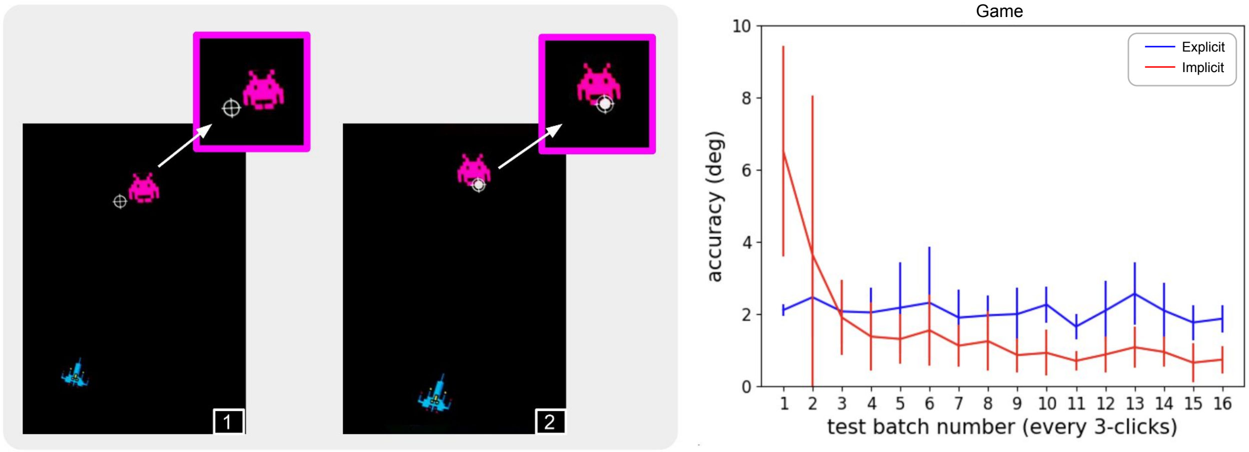


Figure 8: Application 1: appearing targets in the space invaders game. Participants start by using controller to steer the reticle onto the target. Once online calibration is good enough, gaze pointing controls the reticle, with a white dot appearing in the reticle center as feedback (unknown to the participants). Participants could use gaze to aim at targets and pull the trigger to hit. On the right plot an example of the improvement on implicit calibration with Online-EYE: already after 3 evaluations, performing as good as native explicit calibration.

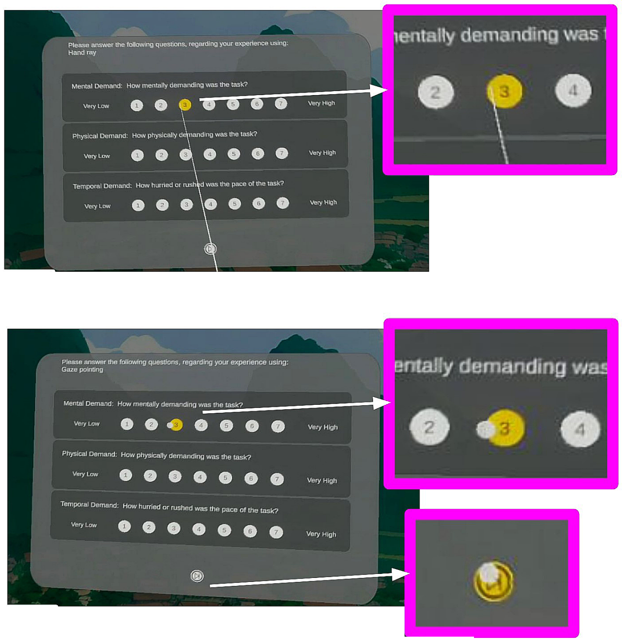
informed about the meaning of the white dot or that gaze-and-pinch has been switched on. We simply asked them to notice any changes in the aiming experience when the white dot appeared.

If the error exceeded the threshold, the system switched back to controller-based aiming, for feedback, the white dot would disappear. If the reticle was not aligned on the target when the trigger was clicked, the system would record an error event. In such cases, the trigger click wouldn't destroy the target, and it would remain on the screen until successfully targeted.

6.1.2 Application 2: Experience Questionnaire. In the second application (Figure 9), participants completed NASA TLX surveys to rate their experience with using controller versus gaze pointing in the space invaders game, as well as a demographic survey.

This application serves three purposes - to collect user feedback for using online-calibrated gaze pointing vs. controller during the space invader game, collect participant demographic information, and to demonstrate the feasibility of using survey (button selection) to enable online-calibrated gaze pointing.

The interface led participants through a series of navigation clicks to launch the surveys. Participant first click on the "Home" button to launch an apps menu. They then open each of the three apps: one for launching a user feedback survey for using controller



Please answer the following questions, regarding your experience using:
Hand ray

Mental Demand: How mentally demanding was the task?
Very Low 1 2 3 4 5 6 7 Very High

Physical Demand: How physically demanding was the task?
Very Low 1 2 3 4 5 6 7 Very High

Temporal Demand: How hurried or rushed was the pace of the task?
Very Low 1 2 3 4 5 6 7 Very High

Please answer the following questions, regarding your experience using:
Gaze pointing

Mental Demand: How mentally demanding was the task?
Very Low 1 2 3 4 5 6 7 Very High

Physical Demand: How physically demanding was the task?
Very Low 1 2 3 4 5 6 7 Very High

Temporal Demand: How hurried or rushed was the pace of the task?
Very Low 1 2 3 4 5 6 7 Very High

Figure 9: Application 2: Participants complete the Likert scale survey to rate their experience using controller and gaze pointing in the space invader game. (top) Participants start by using controller to select answers, with controller ray as feedback. (bottom) Once online calibration is good enough, gaze pointing controls the reticle, cursor becomes a white dot during gaze pointing. Participants could use their eyes to select answers, click on the bottom "Next" button to advance, and complete the survey using online calibrated gaze.

pointing; one for launching the user feedback survey for gaze pointing; and one for launching a demographic survey. The survey's answers are each 2° in diameter.

Initially, participants used controller pointing to select answers. Once online calibration reached the accuracy threshold (1.2°), the survey transitioned to gaze pointing. Note that in this application we reduced even further the error threshold to achieve higher precision.

In this application participants were offered the option to manually switch back to controller pointing from gaze pointing by pushing a key on the joystick on the controller. Participants could use this option when gaze pointing was not accurate enough. Once clicked, the Online-EYE would take over again, assigning the pointing mode depending on the dynamic accuracy evaluation. We recorded the number of times participants manually switched to controller pointing.

6.2 Results

6.2.1 Performance. 21 out of 22 participants were able to turn on Online-EYE in the Space Invader and the subsequent NASA TLX and demographics surveys. The mean number of clicks to turn on Online-EYE is 8.80 (SD:4.52) in the Space Invader and 9.50 (SD:7.02) clicks for the surveys. The number of times the system automatically switched from Online-EYE back to controller due to detecting higher than threshold error was 3.33 (SD:2.25) times for the Space Invader and 1.95 (SD:2.80) times for the survey. Participants seldom manually switched to controller from Online-EYE, on average 0.25 (SD:0.43) times for the survey, showing a quite stable accuracy on the Online-EYE system.

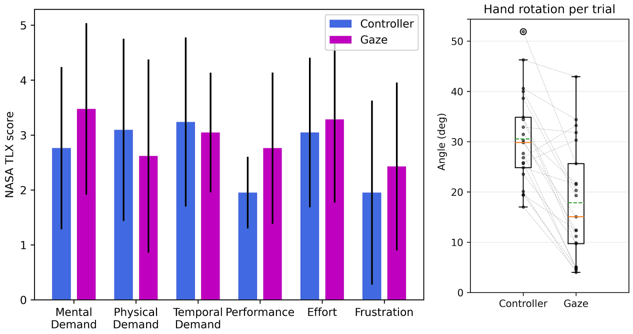


Figure 10: (left) Nasa TLX Likert scale results. (right) Hand rotations performed during the space invaders

6.2.2 Awareness of Transitions. In the Space Invader game, we also explore how participants perceived the automatic transition from controller to gaze, without telling them before the study that gaze was going to take over when online calibration is ready. Instead, we let them figure out how cursor was controlled, and when Online-EYE was switched on. Most participants realized it was gaze pointing immediately. Three participants realized after 3-5 clicks, and two participants never realized until the experimenter told them. Showing that the transition can be quite seamless for users.

6.2.3 Hand Motions. We measured hand rotation as the cumulative angular displacement of the handheld controller per trial. Using Online-EYE on the space invaders, participants showed a significant reduction in hand rotation compared to using controller (Figure 10), from 30.54° (SD:8.97) to 17.85° (SD:11.34) (t-test=3.92, p=0.0003, df=40). We found no significant differences between trial completion time, head rotation, or error rate between using controller or Online-EYE during the Space Invader game ($p > 0.05$). These effects are presumably unconscious as most participants were unaware that they had switched out of the controller to Online-EYE.

6.2.4 NASA TLX. The NASA TLX survey (Figure 10) showed significant difference in Performance rating, participants felt they performed better using controller 1.95 (SD:0.65) than Online-EYE 2.76 (SD:1.38) ($W = 10.5, p = 0.023$). No significant difference between controller and Online-EYE was found for mental, physical, temporal demands, effort, or frustration ($p > 0.05$).

6.2.5 Qualitative Preferences. Overall, 14 out of 21 participants preferred Online-EYE in one or both the applications, 3 of them preferred Online-EYE only for the game but not survey, 2 of them preferred Online-EYE only for the survey but not the game.

The reasons participants preferred gaze-and-pinch included that it was faster (11 participants), accurate (7 participants), easy and natural (5 participants) and required less hand movement (4 participants).

Accuracy was still the main reason participants preferred controller. 2 participants mentioned at edges the accuracy was worse in Online-EYE. 5 participants mentioned they needed to compensate when accuracy was off, by looking away from target so that cursor lands on target. While looking away to compensate allowed participants to complete that trial, it also prevented Online-EYE from learning the correct weights, because they were not looking at the target.

Unfamiliarity with eye tracking was another reason for controller preference, which is not surprising as 19 of the 21 participants have never used an eye tracking device before.

Five participants commented they needed practice to get used to gaze pointing, and three mentioned it got better with time. Two participants were "chasing" the gaze cursor and "went into a spiral" when they tried to look at the cursor and keep chasing it away. P14 mentioned the hand-eye coordination required to operate Online-EYE "I keep on thinking I need to blink to hit". P7 also commented the effect of blinking causes cursor to jitter as "I blink a lot".

There are also comments about the specific implementation of the gaze cursor in this game. Three participants found the gaze cursor always visible to be distracting and caused eye strain, and that it was "weird" when it follows their eyes while reading (during survey). In order to avoid the “late-trigger error” [28] caused by the eyes moving away from target right before selection, we showed the cursor position as the most recent fixation, this trade-off between selection accuracy and delay is noticed by 5 participants who commented that there was a small lag in cursor movement.

7 DISCUSSION

This paper investigated the effectiveness of online calibration for eye tracking in VR, leveraging multimodal inputs to refine the calibration matrix continuously without interrupting user tasks. We compared the accuracy of online calibration to explicit, personalized calibration methods and found that online calibration achieved comparable performance within the interaction area.

7.1 Online Implicit Calibration’s Advantages

A key advantage of online calibration is its potential to reduce the need for explicit personalized calibration, allowing users to start tasks immediately. We simulated a generic eye tracking model that future HMDs could include out-of-the-box by having each participant begin the experiment with the previous participant’s calibration, introducing a diverse, randomized deviation from the generic model. Our results demonstrated a significant improvement in accuracy compared to the generic model, supporting the feasibility of this approach. This could be also beneficial for shared headsets, potentially eliminating the time-consuming process of recalibrating for each user or session.

In the applications, we demonstrated the dynamic evaluation of the eye tracking accuracy to automatically turn on gaze-and-pinch when online calibration reached accuracy threshold. Participants experienced the benefit of speed, reduced hand movement that gaze-and-pinch affords. Additionally, online calibration can effectively calibrate frequently used interaction points, such as the "next" buttons in the surveys application, This can save effort and improve overall user experience, especially when these points are frequently accessed. Online-EYE offers a significant advantage by enabling users to quickly turn on gaze-and-pinch functionality within 8-9 clicks, as demonstrated in our applications. This streamlined process widens the applicability of eye tracking in XR, making it more accessible and user-friendly.

7.2 Robustness Across Tasks

To assess the impact of different interaction types, we explored the performance of online calibration with discrete, continuous, and typing tasks. Our findings revealed no significant differences between these interaction types, suggesting that online calibration is effective across a variety of user interactions. Specifically, we compared the performance of typing tasks with and without visible letter feedback (keyboard vs. password). We did not find a significant difference with visibility of input letters, which could represent task-related distractions. Moreover, we compared the performance of tasks with known target positions (survey) to those with unknown target positions (space invader game). Our results showed no significant differences, suggesting that the knowledge of UI positions does not substantially impact calibration accuracy. These findings strengthen the ecological validity of our tasks, as they represent a diverse range of real-world scenarios.

In these UI experiments, we further tested H1 - people look at where they click, in naturalistic settings against diverse interaction types. H1 posits a more general relationship between gaze and click behavior, considering target size and density, we observed larger offsets with larger targets. While buttons use cases being the point of departure for subsequent UI studies, we found significant transfer of the effect, even though the behavior differed across tasks. This aligns with previous research suggesting that eye gaze often leads hand movement. H1 holds in most button scenarios but may break under specific conditions. For example, in tasks like typing, where prior knowledge of UI layout plays a role, the hand may reach the target before the eyes. People may glance off-target for visual feedback, such as in operating sliders; and may look at the destination during drag-and-drop.

The results suggest that, our approach of leveraging the optimal fixations based on UI interaction timing, is effective across many tasks, especially those requiring precise targeting and selection. For continuous tasks, the action of selecting the UI element still relies on precise aiming and clicking. While previous work has proposed effective smooth pursuit-based calibration [11, 55, 65], we are skeptical that it applied when UI is controlled by the user, because people tend to look ahead at the target [30, 31]. Future research could explore target position for additional calibration opportunities.

Furthermore, we found no significant difference in calibration accuracy between tasks performed at 0.5m and 2m distances. Both groups of participants were able to calibrate and enable gaze-and-pinch during the applications. This suggests task distances in the direct touch range does not significantly affect calibration accuracy, and online calibration could potentially benefit direct touch scenarios.

7.3 Generalizability to Other Input Modalities and the XR Spectrum

Online-EYE could be extended to other input modalities beyond controllers, such as direct touch and mouse. The system leverages the natural human tendency to look at a target before interacting with it, regardless of the input device. While the study focused on controllers due to their robustness and reliability compared to hand tracking, the underlying principle of gaze preceding action remains applicable to other input modalities. The coordination of eyes and hand has been studied extensively [14, 27, 30, 31, 61], with HCI works demonstrating their application in various context including selection, mid-air gesture, and 3D interaction [35, 36, 45, 67].

Our approach relies on interacting with UI elements to calibrate the eye tracker. These elements have known positions in both AR and VR. In AR, UI elements can be part of the OS menu or contextually rendered near real-world objects [10]. The principles of eye-hand coordination remains the same across the XR continuum, and we expect the results to be generalizable to any type of XR where the position of the UI elements are known.

7.4 Limitations

While online calibration offers the advantage of continuous adaptation, it can be susceptible to overfitting to the specific interaction area, potentially limiting its accuracy in other regions of the VR environment. To mitigate this, UI elements could be strategically placed to encourage calibration across a wider FOV. It is possible that participants improved their ability to fixate on the evaluation targets over time, potentially contributing to lower errors in the later evaluation for the explicit calibration. Explicit calibration is evaluated at the end of the study because participants started with the previous participant's calibration as a proxy for a generic gaze model, and that every new calibration irreversibly replaces the previous calibration. However, the primary reason for the higher accuracy of explicit calibration is likely due to its coverage of the entire FOV, while implicit calibration only covered the area spanned by the UI elements, and eye tracking is evaluated over the larger FOV. We did not find significant difference between implicit and explicit calibration accuracy, when gaze is inside the UI interaction area.

Another challenge is user's compensating behavior when eye tracking accuracy is not good enough, users may intentionally look away from the target to make the gaze cursor land on the target. This can hinder the learning process for online calibration. During compensation, the user is not looking at the target, therefore the calibration matrix will never learn the correct weights with these positions. Additionally, participants may be reluctant to use manual switching to turn on controller input, preferring to compensate with eyes instead and stay in gaze mode.

To mitigate these limitations, future research could explore cursor refinement techniques [23, 29, 57] and ways to fall back onto manual input [58]. Alternatively, the coexistance of the controller pointer and gaze pointer could allow users to seamless decide when to toggle between the two inputs [63].

Although Online-EYE enabled gaze-and-pinch within 8-9 clicks in the Space Invader game, it is important to consider the visual characteristics of UI elements in the calibration context. The icon of the alien has many salient points, and users may be more likely to look at the mouth than the eyes. This could lead to their true gaze being closer to the bottom of the UI than the center, contributing to calibration error. While we did not explore the effects of visually salient UI appearances or different visual feedback mechanisms, these factors could potentially further enhance the effectiveness of online calibration.

8 CONCLUSION

This research presents Online-EYE, a multimodal approach to implicit calibration of eye tracking in XR that refines the calibration matrix continuously without interrupting user tasks. Our findings demonstrate the effectiveness and accuracy of implicit calibration, achieving comparable performance to traditional explicit calibration methods across a range of tasks and applications. Online-EYE represents a significant step towards improving the user experience and usability of eye tracking in XR, users can easily enable gaze-and-pinch functionality in as few as 8-9 clicks, potentially eliminating the need for explicit calibration and providing a more seamless integration with other input modalities.

All in all, both our scientific exploration of the underlying hypotheses as well as the application built on these foundations highlight the potential for implicit calibration as a contesting alternative to explicit calibration of Eye Tracking (ET), with the capacity to revolutionize how we have done gaze calibration to date.

Acknowledgments

The authors would like to thank Ishan Chatterjee, Jake Popham, and Seeyam Qiu for their contributions.

References

[1] Fares Alnajar, Theo Gevers, Roberto Valenti, and Sennay Ghebreab. 2017. Auto-calibrated gaze estimation using human gaze patterns. *International Journal of Computer Vision* 124 (2017), 223–236.

[2] Christopher C Berger, Mar Gonzalez-Franco, Ana Tajadura-Jiménez, Dinei Florencio, and Zhengyou Zhang. 2018. Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. *Frontiers in neuroscience* 12 (2018), 21.

[3] Kamran Binaee, Gabriel Diaz, Jeff Pelz, and Flip Phillips. 2016. Binocular eye tracking calibration during a virtual ball catching task using head mounted display. In *Proceedings of the acm symposium on applied perception*. 15–18.

[4] Richard A Bolt. 1981. Gaze-orchestrated dynamic windows. *ACM SIGGRAPH Computer Graphics* 15, 3 (1981), 109–119.

[5] Riccardo Bovo, Steven Abreu, Karan Ahuja, Eric J Gonzalez, Li-Te Cheng, and Mar Gonzalez-Franco. 2024. EmBARDiment: an Embodied AI Agent for Productivity in XR. *arXiv preprint arXiv:2408.08158* (2024).

[6] Charlie S Burlingham, Naveen Sendhilnathan, Oleg Komogortsev, T Scott Murdison, and Michael J Proulx. 2024. Motor "laziness" constrains fixation selection in real-world tasks. *Proceedings of the National Academy of Sciences* 121, 12 (2024), e2302239121.

[7] Di Laura Chen, Marcello Giordano, Hrvoje Benko, Tovi Grossman, and Stephanie Santosa. 2023. Gazeraycursor: Facilitating virtual reality target selection by blending gaze and controller raycasting. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*. 1–11.

[8] Mon Chu Chen, John R Anderson, and Myeong Ho Sohn. 2001. What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. In *CHI'01 extended abstracts on Human factors in computing systems*. 281–282.

[9] Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards Gaze-Based Prediction of the Intent to Interact in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. ACM, New York, NY, USA, Article 2, 7 pages. doi:10.1145/3448018.3458008

[10] Mustafa Doga Dogan, Eric J Gonzalez, Karan Ahuja, Ruofei Du, Andrea Colaço, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented Object Intelligence with XR-Objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–15.

[11] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th annual ACM symposium on user interface software & technology*. 457–466.

[12] Ribel Fares, Shaomin Fang, and Oleg Komogortsev. 2013. Can we beat the mouse with MAGIC?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1387–1390.

[13] Ajoy S Fernandes, T Scott Murdison, and Michael J Proulx. 2023. Leveling the playing field: A comparative reevaluation of unmodified eye tracking as an input and interaction modality for VR. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2269–2279.

[14] J Randall Flanagan and Roland S Johansson. 2003. Action plans used in action observation. *Nature* 424, 6950 (2003), 769–771.

[15] Swartz Center for Computational Neuroscience (SCCN) University of California San Diego. [n. d.]. labstreaminglayer. https://github.com/sccn/labstreaminglayer GitHub repository.

[16] Argenis Ramirez Gomez and Hans Gellersen. 2017. GazeBall: leveraging natural gaze behavior for continuous re-calibration in gameplay. In *COGAIN Symposium: 2017 Meeting and Symposium on Communication by Gaze Interaction: 19th European Conference on Eye Movements*.

[17] Katarzyna Harezlak, Pawel Kasprowski, and Mateusz Stasch. 2014. Towards accurate eye tracker calibration–methods and procedures. *Procedia Computer Science* 35 (2014), 1073–1081.

[18] Monson H Hayes. 1996. *Statistical digital signal processing and modeling*. John Wiley & Sons.

[19] Mamoru Hiroe, Michiya Yamamoto, and Takashi Nagamatsu. 2018. Implicit user calibration for gaze-tracking systems using an averaged saliency map around the optical axis of the eye. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. 1–5.

[20] Mamoru Hiroe, Michiya Yamamoto, and Takashi Nagamatsu. 2023. Implicit User Calibration for Gaze-tracking Systems Using Saliency Maps Filtered by Eye Movements. In *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*. 1–5.

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

[22] Baosheng James Hou, Yasmeen Abdrabou, Florian Weidner, and Hans Gellersen. 2024. Unveiling Variations: A Comparative Study of VR Headsets Regarding Eye Tracking Volume, Gaze Accuracy, and Precision. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 650–655.

[23] Baosheng James Hou, Joshua Newn, Ludwig Sidenmark, Anam Ahmad Khan, and Hans Gellersen. 2024. GazeSwitch: Automatic Eye-Head Mode Switching for Optimised Hands-Free Pointing. *Proceedings of the ACM on Human-Computer Interaction* 8, ETRA (2024), 1–20.

[24] Jeff Huang, Ryen White, and Georg Buscher. 2012. User see, user point: gaze and cursor alignment in web search. In *Proceedings of the sigchi conference on human factors in computing systems*. 1341–1350.

[25] Michael Xuelin Huang, Tiffany CK Kwok, Grace Ngai, Stephen CF Chan, and Hong Va Leong. 2016. Building a personalized, auto-calibrating eye tracker from user interactions. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5169–5179.

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

[27] Roland S Johansson, Göran Westling, Anders Bäckström, and J Randall Flanagan. 2001. Eye–hand coordination in object manipulation. *Journal of neuroscience* 21, 17 (2001), 6917–6932.

[28] Manu Kumar, Jeff Klingner, Rohan Puranik, Terry Winograd, and Andreas Paepcke. 2008. Improving the accuracy of gaze input for interaction. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) *(ETRA '08)*. ACM, New York, NY, USA, 65–68. doi:10.1145/1344471.1344488

[29] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billinghurst. 2018. Pinpointing: Precise head-and eye-based target selection for augmented reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[30] Michael Land, Neil Mennie, and Jennifer Rusted. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception* 28, 11 (1999), 1311–1328.

[31] Michael F Land and Mary Hayhoe. 2001. In what ways do eye movements contribute to everyday activities? *Vision research* 41, 25-26 (2001), 3559–3565.

[32] Daniel J Liebling and Susan T Dumais. 2014. Gaze and mouse coordination in everyday work. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: adjunct publication*. 1141–1150.

[33] Jiahui Liu, Jiannan Chi, and Zuoyun Yang. 2024. A review on personal calibration issues for video-oculographic-based gaze tracking. *Frontiers in Psychology* 15 (2024), 1309047.

[34] Meng Liu, Youfu Li, and Hai Liu. 2020. 3D gaze estimation for head-mounted eye tracking system with auto-calibration method. *IEEE Access* 8 (2020), 104207–104215.

[35] Mathias Nørhede Lystbæk, Thorbjørn Mikkelsen, Roland Krisztandl, Eric J Gonzalez, Mar Gonzalez-Franco, Hans Gellersen, and Ken Pfeuffer. 2024. Hands-on, Hands-off: Gaze-Assisted Bimanual 3D Interaction. In *UIST'24-The 37th Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery.

[36] Mathias N Lystbæk, Peter Rosenberg, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proceedings of the ACM on Human-Computer Interaction* 6, ETRA (2022), 1–18.

[37] Sebastian Marwecki, Andrew D Wilson, Eyal Ofek, Mar Gonzalez Franco, and Christian Holz. 2019. Mise-unseen: Using eye tracking to hide virtual reality scene changes in plain sight. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 777–789.

[38] Antonella Maselli, Eyal Ofek, Brian Cohn, Ken Hinckley, and Mar Gonzalez-Franco. 2023. Enhanced efficiency in visually guided online motor control for actions redirected towards the body midline. *Philosophical Transactions of the Royal Society B* 378, 1869 (2023), 20210453.

[39] Aunnoy K Mutasim, Anil Ufuk Batmaz, and Wolfgang Stuerzlinger. 2021. Pinch, click, or dwell: Comparing different selection techniques for eye-gaze-based pointing in virtual reality. In *Acm symposium on eye tracking research and applications*. 1–7.

[40] Takashi Nagamatsu, Junzo Kamahara, and Naoki Tanaka. 2008. 3D gaze tracking with easy calibration using stereo cameras for robot and human communication. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 59–64.

[41] Diederick C Niehorster, Thiago Santini, Roy S Hessels, Ignace TC Hooge, Enkelejda Kasneci, and Marcus Nyström. 2020. The impact of slippage on the data quality of head-worn eye trackers. *Behavior research methods* 52 (2020), 1140–1160.

[42] Alexandra Papoutsaki, Patsorn Sangkloy, James Laskey, Nediyana Daskalova, Jeff Huang, and James Hays. 2016. Webgazer: scalable webcam eye tracking using user interactions. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI Press, 3839–3845.

[43] David Perra, Rohit Kumar Gupta, and Jan-Michael Frahm. 2015. Adaptive eye-camera calibration for head-worn devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4146–4155.

[44] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 509–518.

[45] Ken Pfeuffer, Hans Gellersen, and Mar Gonzalez-Franco. 2024. Design principles and challenges for gaze+ pinch interaction in xr. *IEEE Computer Graphics and Applications* 44, 3 (2024), 74–81.

[46] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th symposium on spatial user interaction*. 99–108.

[47] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. 2013. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 261–270.

[48] Jimin Pi and Bertram E Shi. 2019. Task-embedded online eye-tracker calibration for improving robustness to head motion. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–9.

[49] Kun Qian, Tomoki Arichi, A David Edwards, and Joseph V Hajnal. 2024. Instant interaction driven adaptive gaze control interface. *Scientific Reports* 14, 1 (2024), 11661.

[50] Argenis Ramirez Gomez and Hans Gellersen. 2018. Smooth-i: smart re-calibration using smooth pursuit eye movements. *citado en la* (2018), 23.

[51] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. ACM, New York, NY, USA, 71–78. doi:10.1145/355017.355028

[52] Shreshth Saxena, Elke Lange, and Lauren Fink. 2022. Towards efficient calibration for webcam eye-tracking in online experiments. In *2022 symposium on eye tracking research and applications*. 1–7.

[53] Immo Schuetz and Katja Fiehler. 2022. Eye tracking in virtual reality: Vive pro eye spatial accuracy, precision, and calibration reliability. *Journal of Eye Movement Research* 15, 3 (2022).

[54] Rongkai Shi, Yushi Wei, Xueying Qin, Pan Hui, and Hai-Ning Liang. 2023. Exploring gaze-assisted and hand-based region selection in augmented reality. *Proceedings of the ACM on Human-Computer Interaction* 7, ETRA (2023), 1–19.

[55] Ludwig Sidenmark, Christopher Clarke, Joshua Newn, Mathias N Lystbæk, Ken Pfeuffer, and Hans Gellersen. 2023. Vergence matching: Inferring attention to objects in 3d environments for gaze-assisted selection. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[56] Ludwig Sidenmark and Anders Lundström. 2019. 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* (Denver, Colorado) *(ETRA '19)*. ACM, New York, NY, USA, Article 6, 9 pages. doi:10.1145/3314111.3319815

[57] Ludwig Sidenmark, Diako Mardanbegi, Argenis Ramirez Gomez, Christopher Clarke, and Hans Gellersen. 2020. Bimodalgaze: Seamlessly refined pointing with gaze and filtered gestural head movement. In *ACM Symposium on Eye Tracking Research and Applications*. 1–9.

[58] Ludwig Sidenmark, Mark Parent, Chi-Hao Wu, Joannes Chan, Michael Glueck, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. 2022. Weighted pointer: Error-aware gaze-based interaction through fallback modalities. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3585–3595.

[59] Sophie Stellmach and Raimund Dachselt. 2012. Look & touch: gaze-supported target acquisition. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2981–2990.

[60] Yusuke Sugano and Andreas Bulling. 2015. Self-calibrating head-mounted eye trackers using egocentric visual saliency. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 363–372.

[61] Brian Sullivan, Casimir JH Ludwig, Dima Damen, Walterio Mayol-Cuevas, and Iain D Gilchrist. 2021. Look-ahead fixations during visuomotor behavior: Evidence from assembling a camping tent. *Journal of vision* 21, 3 (2021), 13–13.

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

[63] Tobii. [n. d.]. Gaze UI. https://developer.tobii.com/xr/explore/gaze-ui/ Tobii XR Devzone.

[64] Subarna Tripathi and Brian Guenter. 2017. A statistical approach to continuous self-calibrating eye gaze tracking for head-mounted virtual reality systems. In *2017 IEEE winter conference on applications of computer vision (WACV)*. IEEE, 862–870.

[65] 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*. 439–448.

[66] Uta Wagner, Matthias Albrecht, Haopeng Wang, Ken Pfeuffer, and Hans Gellersen. 2024. Gaze, Wall, and Racket: Combining Gaze and Hand-controlled Plane for 3D Selection in Virtual Reality. In *ISS'24 ACM Interactive Surfaces and Spaces*.

[67] Uta Wagner, Mathias N Lystbæk, Pavel Manakhov, Jens Emil Sloth Grønbæk, Ken Pfeuffer, and Hans Gellersen. 2023. A fitts' law study of gaze-hand alignment for selection in 3d user interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[68] Zhonghua Wan, Hanyuan Zhang, Mingxuan Yang, Qi Wu, and Shiqian Wu. 2024. Self-Calibrating Gaze Estimation via Matching Spatio-Temporal Reading Patterns and Eye-Feature Patterns. *IEEE Transactions on Industrial Informatics* (2024).

[69] Shu Wei, Desmond Bloemers, and Aitor Rovira. 2023. A preliminary study of the eye tracker in the meta quest pro. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*. 216–221.

[70] Yushi Wei, Rongkai Shi, Difeng Yu, Yihong Wang, Yue Li, Lingyun Yu, and Hai-Ning Liang. 2023. Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[71] Pierre Weill-Tessier, Jayson Turner, and Hans Gellersen. 2016. How do you look at what you touch? A study of touch interaction and gaze correlation on tablets. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 329–330.

[72] Zihan Yan, Yue Wu, Yifei Shan, Wenqian Chen, and Xiangdong Li. 2022. A dataset of eye gaze images for calibration-free eye tracking augmented reality headset. *Scientific Data* 9, 1 (2022), 115.

[73] Songzhou Yang, Meng Jin, and Yuan He. 2022. Continuous gaze tracking with implicit saliency-aware calibration on mobile devices. *IEEE Transactions on Mobile Computing* 22, 10 (2022), 5816–5828.

[74] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 246–253.

A APPENDIX

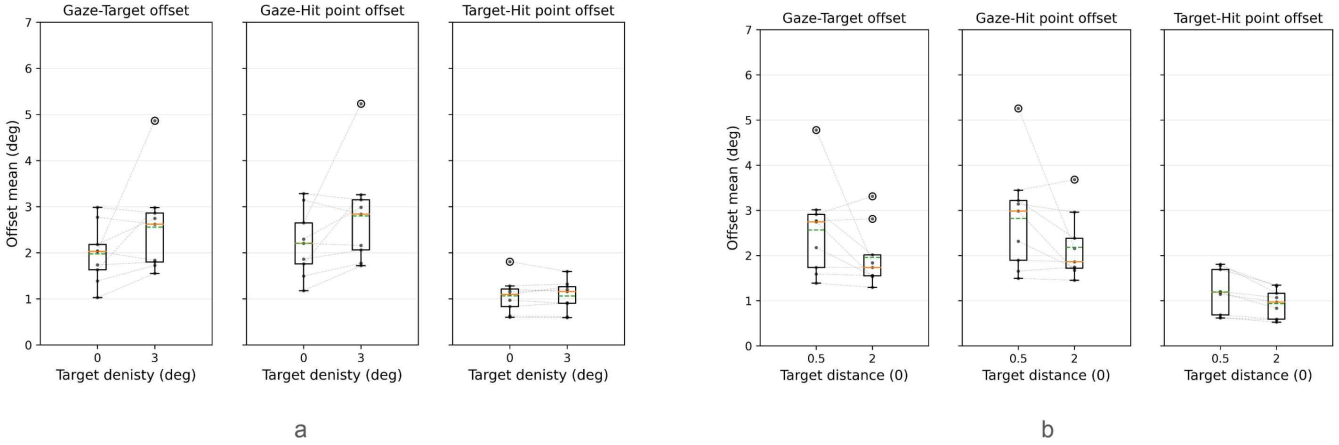


Figure 11: a) No significant differences were found for gaze-target, gaze-hit point offsets between the target densities. b) No significant differences were found for gaze-target, gaze-hit point offsets between the target render distances

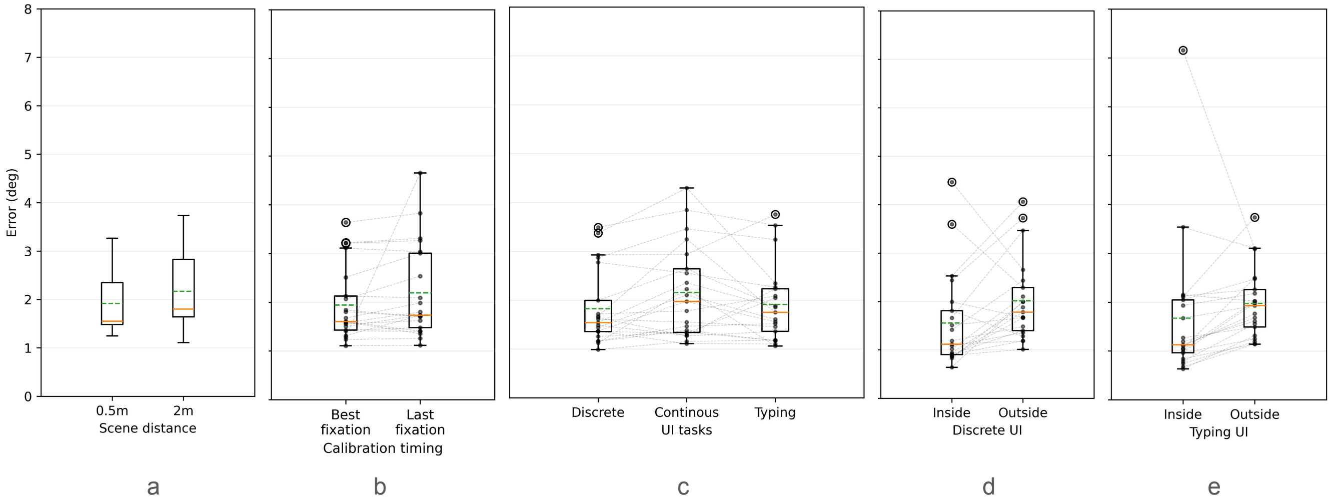


Figure 12: No significant differences were found for: a)0.5m and 2m task distance. b) best or last fixation for the calibration timing. c) Discrete, continuous, and typing task types. c). Inside or outside interaction area of the discrete tasks. d) inside or outside the interaction area of the typing tasks.

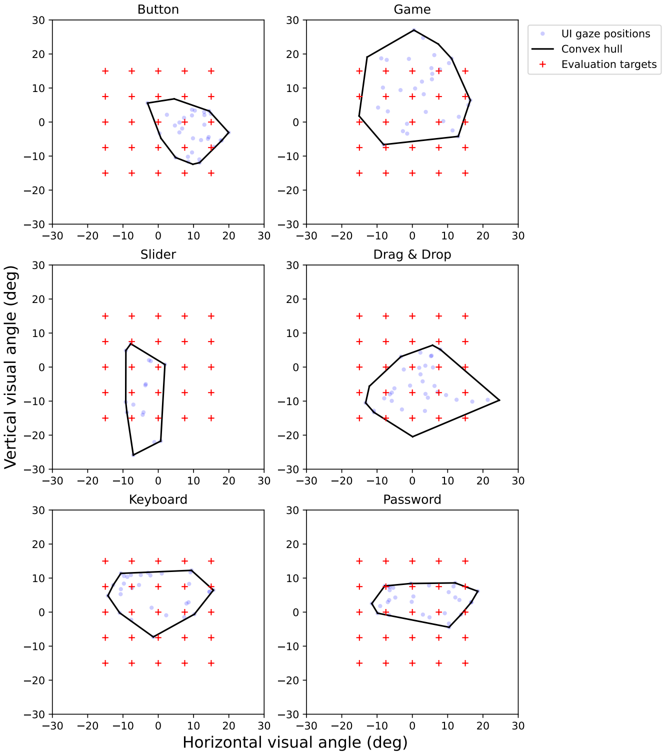


Figure 13: Interaction Region Convex Hulls for Different Tasks for an example participant. Convex hulls are fitted from gaze samples collected during the UI tasks, overlaid over evaluation target positions.