Comparing Gaze, Head and Controller Selection of Dynamically Revealed Targets in Head-mounted Displays

Ludwig Sidenmark, Franziska Prummer, Joshua Newn and Hans Gellersen

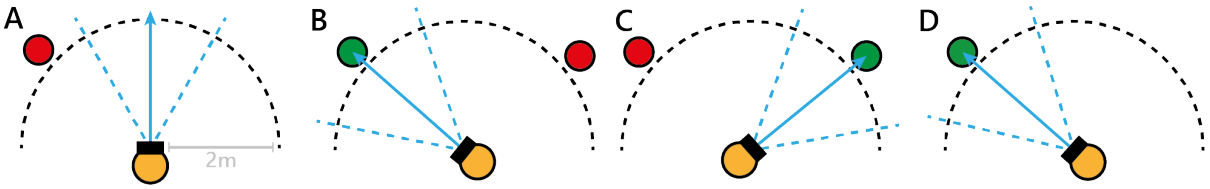

Fig. 1: Our study task sequence to investigate the selection of dynamically revealed targets. A: To start the task, the participant has to search for a start target presented outside their field of view. B: Upon selection of the start target, a first target at an unknown position in the opposite direction, for the participant to select without prior knowledge. C: Upon selection of the first target, a second target is presented at the start position, for the participant to select with prior knowledge of target position. D: Selection of the second target completes the task.

Abstract—This paper presents a head-mounted virtual reality study that compared gaze, head, and controller pointing for selection of dynamically revealed targets. Existing studies on head-mounted 3D interaction have focused on pointing and selection tasks where all targets are visible to the user. Our study compared the effects of screen width (field of view), target amplitude and width, and prior knowledge of target location on modality performance. Results show that gaze and controller pointing are significantly faster than head pointing and that increased screen width only positively impacts performance up to a certain point. We further investigated the applicability of existing pointing models. Our analysis confirmed the suitability of previously proposed two-component models for all modalities while uncovering differences for gaze at known and unknown target positions. Our findings provide new empirical evidence for understanding input with gaze, head, and controller and are significant for applications that extend around the user.

Index Terms—Pointing; Selection Performance; Virtual Reality; 3D Interaction

1 Introduction

Head-mounted displays (HMDs) afford a partial view of virtual environments that can extend further around the user. As a consequence, only targets that appear in the viewport are directly selectable by pointing, while other targets must first be revealed through rotation of the head relative to the environment. The two-step process of first bringing a target into view before selecting it has been described as *Peephole pointing* or *Magic Lens interaction*, originally studied with spatially aware mobile devices [8, 49]. The process is of particular interest for immersive virtual and augmented reality (VR/AR) as HMDs vary in the field of view (FOV) they provide and, therefore, in how much of the environment they reveal [13].

In this work, we compare gaze, head, and controller as modalities for selecting dynamically revealed targets in HMDs. The three modalities constitute alternatives for raycasting, the most common pointing technique for VR/AR [1, 40]. Prior comparison of the modalities has been limited to pointing at targets within view, where gaze has been found faster but less accurate than controller input [18, 58, 61], while head pointing is found more stable and precise than gaze [4, 18, 26, 33]. However, we cannot readily extrapolate from these findings to pointing at dynamically revealed targets, as the three modalities vary in how they are coupled with an HMD. With a head pointer, the same modality is used for both target search and selection, whereas a controller can point independently from the HMD. Gaze is a special case as it can freely point within the current HMD view but not beyond it, and as it intrinsically relies on eye-head coordination [53]. We contribute fundamental insights into how these differences affect the selection of targets that are initially beyond the FOV. These insights are significant for cross-device, spatial applications where users employ devices with varying FOVs and utilise the whole surrounding enviornment.

Pointing performance is usually modelled with Fitts' Law, predicting movement time depending on the distance and width of a target, assuming that the targets are in view and directly reachable [19, 37]. Prior work on Peephole and Magic Lens pointing has proposed extensions of the model to account for search time to reveal targets that are initially outside the screen area, where movement time is additionally dependent on screen width [8, 49]. Recent work has shown that these models also provide a better fit for manual and controller pointing at dynamically revealed targets in HMDs [13]. In this work, we evaluate the applicability of these models also for head and gaze pointing. Specifically for gaze, it has been argued that movement time should not depend on target width as gaze saccades are pre-programmed by the visual system as a ballistic movement [11, 21]. We have therefore also considered Carpenter's model of saccade duration, solely depending on target distance, for the prediction of gaze performance [9].

For our study, we adopted a reciprocal one-dimensional pointing task from previous work on peephole pointing [8, 13, 30]. The task, illustrated in Figure 1, reflects that head movement to reveal targets in virtual environments is predominantly horizontal. It also tests for the effect of prior knowledge of target position, by including a return step from selection of a new target to a previously selected one. The study provides insight into the relative performance with different modalities, notably finding both gaze and controller significantly faster and more accurate than the head. We also gain insight into the effect of FOV, where the increase from 40° to 70° made a significant difference, whereas further increase to 100° did not. Further, we found existing models of Peephole and Magic Lens pointing to be a strong fit for all three modalities. However, for gaze we found selection performance of targets to differ for known versus unknown target positions, with the latter predicted better by Carpenter's model than Fitts' Law.

- *Ludwig Sidenmark is with University of Toronto. E-mail: lsidenmark@dgp.toronto.edu.*
- *Franziska Prummer and Joshua Newn are with Lancaster University. Email: {f.prummer, j.newn}@lancaster.ac.uk.*
- *Hans Gellersen is with Lancaster University and Aarhus University. Email: h.gellersen@lancaster.ac.uk.*

In summary, this work provides the following main contributions:

- Extending understanding of input with gaze, head and controller in VR/AR to pointing at dynamically revealed targets.
- Confirming fit of Peephole and Magic Lens pointing models for pointing with different modalities in HMDs, of significance for prediction of input in 3D interfaces that extend around the user.
- Uncovering differences in gaze pointing at known versus unknown target positions beyond view, of relevance to ongoing debates of gaze pointing models, and warranting further study.

2 Pointing Models

Fitts' Law (Equation 1) has been widely used in human-computer interaction for performance prediction and to assess and compare the efficiency of pointing devices and techniques [37, 52]. Fitts' Law enables estimation of users' performance with a pointing modality by fitting a small set of parameters that predict movement time (MT) for a range of target amplitudes (A) and sizes (W). It can also be used to effectively measure the trade-off between speed and accuracy by calculating throughput from the movement time and index of difficulty (ID, Equation 1) of the performed selection. This aspect is essential in comparing different modalities and techniques through a single metric and has contributed to its widespread usage.

$$MT = a + b\ log_2(A/W + 1) = a + b\ ID \tag{1}$$

Fitts' Law was initially observed for a 1D pointing task with the hand and has since been used to model performance with a wide variety of devices for 2D input, such as the computer mouse [14], and touchscreen [20]. Previous work has shown that Fitts' law extends well to 3D environments, both for direct touch input with a virtual hand, and for virtual pointer input using a controller and raycasting [1]. In this work, we focus on raycasting as a general pointing technique for 3D interfaces, as it enables selection of objects in the environment at any distance from the user [34]. Using visual angles as a unit of target amplitude and width instead of unit metres has proven useful in abstracting performance away from target depth [46]. This has enabled the use of Fitts' law to model controller input in various 3D settings such as Fish-tank VR [59], real-world settings [32], volumetric displays [23], and HMD-based VR [2, 3, 36, 54, 58, 62].

2.1 Pointing with Head and Gaze

Head direction and gaze have long been considered as hands-free alternatives to pointing with manual devices. Users can move their head and eyes with less effort than their arms and hands, and head and eye movement naturally reflect orientation and attention to objects of interest in the environment [5, 40]. HMD displays intrinsically rely on fast and accurate tracking of head movement to provide a continuous experience of VR/AR environments around the user, readily facilitating head pointing by aligning the HMD with a target, usually guided by a cursor shown in the centre of the display [40]. Head pointing has proven to be a stable and accurate modality for selection [4, 26, 47], and has also been shown to conform to Fitts' Law [29].

Eye tracking, in addition to head-tracking, enables selection of a target by gaze, anywhere on a display without need for alignment in the centre. As gaze is used to guide our movement and action in the world, it naturally precedes pointing actions by other modalities and is faster in reaching targets than head or hand [41, 60]. However, eye movement and, in extension, eye tracking is noisy, negatively influencing performance due to accuracy issues [16, 56]. Evaluations of gaze pointing in 3D environments have confirmed that gaze is fast but more prone to error compared to head and controller [18, 26, 47].

Whether gaze pointing is appropriately modelled by Fitts Law has become a matter of considerable debate [11, 21, 51]. Eye movement, unlike head and hand movement, is ballistic in nature. Once the visual system has determined a target, it moves the eyes in a fast saccade without feedback control. It has therefore been argued that gaze performance is better predicted by Carpenter's model of saccade duration (Equation 2), where performance depends on distance (A) but not width of a target [9, 11].

$$MT = a + bA \tag{2}$$

In practice, studies have routinely found gaze to perform in accordance with Fitts' Law [26, 41]. Researchers have suggested that these results are due to secondary corrective saccades when targets are small [51], or artefacts of experimental conditions or analysis methodology [21]. For this work, it also has to be noted that gaze is not always performed by the eyes alone. In particular for interaction over wider fields of view, gaze shifts are frequently supported by head movement [53]. Larger gaze shifts appear pre-programmed to factor in a contribution by the head, reducing the amplitude of the eye saccade. In this work, we specifically consider pointing at dynamically revealed targets which inherently requires users to perform head movements in coordination with eye movements, for gaze selection of targets.

2.2 Pointing at Dynamically Revealed Targets

A key assumption for the applicability of Fitts' law is that targets are visible and known in advance, to eliminate any element of search from the prediction of movement time. The model has been extended for situations in which targets are initially not visible, and only dynamically brought into view. Cao et al. [8] considered the problem of selection in 2D workspaces that are larger than the display through which they can be viewed. Their study was based on a reciprocal 1D pointing task on which we also base our work, and contributed a model for "pointing through a peephole" (Equation 3). Rohs and Oulasvirtta [49] proposed a similar model for pointing through "magic lenses", where a mobile display is moved in the world to reveal augmentations (Equation 4).

$$MT = a + b(n\ log_2(A/S + 1) + (1 - n)\ log_2(A/W + 1)) \tag{3}$$

$$MT = a + b\ log_2(A/S + 1) + c\ log_2(S/2/W + 1) \tag{4}$$

The two-component Peephole and Magic Lens pointing models both reflect the two phases of moving the view to reveal the target, and moving a pointer within the view to complete the selection. In addition to amplitude and target width, screen width (S) is introduced as a third variable on which performance depends. Screen width presents a trade-off, as search time for a target is shorter when the screen is wider, while pointing time is longer when the screen is larger. The models have proven to accurately predict pointing at dynamically revealed targets in a range of 2D settings, including phone pointing [50], desktop panning [39], smartphones and smartwatches [28, 31, 50], handheld projectors [30] and map-navigation [48].

In 3D environments, selection of dynamically revealed targets has not received much attention. Grinyer and Teather investigated out-of-view target search at varying FOVs and amount of targets in the environment and found increased search performance with a wider FOV [22]. Ens at al. have produced the only work formally investigating the problem of dynamically revealed target selection [13]. Their study simulated varying headset FOV with a viewport projected in a CAVE environment, and compared controller raycasting and direct touch as pointing techniques. The results showed a strong fit with the Peephole and Magic Lens models, while performance was observed to increase with screen width, for all width investigated from 8° to 128° visual angle. In our work we adopt the same task design for comparability but study effect of FOV in an actual HMD and with different modalities, i.e. head and gaze in addition to controller.

Head, gaze, and controller differ fundamentally in how they are coupled with a HMD. Head pointing is coupled with viewport control, whereas controller pointing is world-based and independent of the viewport. Gaze in turn is free to point anywhere within an HMD display but not able to point beyond. Prior work on AR pointing found techniques that are reliant on a cursor that moves with the viewport to be less performant than techniques that are decoupled from the viewport [10]. This contrasts with the finding of Cao et al. where

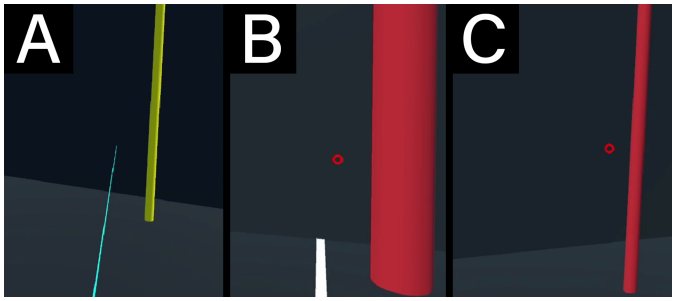

Fig. 2: Implementation of pointing modalities. A: The controller's pointing direction is visualised with a ray emanating from the controller. B: The head's pointing direction is visualised with a small cursor of 1° diameter. C: Similar to the head, gaze is visualised with a cursor.

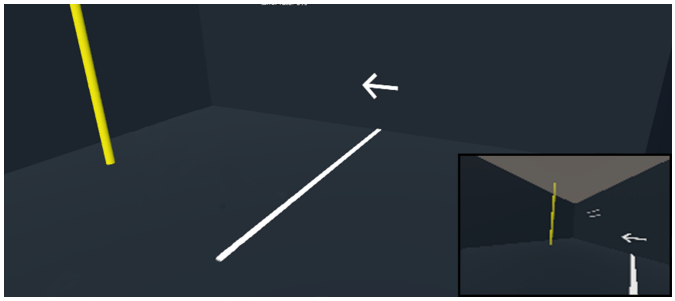

Fig. 3: Study scene and participant point of view. The figure shows the middle direction divider (white line) and the arrow that shows the start target's direction.

a coupled cursor controlled with one hand was faster for peephole pointing than a decoupled cursor in a condition where one hand moved the viewport, while the other was decoupled to point with a stylus on the viewport [8].

There is a variety of other work that has considered targets around the user. Petford et al. compared mouse pointing versus raycasting in a real environment where targets were positioned all around the user, and observed a strong interaction between performance, target location and whether users new target locations in advance [45]. In HMD-based pointing, some studies have included conditions where targets are presented both within and outside the FOV, but these have not been treated differently for analysis [33, 55]. In other work on HMD-based AR/VR, the focus has been on providing visual guidance to targets that are out-of-view [6, 24, 25, 35].

3 Study

Our study aims to extend our understanding of dynamically revealed target selections with commonly found pointing modalities. Specifically, we examine the following modalities, which are frequently employed or becoming more common for interaction in modern HMD-based VR:

Controller-pointing. The controller pointer is a decoupled pointer i.e. it is not related to the HMD. Pointing is performed by pointing via a 6-degree of freedom hand-held controller. In our implementation, the forward vector of the controller is displayed via a ray that intersects with targets (Figure 2a).

Head-pointing. The head pointer is coupled to the centre of the HMD and follows the user's head movements. In our implementation, the head direction is visualised via a cursor (Figure 2b).

Gaze-pointing. The gaze pointer is a semi-decoupled pointer where the user can point freely inside the current view but is not able to point outside the view as with the controller. In our implementation a cursor is displayed at the current gaze direction (Figure 2c).

For all modalities, we used a controller button press for selection confirmation to ensure a consistent confirmation technique. Previous work has also shown higher performance for button clicks compared to dwell-based techniques prevalent in hands-free interaction [42]. The cursor for each modality was always visible during the study.

3.1 Task

We used a similar reciprocal 1D pointing task in the horizontal plane as previous work [8, 13, 30]. Each selection sequence consists of two target selections for data analysis. The first selection is made without prior knowledge of target location, and the second with knowledge of the target location (Figure 1). Knowledge of the second target location is gained by placing the second target at the same location as an initial starting target. All targets appear as cylinders with a significantly higher height than the display height. Each selection sequence varies in amplitude (A) between targets, target widths (W), and display width (S). These parameters were all measured in visual degrees from the participants' perspective. Modality is altered between blocks as per the procedure design. Shown in Figure 1, all targets were positioned 2m from the user in a circular arrangement.

For each selection sequence, participants perform three selections: a start selection to initiate a selection sequence, and two subsequent selections, which are used for data analysis. At the start of a selection sequence, the participant searches for the start target (Figure 1a). The start target's placement to the left or right of a middle direction divider (Figure 3) informs the participant in which direction the start target will be located. When found, participants select the start target to start the selection sequence (Figure 1b). The participant then searches for the first target in the direction opposite to the start target's direction until the target is found and selected (Figure 1c, no prior knowledge of the target position). The participant then returns to the position of the start target to select the second target (Figure 1d, prior knowledge of the target position), which ends the selection sequence. Selection success is indicated via colour feedback. The selection sequence continues when the user successfully selects the target. However, any missed selections before the successful selection of a target renders the sequence to be considered erroneous. Erroneous sequences were re-queued for selection in line with previous work [13, 30].

Our study employed a within-subjects design with five independent variables: pointing modality, target amplitude A, target width W, screen width S, and prior knowledge of target location. Each participant completed a session of selections with each pointing modality. Each session had three sequences for each combination of A, W, and S in random order, resulting in 81 selection sequences. Pointing modality order was counterbalanced with a latin square. Participants performed all blocks with one pointing modality before moving on to the next modality. Each participant performed in total a minimum of 2 selections $\times$ 3 modalities $\times$ 81 sequences = 486 selections for data analysis, not including additional selection sequences required due to missed selections. Participants performed in total 6374 selection sequences (12748 selections for analysis). The study used the following independent variables and levels:

- Pointing Modality: {Gaze, Controller, Head}
- Screen Width (S): {40, 70, 100°}
- Target Amplitude (A): {30, 60, 120°}
- Target Width (W): {2, 4, 8°}
- Prior Knowledge: {Yes (PK), No (NoPK)}

3.2 Apparatus

We developed the study environment in Unity (version 2017.4.3f1). We used a HTC Vive with an integrated Tobii Pro Eye Tracker (120Hz) to record eye and head movements, and the HTC Vive Controller to record controller movements. Participants used the touchpad button on the HTC Vive controller for selection confirmation for all three modalities. The Tobii SDK synchronised the eye and head data. We recorded data at full frame rate and mean gaze accuracy of $0.82 \pm .52°$. The HTC

Vive has a FOV of 100° in the horizontal plane, 110° in the vertical plane, and a frame rate of 90Hz.

3.3 Participants

We recruited 24 participants (11 female, 13 male, 24.21±4.7 years) from our local university. Eight participants had no prior VR experience, 12 reported occasional, 1 reported weekly, and 2 reported daily VR experience. Eleven participants had no prior eye tracking experience, 11 reported occasional, and 2 reported daily eye tracking experience.

3.4 Procedure

On arrival, participants received a short briefing on the study procedure. The participants then signed a consent form and answered a simple demographic questionnaire. Then they were asked to put on the HMD and performed a short training session with the current modality. The researcher began the study once they were comfortable. Participants completed a five-point eye tracking calibration at the start of each modality session before starting the selection sequences. After completing all selections with a pointing modality, participants removed the HMD and completed a Raw NASA TLX questionnaire [7] to record the modality's perceived workload. The participants were given the opportunity to take a short break before continuing with the next modality. The study took 45-60 minutes to complete. The study procedure was approved by Lancaster University's FST Research Ethics Committee. Participants received a £10 Amazon voucher for their participation.

The metrics of interest were:

Error rate: The number of selections that result in an error divided by the total number of selections. An error is defined as a participant that misses the target during a selection prior to the correct selection.

Reach time: The time between the selection start and when the pointer first reaches the target.

Movement time: The time between selection start and a successful selection.

Target overshooting time: The time from when the cursor moves beyond the target until the target is selected.

Head movement: Head rotation from start to successful selection.

Perceived workload: Raw NASA TLX metrics.

4 Results

Unless otherwise stated, the analysis was performed with a 5-way repeated measures ANOVA (α=.05) with Modality, Display width (S), Amplitude (A), Target width (W) and Prior knowledge as independent variables. When the assumption of sphericity 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. The effect sizes are reported as partial eta squared (η_p^2). Raw NASA-TLX scores were analysed using Friedman tests, and Bonferroni-corrected Wilcoxon signed-rank tests were used for the post-hoc analysis.

4.1 Errors

In total, 570 selections were recorded in which participants missed the target before a successful selection (4.4% of all selections). The number of errors was positively skewed and violated the repeated measures ANOVA's assumption of normality after the usual transformations, and the Align Rank Transform technique [63] showed that the aligned responses did not sum to ≈0. Using the number of errors as count data, we fit a Negative Binomial regression model [38] because the variance of errors was larger than the error mean. We report the number of errors as the error rate, i.e. the number of selections resulting in an error divided by the total number of selections.

We included all main effects and all interactions involving Modality in the regression and found that the overall model was significant ($\chi^2(137, N{=}3888){=}2543.71$, p<.001). Investigation of model effects revealed a significant three-way interaction for Modality× S × A ($\chi^2(6){=}14.38$, p=.026), and main effects for Modality ($\chi^2(2){=}13.76$, p=.026), and Prior knowledge ($\chi^2(1){=}8.57$, p=.003). Further analysis did not show significant results for the three-way interaction. For the main effects, sequential Šidák pairwise comparisons showed that controller (1.9%) was significantly more accurate than head (4.5%, p<.001) and gaze (3.4%, p<.001). Gaze was also more accurate than the head (p<.001). Finally, prior knowledge of target position led to fewer errors (2.80%) than NoPK (3.8%, p<.001). In contrast to previous work, we found that target width did not significantly impact the prevalence of errors [13]. These results provide the following insights:

- In contrast to previous work, head pointing is significantly more erroneous than gaze and controller when selecting dynamically revealed targets. This is possibly due to using the head for both search and pointing (coupled).
- The controller was more accurate than gaze, as shown in previous work [18].

4.2 Reach time

We investigated reach time to understand how fast participants could potentially select targets and how much time is needed for the initial pointing phase in relation to the whole selection. We found no 5-way, 4-way, or 3-way interactions. However, the results showed a significant Modality × S interaction ($F_{2.74,63.06}$=29.39, p<.001, η_p^2=.561, Figure 4a). Further analysis showed that gaze reached the target faster than the controller and head at all screen widths (all p<.001). For gaze and controller pointing, an increased screen width led to significantly faster reach times (all p<.001) but no significance was found for the head. The controller only showed significantly faster reach times than the head at 70° and 100° screen widths (both $p\leq$.032), indicating that the decoupled controller benefits from an increased screen width in contrast to the coupled head (Figure 4a).

We also found a significant Modality × W interaction ($F_{4,92}$=9.76, p<.001, η_p^2=.298, Figure 4b). At all target widths, gaze was faster than the head and controller (all p<.001). However, the controller was only faster than the head at target widths of 4° (p=.003) and 8° (p=.004), presumably as more refined movements are performed for 2° targets, thus slowing the controller. Meanwhile, increasing target width led to faster reach times for all modalities (all p<.001).

Finally, we found a significant Modality × Prior knowledge interaction ($F_{2,46}$=7.85, p=.001, η_p^2=.254, Figure 4c). Prior knowledge only affected the reach time of the head (p=.016), where participants were slower with PK as users would perform faster head movements with high amplitude to quickly find the target in the NoPK condition. This behaviour led to an overshoot and therefore to a faster reach time compared to the PK condition, where overshooting would be less likely. Meanwhile, gaze was again significantly faster than head and controller with both PK and NoPK (all p<.001). The controller was faster than the head with prior knowledge (p=.010).

We also found significant main effects. For Modality ($F_{2,46}$=39.99, p<.001, η_p^2=.635), gaze (.66s) was significantly faster than the controller (.78s, p<.001) and head (.82s, p<.001), while the controller was significantly faster than the head (p=.026). For screen width ($F_{1.31,30.13}$=98.76, p<.001, η_p^2=.811), a larger screen size led to faster reach time at all levels (40°: .81s, 70°: .74s, 100°: .71s, all p<.001). For target width ($F_{1.50,34.59}$=341.99, p<.001, η_p^2=.937), larger target widths led to faster reach time at all levels (2°: .81s, 4°: .74s, 8°: .71s, all p<.001). Finally, for target amplitude ($F_{1.44,33.10}$=1143.55, p<.001, η_p^2=.980), larger amplitudes led to longer reach times (30°: .48s, 60°: .70s, 120°: 1.08s, all p<.001). In summary, these results show that:

- Gaze is the fastest to reach the target, while the controller is faster than the head. These results are expected as the eyes move faster than other body parts to guide our actions.
- An increased screen width only benefits gaze and controller pointing. This may be because these modalities are more independent from viewport control than the head.

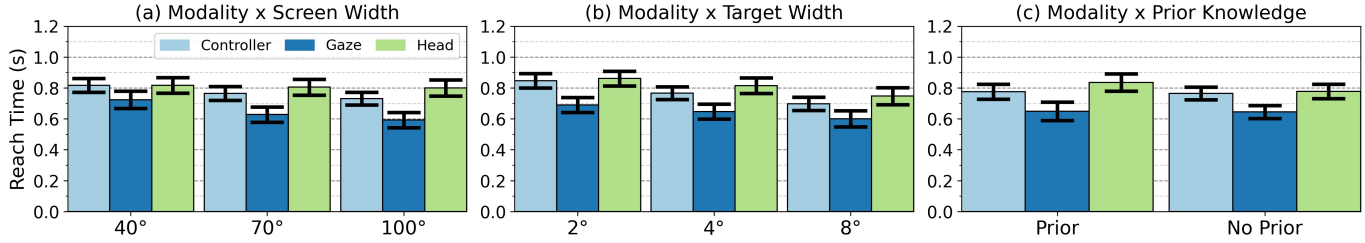

Fig. 4: Reach time interactions. Error bars represent the mean 95% confidence interval.

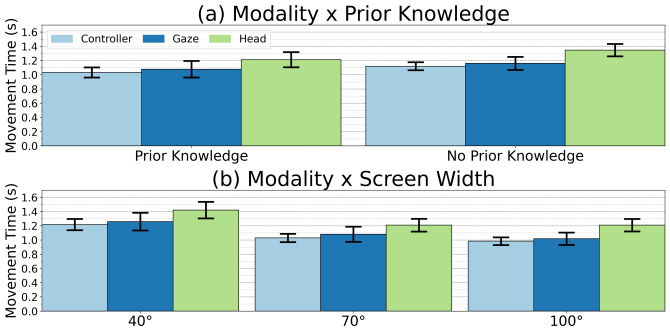

Fig. 5: A: Modality and Prior Knowledge movement time. B: Modality and Screen Width movement time. Error bars represent the mean 95% confidence interval.

- Users are slower in reaching targets when they have prior knowledge of the target location, as the initial movement is more accurate. No prior knowledge leads to users reaching the target faster, as users perform large search movements.

4.3 Movement Time

Only successful sequences with no errors were considered for the analysis of movement time. We found significant main effects for Modality ($F_{2,46}$=19.42, p<.001, η_p^2=.458), screen width ($F_{1.22,28.12}$=131.30, p<.001, η_p^2=.851), Amplitude ($F_{1.33,30.59}$=511.47, p<.001, η_p^2=.957), Target Width ($F_{1.28,29.38}$=285.87, p<.001, η_p^2=.926), and Prior Knowledge ($F_{1,23}$=106.966, p<.001, η_p^2=.823). Pairwise comparisons showed that increased target amplitude led to longer movement times (30°: .85s, 60°: 1.11s, 120°: 1.48s, all p<.001). Similarly, a decreased target width (2°: 1.29s, 4°: 1.15s, 8°: 1.00s, all p<.001), and a decreased screen width (40°: 1.29s, 70°: 1.09s, 100°: 1.06s, all p<.001) also led to longer movement times. Furthermore, participants were faster with PK of the target position than NoPK (Figure 5a). Finally, the results showed that the head (1.27s) was significantly slower than gaze (1.10s, p<.001) and controller (1.07s, p<.001). We found no significant difference between gaze and controller (p=1.000). These results are aligned with previous work on Fitts' law [18], and comparison between coupled and decoupled techniques [10].

We found no 5-way interaction. However, the results showed a significant Modality × S × W × A 4-way interaction ($F_{16,368}$=1.70, p=.045, η_p^2=.069). Further analysis showed that the head was significantly slower than gaze and controller for 70° and 100° screen widths for all target sizes and amplitudes (all $p\leq$.013) but for 40° screen width, the head was only significantly slower for 2° targets (all $p\leq$.025). We found no significant differences between gaze and controller. These results imply that there is less difference between modalities at narrow screen widths but that the decoupled and semi-decoupled pointers can leverage the increased screen width at a larger capacity than coupled cursors. Regarding screen width, selections with 40° screen width were consistently slower than 70° and 100° screen widths under all conditions (all $p\leq$.040). However, we did not find significant differences between the 70° and 100° screen widths. These results imply that increasing screen widths only positively impacts movement time up to a certain point and that the increased screen width will then have a negligible effect on movement time (Figure 5b). Increased target width (all $p\leq$.012) and decreased target amplitude (all $p\leq$.001) both led to significantly shorter movement times for all conditions. Summarising, these results show the following:

- The head is the slowest to select the target. Although the gaze reaches the target before the controller, there is no significant difference in movement time.
- An increase in screen width only significantly affects movement time between 40° and 70° screen widths. We found no significant difference between 70° and 100° which implies a limit to the performance gained from increasing the screen width.
- The head's movement time was not significantly slower than gaze and controller for larger targets in a smaller FOV.

4.4 Target Overshoot Time

Previous work has shown that overshooting time is a key characteristic of the selection of dynamically revealed targets [30]. Hence, we were interested to see how this affects the modalities in our study. The data was positively skewed and we log-transformed the data before statistical tests to comply with normality assumptions. Note that values presented in text and figures represent non-transformed data.

We found no 5-way or 4-way interaction. However, we found a 3-way interaction for Modality × A × Prior knowledge ($F_{4,92}$=10.75, p<.001, η_p^2=.319) and for Modality × S × A ($F_{4.82,110.95}$=2.94, p=.017, η_p^2=.113). Further post-hoc tests for the Modality × A × Prior knowledge interaction showed that at all amplitudes, the controller had a shorter overshooting time than gaze and head in PK and NoPK conditions (all $p\leq$.034). However, gaze had a significantly shorter overshooting time than the head only with NoPK (p<.001). With the controller, participants have time to adapt to the target selection as pointer movement is independent from display movement, while gaze seems faster at recovering from overshoots than head pointing. For all modalities, amplitude only led to increased overshooting time for selections with NoPK, where the highest was for 60° amplitude in comparison to 30° and 120° (all $p\leq$.028). These results are most likely a consequence of users being able to see the targets in the periphery for some conditions with 30° amplitudes and that participants slow down for the 120° amplitudes, thus minimising overshooting. PK led to less overshooting at all conditions (all $p\leq$.013, Figure 6a).

For the Modality × S × A interaction, at 40° screen width (Figure 6b), gaze and controller had significantly less overshooting time than the head at all amplitudes (all p<.001), but we found no difference between gaze and controller. At 70° screen width (Figure 6c), the controller had less overshooting time than the head ($p\leq$.0012) and gaze ($p\leq$.006), except for gaze at the longest amplitude (p=.121). Here we found no significance between gaze and head. At 100° screen width (Figure 6d), the controller had significantly less overshooting time than gaze and head at 30° and 60° amplitudes (all $p\leq$.018). However, there was no significance at 120° amplitude. There was again no significant difference between gaze and head. Regarding screen widths, the head and controller had significantly longer overshooting times at 40° than

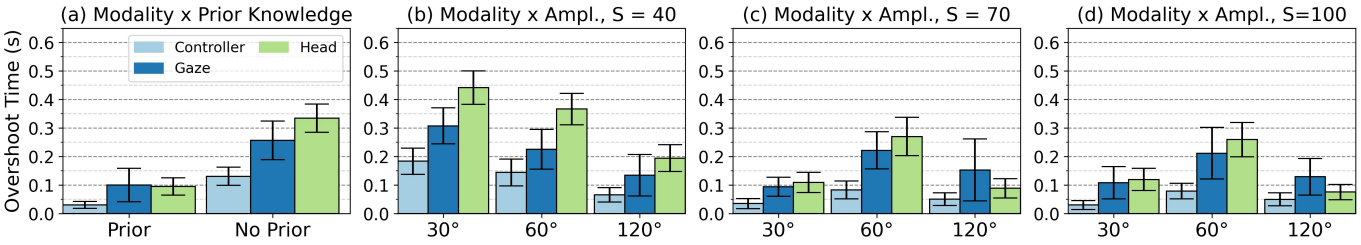

Fig. 6: Overshooting time interactions. Error bars represent mean 95% confidence interval.

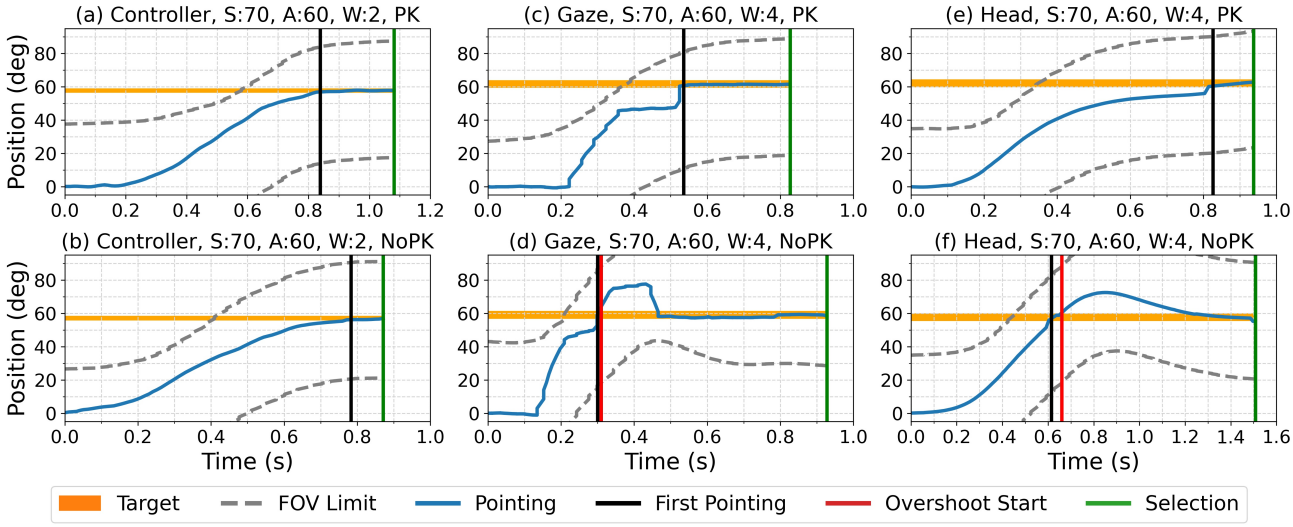

Fig. 7: Example selections. The plots show differences in user behaviour between modalities and with or without prior knowledge of target position.

70° (all p≤.003) and 100° screen widths (all p≤.003) at all amplitudes but no difference between 70° and 100° screen widths. For gaze however, significant differences were only found between 40° and 70°, and 40° and 100° screen widths at the smallest target amplitude (all p≤.001). Presumably, participants had a clearer view of the target with the shortest amplitude from the start with larger screen widths leading to significantly less overshooting times for these conditions. Otherwise, larger screen width did not seem to impact gaze-based overshooting. For the controller and head, we found that increased screen width can reduce overshooting time up to a certain point. Regarding amplitude, we found that for the 40° screen size, increased amplitude led to less overshooting time (all p≤.002). However, for 70° and 100° screen widths, the middle amplitude had larger overshooting time than the others (all $p<.001$). This is again because an increased screen width led to the participants seeing the closest target at the start of the selection, and due to participants having slowed down during the long amplitudes as they were expecting the target to appear soon.

We also found significant main effects. For Modality ($F_{2,46}$=22.33, $p<.001$, η_p^2=.493), the controller (.08s) had less overshooting time than gaze (.18s, p=.003) and head (.22s, $p<.001$) but no difference between gaze and head. For screen width, ($F_{1.55,35.66}$=181.38, $p<.001$, η_p^2=.887), the 40° screen width (.23s) had significantly longer overshooting time than 70° (.13s) and 100° (.13s) screen widths (both $p<.001$). For target width ($F_{1.29,29.56}$=170.32, $p<.001$, η_p^2=.881), larger target widths led to less overshooting time (2°: .24s, 4°: .16s, 8°: .08s, $p<.001$). Similarly for target amplitude ($F_{2,46}$=76.69, $p<.001$, η_p^2=.769), larger amplitude led to less overshooting time (30°: .16s, 60°: .23s, 120°: .11s, $p<.001$). However, targets just outside the FOV (60°) led to higher overshooting as targets occurred earlier in the head shift, requiring an early change of direction. Finally, prior knowledge of target location led to less overshoot time (PK: .08s, NoPK: .25s, $F_{1,23}$=295.52, $p<.001$, η_p^2=.928). These results show the following:

- Overshooting of targets is less prevalent with prior knowledge of the target position, as shown in previous work.
- Increasing the screen width only significantly affects the overshooting time at smaller screen widths. We found significant differences between 40° and 70°, but not between 70° and 100°.
- Targets just outside the FOV had longer overshooting times than targets at further amplitudes as targets appear early in the search movement.
- The controller had significantly less overshooting than gaze and head, while gaze and head had no significant differences.

4.5 Trial Analysis

To understand the nature of selection and overshooting for each modality, we plotted individual selections of each modality (Figure 7). Due to its decoupled nature, the controller rarely overshoots the target, as users can first identify the target and then move the controller in both PK (Figure 7a) and NoPK (Figure 7b) conditions. In case of overshooting, users can quickly adjust, and continuous visual feedback allows participants to anticipate when the pointer will hover over the target.

For gaze pointing, we found that users generally perform large saccades followed by a corrective saccade to point at the target (Figure 7c). Participants commonly overshoot the target without prior knowledge, yet gaze would quickly readjust to hover over the target (Figure 7d). In contrast to controller pointing, users need significant time to visually confirm that they are pointing at the target (Figure 7d). This extra time is likely needed as there is no continuous feedback during pointer movement. The required processing time is independent of target width and may explain why gaze reaches targets significantly faster than the

controller, yet there is no significant difference in movement time. Finally, for the head with PK, users can anticipate target selection just as with the controller through continuous visual feedback to make selections efficiently and accurately (Figure 7e). With no prior knowledge, participants commonly perform overshooting due to its coupled nature with search. In contrast to gaze, the head is significantly slower to readjust the pointing direction, which causes long overshooting times (Figure 7f). However, users are able to preemptively select targets due to visual feedback. These results show:

- Overshooting is rare with the controller, and users can quickly select targets after reaching the target due to visual feedback during pointing.
- Gaze is quick to recover from overshooting. However, participants require a significant amount of time to process the final cursor position before selection, as there is no visual feedback during gaze movement.
- For head pointing, significant time is spent changing the direction of the movement and hovering over the target to recover from the overshooting.

4.6 Model Fitting

We separately analysed the movement time data from the three modalities and the prior knowledge conditions as in previous work for each model [8, 13, 30], and obtained the model fits as presented in Table 1a. Results show that Peephole and Magic Lens's two-stage models fit exceptionally well with prior and without prior knowledge. Meanwhile, Fitts' law and Carpenter's model had comparatively worse fits. These results align with previous results that have shown that the two-stage models are significantly better for predicting the selection of dynamically revealed targets at different screen widths [8, 13, 49]. We show that these models can accurately predict selections with gaze, head and controller pointing in HMD-based VR. We also investigated only the conditions where the target is guaranteed to be outside the FOV (Table 1b). For these conditions, the two-component models show marginally better results for NoPK selections, while selections with prior knowledge show similar or marginally worse performance. Fitts' law and Carpenter's model show similar or marginally worse results.

Further, we were interested in how accurately each model predicts selection performance at individual screen widths. Table 1c shows the model results for all separate screen widths. Peephole and Magic Lens yield exceptionally high R^2 for all modalities and screen widths compared to Carpenter and Fitts' Law. Although it is expected that adding any extra parameter in a regression analysis will always improve the correlation [44], the high correlations imply that both Peephole and Magic Lens are well suited to predicting dynamically revealed selection at both single and multiple screen widths.

For Fitts' Law, we find similar results as previous work that show that Fitts' produces high correlations for selections with prior knowledge at different screen widths [13, 30]. However, as also previously noted, Fitts' law struggles with no prior knowledge of the target location. These results indicate that the two-step selection process is more pronounced when the user has no previous knowledge of the target location and, therefore, better predicted by the two-component models.

For gaze-based pointing, we found very low correlations for Carpenter's formula with prior knowledge but much higher without prior knowledge ($>$.85). This result may be due to the cause of overshooting. With prior knowledge, overshoots are most likely caused by natural overshoots of saccades, and the need for corrective eye movements may then depend on the target width [51]. Without prior knowledge, the user is more likely to traverse the target during search with their gaze, and the user then has to perform corrective saccades when finding the target. As shown previously in Figure 7, it is possible that the participant needs a longer time to visually process that they are pointing at the target and that this processing time is independent of the target width. Our results show that:

- Two-component Peephole and Magic Lens models accurately describe movement time for all modalities.
- Fitts' law accurately describes the movement time for all modalities ($>$.85) at individual screen widths and prior knowledge of the target position.
- Carpenter's model accurately ($>$.85) describes gaze selection without prior knowledge of individual screen widths.

4.7 Head Movement

We were interested to see if the relationship between the pointer and the display impacted performed head movement (i.e display movement). We found no significant 5-way or 4-way interactions. However, we found a significant Modality $\times$ S $\times$ Prior knowledge interaction ($F_{4,92}$=8.17, p<.001, η_p^2=.262). Further analysis showed that participants moved more with the head than the controller and gaze only with NoPK at the 40° screen width (both $p\leq$.006). At 70° and 100° screen widths, the head performed more movement than gaze and controller with PK and NoPK (all $p\leq$.015). Screen width significantly impacted head movement at all levels (all $p\leq$.012). NoPK led to more head movement for all screen widths and modalities (all p<.001).

We also found multiple main effects. For Modality ($F_{2,46}$=11.37, p<.001, η_p^2=.331), head pointing (77.4°) had significantly more head movement than controller (70.3°, p<.001) and gaze (71.7°, p=.007), but we found no significant difference between gaze and controller. For screen width ($F_{2,46}$=441.67, p<.001, η_p^2=.950), smaller screen width led to more head movement (40°: 79.6°, 70°: 70.7°, 100°: 69.0°, all p<.001). Note that the difference is significant but only marginal between 70° and 100° screen widths. For Target width ($F_{1.45,33.39}$=34.89, p<.001, η_p^2=.604) larger targets led to less movement (2°: 74.3°, 4°: 73.6°, 8°: 71.5°, all $p\leq$.007). Regarding target amplitude ($F_{2,46}$=6076.02, p<.001, η_p^2=.996), larger amplitudes led to more movement (30°: 34.9°, 60°: 67.2°, 120°: 117.3°, all p<.001). Finally, participants performed less head movement with PK (69.6°) than NoPK (76.7°) ($F_{1,23}$=192.03, p<.001, η_p^2=.893). These results show:

- Prior knowledge of target position decreases head movement as movement can be planned in advance.
- Gaze and controller pointing do not increase the amount of head movement performed in contrast to head pointing.

4.8 NASA TLX

Friedman test on the overall workload from the Raw NASA TLX questionnaire (Figure 9) showed no significance ($\chi^2(2)$=3.46, p=.177). However, Friedman tests on the results of each sub-scale showed significant differences in Physical demand ($\chi^2(2)$=15.16, p<.001) and Performance ($\chi^2(2)$=7.53, p=.023). Bonferroni corrected Wilcoxon analysis showed that gaze had significantly lower perceived physical demand than both the controller (p=.035) and head (p=.006). Post hoc analysis showed no significant differences for Performance.

5 Discussion

Our results demonstrate how the affordances and differences of gaze, head and controller pointing fit to the specific context of dynamically revealed target selection in HMD-based VR. The main observations of our study can be summarised into two key considerations. First, a decoupled pointer (gaze and controller) is preferable for pointing and selection due to decreased overshooting. Second, the differences between modalities become more apparent at larger screen widths.

5.1 Modalities and Pointing Models

The study results showed significant differences between the three modalities. In contrast to previous Fitts' law-based work, the usually accurate head proved to have more errors, and was also slower than gaze and controller. These results indicate that having the pointer coupled to the display was risk-free when the interaction was performed in front of the user as shown in previous work, but became problematic for the selection of dynamically revealed targets. As highlighted in the amount of overshooting time, with no prior knowledge of the target location, users had to switch head movements to be part of the search

	Peephole								Magic Lens								Fitts' Law						Carpenter					
	Prior Knowl.				No Prior Knowl.				Prior Knowl.				No Prior Knowl.				Prior Knowl.			No Prior Knowl.			Prior Knowl.			No Prior Knowl.		
Modality	R^2	a	b	n	R^2	a	b	n	R^2	a	b	c	R^2	a	b	c	R^2	a	b	R^2	a	b	R^2	a	b	R^2	a	b
Controller	.986	.219	.346	.573	.964	.131	.527	.730	.979	.071	.439	.158	.958	-.009	.617	.152	.849	.296	.184	.664	.280	.214	-	-	-	-	-	-
Gaze	.959	.259	.335	.588	.953	.281	.495	.780	.952	.120	.422	.148	.950	.172	.564	.117	.814	.335	.174	.582	.430	.181	.701	.702	.005	.718	.733	.006
Head	.979	.361	.287	.413	.955	.449	.387	.604	.976	.187	.395	.182	.950	.293	.484	.164	.928	.407	.191	.797	.539	.197	-	-	-	-	-	-

(a) Model fitting results across all screen widths.

	Peephole								Magic Lens								Fitts' Law						Carpenter					
	Prior Knowl.				No Prior Knowl.				Prior Knowl.				No Prior Knowl.				Prior Knowl.			No Prior Knowl.			Prior Knowl.			No Prior Knowl.		
Modality	R^2	a	b	n	R^2	a	b	n	R^2	a	b	c	R^2	a	b	c	R^2	a	b	R^2	a	b	R^2	a	b	R^2	a	b
Controller	.975	.162	.402	.620	.973	.053	.624	.772	.949	-.008	.529	.155	.971	-.130	.746	.151	.827	.155	.218	.675	.040	.267	-	-	-	-	-	-
Gaze	.925	.189	.391	.585	.961	.089	.634	.779	.890	.019	.525	.162	.957	-.085	.753	.147	.808	.183	.222	.658	.076	.268	.616	.690	.006	.748	.585	.008
Head	.954	.300	.339	.447	.955	.134	.608	.678	.914	.093	.494	.190	.938	-.090	.771	.200	.899	.296	.227	.763	.123	.303	-	-	-	-	-	-

(b) Model fitting results across screen widths where conditions with targets appearing within the initial FOV are excluded ($A = 30°$ and $S = 70°, 100°$).

		Peephole								Magic Lens								Fitts' Law						Carpenter					
		Prior Knowl.				No Prior Knowl.				Prior Knowl.				No Prior Knowl.				Prior Knowl.			No Prior Knowl.			Prior Knowl.			No Prior Knowl.		
S	Modality	R^2	a	b	n	R^2	a	b	n	R^2	a	b	c	R^2	a	b	c	R^2	a	b	R^2	a	b	R^2	a	b	R^2	a	b
40	Controller	.987	.271	.322	.499	.966	.074	.603	.757	.990	.101	.409	.182	.973	.093	.683	.170	.931	.283	.213	.770	.109	.294	-	-	-	-	-	-
	Gaze	.948	.413	.267	.440	.933	.170	.612	.799	.956	.252	.348	.170	.941	.021	.679	.147	.910	.422	.188	.708	.208	.281	.659	.859	.005	.895	.708	.009
	Head	.983	.400	.248	.160	.962	.297	.453	.483	.987	.183	.361	.234	.973	.041	.580	.268	.980	.403	.221	.912	.314	.305	-	-	-	-	-	-
70	Controller	.980	.108	.395	.575	.994	.019	.665	.805	.966	-.094	.560	.171	.995	-.152	.792	.136	.920	.094	.225	.787	-.013	.264	-	-	-	-	-	-
	Gaze	.952	.049	.493	.646	.995	.017	.650	.771	.938	-.157	.664	.176	.993	-.172	.795	.154	.865	.030	.255	.820	-.013	.274	.720	.592	.007	.860	.531	.008
	Head	.972	.251	.352	.464	.996	.015	.645	.686	.953	.026	.537	.191	.992	-.238	.843	.208	.942	.241	.230	884	-.011	.230	-	-	-	-	-	-
100	Controller	.995	.121	.482	.698	.976	.080	.706	.823	.991	-.098	.683	.146	.981	-.124	.882	.130	.900	.088	.215	.785	.025	.244	-	-	-	-	-	-
	Gaze	.964	.121	.413	.579	.969	.088	.719	.821	.948	-.129	.658	.171	.967	-.106	.900	.129	.918	.098	.223	.781	.031	.250	.659	.619	.005	.855	.515	.007
	Head	.991	.113	.682	.723	.997	.113	.682	.723	.938	.005	.656	.181	.984	-.165	.947	.188	.932	.244	.230	.882	.006	.290	-	-	-	-	-	-

(c) Model fitting results across separate screen widths.

Table 1: Model fitting results.

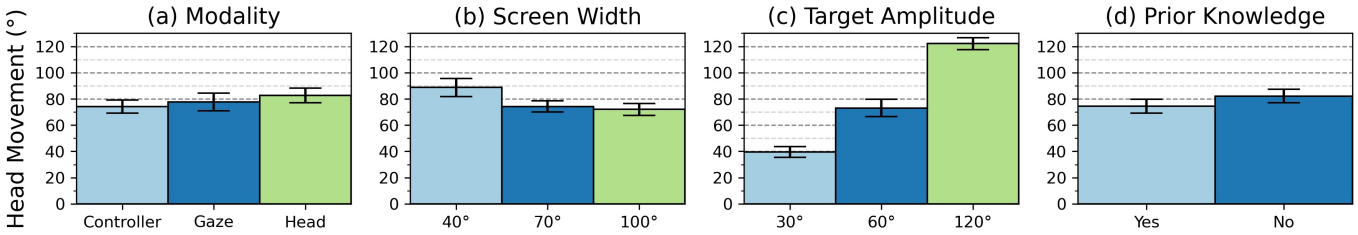

Fig. 8: Head movement main effects. Error bars represent mean 95% confidence interval.

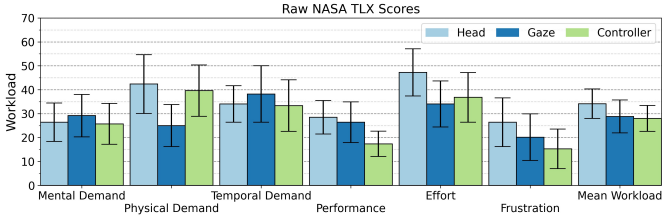

Fig. 9: Average overall workload and NASA TLX sub-scales. Error bars represent mean 95% confidence interval.

and selection. As shown in the high number of errors and longer movement times, this transition appeared to have been a struggle for users. This issue lessened for selection with prior knowledge. Still, the coupling to the display led to longer movement times as users had to slow down their head movement earlier compared to decoupled pointers so that the head pointer would become stationary on the target.

Controller pointing had significantly less overshooting time than both the head and gaze due to the decoupled nature of the pointer. Participants would perform head movements to find the targets, followed by the pointer. Participants had time to adjust their pointing direction to minimise target overshooting and movement time. The small difference between pointing and movement times showed the controller's efficiency, and the gap was primarily due to the time taken to perceive the pointer landing on the target and clicking the button.

For gaze, we found that users first reached the target with their gaze. Similar to head, participants tended to overshoot targets with gaze. However, participants were faster at adjusting their pointing with gaze in comparison with the head, resulting in quicker selection times. In contrast to the head and controller, where participants benefited from continuous visual feedback during movement, significant time was required after adjusting the gaze position to visually confirm that the user was pointing at the target. In addition, gaze had a higher error rate than the controller and a larger difference between reach time and movement time. A possible reason that the controller and gaze reached similar movement times could be that participants were more careful with performing the final click due to eye tracking noise.

In line with previous work on selecting dynamically revealed targets, we found that the two-stage Peephole and Magic Lens models were accurate in predicting performance across multiple screens widths and at individual screen widths [13, 30]. This result was significant as it showed that gaze, head, and controller pointing could be predictably estimated and could be used by designers to decide what modality to use. Further, in line with previous results, we found that Fitts' law presented poor results at various screen widths and only showed good performance with prior knowledge as it did not account for the search element in these types of selections.

For Carpenter's formula, we found high correlations for gaze without prior knowledge at separate screens. A potential explanation for these results may lie in the processing time required before selection with gaze. With prior knowledge, overshoots or undershoots were likely caused by the natural inaccuracy of large saccades. They had to be corrected by following corrective saccades as shown in previous work [51]. These were less likely to occur with larger targets, and the movement time might have been affected by target width. Furthermore, with prior knowledge, less time might have been required for visual processing, which lowered its effect. Without prior knowledge, overshooting occurred due to the user moving past the target with their gaze during the search. This overshooting was independent of the target width. Furthermore, it was possible that more visual processing independent

of target width was required to gain enough information before accurate selection. Further analysis of this aspect of gaze selection would have been an exciting avenue for further research.

5.2 Screen Width

We found that screen width had a significant impact on performance. At the smallest screen width, we observed slight differences between modalities as there was little room for decoupled cursors to operate. However, as the screen width increased, the decoupled gaze and controller achieved significantly higher performance. Interestingly, we observed little performance gain for all modalities between the middle (70°) and wide (100°) screen widths. A possible reason is that the screen width became large enough for participants to react in time, and overshooting time became negligible for overall movement time. A significant insight from these results is that larger screen width was not always better for performance.

As most HMDs did not employ dynamic screen widths, it was important to question the practical usability of knowing performance at varying screen widths. As current HMDs can significantly vary in FOV (e.g. StarVR One has a horizontal FOV of ∼175°, while the HoloLens employs 40°), it would be helpful to know a modality's performance across multiple FOVs, thus making movement time prediction device independent. Furthermore, restricting the FOV has been shown to be an effective tool for combating VR sickness [17] and knowing the impact such restrictors have on performance could be useful for including users prone to VR sickness in performance modelling.

5.3 Prior Knowledge of Target Position

Prior knowledge of the target position also proved to have a significant impact on performance. In line with previous work [8, 13, 30], our results showed that prior knowledge led to faster selections. All modalities were affected by prior knowledge during the search phase. Without prior knowledge, users performed more searching movements, as shown by the significant increase in head rotation and movement time. However, the effect on the pointing phase of the selection varied between modalities. It had the smallest effect on the decoupled controller, where users could search without moving the pointer, only moving it when the target had been located (Figure 7c).

Based on our results, this is a key advantage for the controller when selecting dynamically revealed targets. For the head, we observed that prior knowledge had a significant effect on selections, as users are likely to overshoot without prior knowledge. Possibly due to the head's velocity and mass, adjusting to overshooting is time-consuming and a significant reason why the head is less performant than the controller and gaze (Figure 7f). Finally, gaze quickly recovers from overshooting caused by no prior knowledge of the target location (Figure 7c). Yet, significant time was required to visually process target and gaze positions before selection (Figure 7d). These results could explain why gaze is significantly faster than the controller to reach the target, yet their movement time is not significantly different. Previous work has attributed these results to the common overshooting of gaze [51]. However, our results indicate that visual processing is also a significant factor, as users do not receive visual feedback during gaze movement as they do during controller and head movements. Future work could investigate the effect of visual processing during selection and techniques that could lower the required processing time, which could lead to a significant performance increase for gaze-based pointing and selection.

5.4 Ergonomic Considerations

Several ergonomic issues must be considered when designing for VR [15]. While this study did not mainly focus on ergonomics, our study provided notable results regarding modality differences. Gaze proved to require less perceived physical demand than both the head and controller, according to Raw NASA TLX scores. Expectedly, our head movement results showed that users perform more head movement with the head than with gaze and controller pointing. In addition, gaze does not rely on controller movement in contrast to controller pointing, and we found no significant differences in head movement between gaze and controller. Furthermore, previous results have shown that laser pointers can cause fatigue [43], which can be further exasperated at large selection amplitudes. As such, gaze could be considered a favourable modality to minimise strain.

5.5 Limitations and Future Work

A significant limitation of our work is that it only considers pointing in the one-dimensional horizontal direction. Fitts' law-based procedures are commonly based on two-dimensional movements to capture performance differences in varying pointing directions [37]. Our study methodology is directly based on the original Peephole procedure, which consisted only of one-dimensional pointing [8], and which has been adapted to HMDs [13]. As humans have a bias of searching and moving our heads horizontally [12,57], we believe that our study is still ecologically valid for the search part of the selections. However, our study did not encompass changes in directions that are likely required when the user has found the target. We hypothesise that the decoupled controller and semi-decoupled gaze would perform better than the head in these scenarios. Furthermore, the head also has range of motion and ergonomic limitations in the vertical directions [27, 53]. We believe that the controller and gaze would again be more performant in the event of vertical search. Future work should investigate these scenarios to quantify the differences between the modalities.

Another limitation of our work is the relatively small number of participants. We employed a 5-way RM-ANOVA to cover all study parameters, but only 24 participants were recruited for our study. Our comparable number of data points to previous studies on dynamically revealed targets [8,10,30], and the results aligned with previous studies give us confidence in our results [8, 13, 30, 50]. However, we acknowledge that a higher number of participants would lend more confidence to our statistical results. Future studies should employ a higher participant count to ensure stronger statistical validity.

There are numerous additional opportunities for future research on dynamically revealed target selection in HMD-based VR, some of which arrive from limitations with the present work. For practical reasons, our study contained a relatively low amount of factors within each independent variable. Previous work on selection modelling usually includes more levels to get more data points for model fitting [13]. Finally, our work only considers basic modalities commonly found in VR. VR research includes a plethora of techniques that have improved the performance of each modality and various visualisations to ease selection. Understanding these works in the context of dynamically revealed target selections would be an exciting avenue for future research.

6 Conclusion

As HMD-based VR becomes more prevalent for interaction, it becomes increasingly important to evaluate all aspects of interface design to ensure that these new types of systems are usable and practical for various contexts and tasks. One critical action to support HMD-based VR is efficient pointing and selection.

We carried out a study that explored the selection of dynamically revealed target selection for commonly deployed pointing modalities in HMD-based VR: controller, gaze and controller. We found that controller and gaze pointing provide the fastest selection interaction while the head was the poorest in terms of accuracy and speed, and that an increased screen width can decrease selection time. Additionally, we found that existing models dynamically revealed target selection and accurately described performance.

Our findings can inform interaction designers on choosing interaction modalities that best suit the intended environment and subsequently support users in interacting with environments that leverage the full environment provided by modern AR and VR systems.

Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 101021229, GEMINI: Gaze and Eye Movement in Interaction).

REFERENCES

[1] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013. doi: 10.1016/j.cag.2012.12.003 1, 2

[2] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. Do head-mounted display stereo deficiencies affect 3d pointing tasks in ar and vr? In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 585–592, 2019. doi: 10.1109/VR.2019.8797975 2

[3] A. U. Batmaz and W. Stuerzlinger. Effect of fixed and infinite ray length on distal 3d pointing in virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, p. 1–10. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3334480.3382796 2

[4] J. Blattgerste, P. Renner, and T. Pfeiffer. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3206343.3206349 1, 2

[5] R. A. Bolt. Gaze-orchestrated dynamic windows. In *Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '81, pp. 109–119. ACM, New York, NY, USA, 1981. doi: 10.1145/800224.806796 2

[6] F. Bork, C. Schnelzer, U. Eck, and N. Navab. Towards efficient visual guidance in limited field-of-view head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2983–2992, 2018. doi: 10.1109/TVCG.2018.2868584 3

[7] J. C. Byers, A. C. Bittner Jr., and S. G. Hill. Traditional and raw task load index (tlx) correlations: Are paired comparisons necessary. *Advances in industrial ergonomics and safety*, 1:481–485, 1989. 4

[8] X. Cao, J. J. Li, and R. Balakrishnan. Peephole pointing: Modeling acquisition of dynamically revealed targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 1699–1708. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1357054.1357320 1, 2, 3, 7, 9

[9] R. H. S. Carpenter. *Movements of the eyes, 2nd rev. & enlarged ed.* Pion Limited, London, United Kingdom, 1988. 1, 2

[10] Y. Chen, K. Katsuragawa, and E. Lank. Understanding viewport- and world-based pointing with everyday smart devices in immersive augmented reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376592 2, 5, 9

[11] H. Drewes. *Eye Gaze Tracking for Human Computer Interaction*. PhD thesis, LMU Munich, 2010. 1, 2

[12] W. Einhäuser, F. Schumann, S. Bardins, K. Bartl, G. Böning, E. Schneider, and P. König. Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, 18(3):267–297, jan 2007. doi: 10.1080/09548980701671094 9

[13] B. Ens, D. Ahlström, and P. Irani. Moving ahead with peephole pointing: Modelling object selection with head-worn display field of view limitations. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, SUI '16, p. 107–110. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2983310.2985756 1, 2, 3, 4, 7, 8, 9

[14] B. W. Epps. Comparison of six cursor control devices based on fitts' law models. *Proceedings of the Human Factors Society Annual Meeting*, 30(4):327–331, 1986. doi: 10.1177/154193128603000403 2

[15] J. a. M. Evangelista Belo, A. M. Feit, T. Feuchtner, and K. Grønbæk. Xrgonomics: Facilitating the creation of ergonomic 3d interfaces. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445349 9

[16] 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. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025599 2

[17] A. S. Fernandes and S. K. Feiner. Combating vr sickness through subtle dynamic field-of-view modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 201–210, 2016. doi: 10.1109/3DUI.2016.7460053 9

[18] A. S. Fernandes, T. S. Murdison, and M. J. Proulx. 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*, pp. 1–11, 2023. doi: 10.1109/TVCG.2023.3247058 1, 2, 4, 5

[19] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, 1954. doi: 10.1037/h0055392 1

[20] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan. Direct-touch vs. mouse input for tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, p. 647–656. Association for Computing Machinery, New York, NY, USA, 2007. doi: 10.1145/1240624.1240726 2

[21] J. Gori, O. Rioul, Y. Guiard, and M. Beaudouin-Lafon. The perils of confounding factors: How fitts' law experiments can lead to false conclusions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–10. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173770 1, 2

[22] K. Grinyer and R. J. Teather. Effects of field of view on dynamic out-of-view target search in virtual reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 139–148, 2022. doi: 10.1109/VR51125.2022.00032 2

[23] T. Grossman and R. Balakrishnan. Pointing at trivariate targets in 3d environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, p. 447–454. Association for Computing Machinery, New York, NY, USA, 2004. doi: 10.1145/985692.985749 2

[24] U. Gruenefeld, A. E. Ali, W. Heuten, and S. Boll. Visualizing out-of-view objects in head-mounted augmented reality. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3098279.3122124 3

[25] U. Gruenefeld, D. Ennenga, A. E. Ali, W. Heuten, and S. Boll. Eyesee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 109–118. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132175 3

[26] J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard. A fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3206343.3206344 1, 2

[27] B. J. Hou, J. Newn, L. Sidenmark, A. A. Khan, P. Bækgaard, and H. Gellersen. Classifying head movements to separate head-gaze and head gestures as distinct modes of input. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581201 9

[28] W. Hürst and T. Bilyalov. Dynamic versus static peephole navigation of vr panoramas on handheld devices. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, MUM '10. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1899475.1899500 2

[29] R. J. Jagacinski and D. L. Monk. Fitts' law in two dimensions with hand and head movements movements. *Journal of Motor Behavior*, 17(1):77–95, 1985. doi: 10.1080/00222895.1985.10735338 2

[30] B. Kaufmann and D. Ahlström. Revisiting peephole pointing: A study of target acquisition with a handheld projector. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '12, p. 211–220. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2371574.2371607 1, 2, 3, 5, 7, 8, 9

[31] F. Kerber, A. Krüger, and M. Löchtefeld. Investigating the effectiveness of peephole interaction for smartwatches in a map navigation task. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services*, MobileHCI '14, p. 291–294. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2628363.2628393 2

[32] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International Journal of Human-Computer Studies*, 68(10):603 – 615, 2010. doi: 10.1016/j.ijhcs.2010.05.001 2

[33] 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. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173655 1, 3

[34] J. J. LaViola Jr., E. Kruijff, D. A. Bowman, McMahan, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, USA, 2017. 2

[35] Y.-T. Lin, Y.-C. Liao, S.-Y. Teng, Y.-J. Chung, L. Chan, and B.-Y. Chen. Outside-in: Visualizing out-of-sight regions-of-interest in a 360° video using spatial picture-in-picture previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, p. 255–265. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3126594.3126656 3

[36] P. Lubos, G. Bruder, and F. Steinicke. Analysis of direct selection in head-mounted display environments. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 11–18. IEEE, mar 2014. doi: 10.1109/3DUI.2014.6798834 2

[37] I. S. MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Human–Computer Interaction*, 7(1):91–139, 1992. doi: 10.1207/s15327051hci0701_3 1, 2, 9

[38] P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall/CRC, 1989. 4

[39] S. Mehra, P. Werkhoven, and M. Worring. Navigating on handheld displays: Dynamic versus static peephole navigation. *ACM Trans. Comput.-Hum. Interact.*, 13(4):448–457, Dec. 2006. doi: 10.1145/1188816.1188818 2

[40] M. R. Mine. Virtual environment interaction techniques. Technical report, UNC Chapel Hill CS Dept, 1995. 1, 2

[41] D. Miniotas. Application of fitts' law to eye gaze interaction. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '00, p. 339–340. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/633292.633496 2

[42] A. K. Mutasim, A. U. Batmaz, and W. Stuerzlinger. 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*, ETRA '21 Short Papers. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3448018.3457998 3

[43] M. A. Nacenta, S. Sallam, B. Champoux, S. Subramanian, and C. Gutwin. Perspective cursor: Perspective-based interaction for multi-display environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, p. 289–298. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1124772.1124817 9

[44] R. R. Pagano. *Understanding Statistics in the Behavioral Sciences*. Wadsworth Publishing, 2008. 7

[45] J. Petford, M. A. Nacenta, and C. Gutwin. Pointing all around you: Selection performance of mouse and ray-cast pointing in full-coverage displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3174107 3

[46] I. Poupyrev, S. Weghorst, M. Billinghurst, and T. Ichikawa. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Computer Graphics Forum*, 17, 1998. 2

[47] 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. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132182 2

[48] R. Rädle, H.-C. Jetter, J. Müller, and H. Reiterer. Bigger is not always better: Display size, performance, and task load during peephole map navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 4127–4136. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2557071 2

[49] M. Rohs and A. Oulasvirta. Target acquisition with camera phones when used as magic lenses. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 1409–1418. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1357054.1357275 1, 2, 7

[50] M. Rohs, A. Oulasvirta, and T. Suomalainen. Interaction with magic lenses: Real-world validation of a fitts' law model. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 2725–2728. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1979343 2, 9

[51] I. Schuetz, T. S. Murdison, K. J. MacKenzie, and M. Zannoli. An explanation of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300765 2, 7, 8, 9

[52] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S. Booth. Two-part models capture the impact of gain on pointing performance. *ACM Trans. Comput.-Hum. Interact.*, 19(4), dec 2012. doi: 10.1145/2395131.2395135 2

[53] 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 1, 2, 9

[54] 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. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3332165.3347921 2

[55] 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. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3379155.3391312 3

[56] L. Sidenmark, M. Parent, C.-H. Wu, J. Chan, M. Glueck, D. Wigdor, T. Grossman, and M. Giordano. Weighted pointer: Error-aware gaze-based interaction through fallback modalities. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3585–3595, 2022. doi: 10.1109/TVCG.2022.3203096 2

[57] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein. Saliency in vr: How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1633–1642, 2018. doi: 10.1109/TVCG.2018.2793599 9

[58] V. Tanriverdi and R. J. K. Jacob. Interacting with eye movements in virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, p. 265–272. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/332040.332443 1, 2

[59] R. J. Teather and W. Stuerzlinger. Pointing at 3d targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 87–94, 2011. doi: 10.1109/3DUI.2011.5759222 2

[60] R. Vertegaal. A fitts law comparison of eye tracking and manual input in the selection of visual targets. In *Proceedings of the 10th International Conference on Multimodal Interfaces*, ICMI '08, p. 241–248. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1452392.1452443 2

[61] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. Grønbæk, K. Pfeuffer, and H. Gellersen. 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*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581423 1

[62] C. A. Wingrave and D. A. Bowman. Baseline factors for raycasting selection. In *Proceedings of HCI International*, pp. 61–68. Citeseer, 2005. 2

[63] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963 4