Gaze, Wall, and Racket: Combining Gaze and Hand-Controlled Plane for 3D Selection in Virtual Reality

UTA WAGNER, Aarhus University, Denmark
MATTHIAS ALBRECHT, University of Konstanz, Germany
ANDREAS ASFERG JACOBSEN, Aarhus University, Denmark
HAOPENG WANG, Lancaster University, United Kingdom
HANS GELLERSEN, Lancaster University, United Kingdom and Aarhus University, Denmark
KEN PFEUFFER, Aarhus University, Denmark

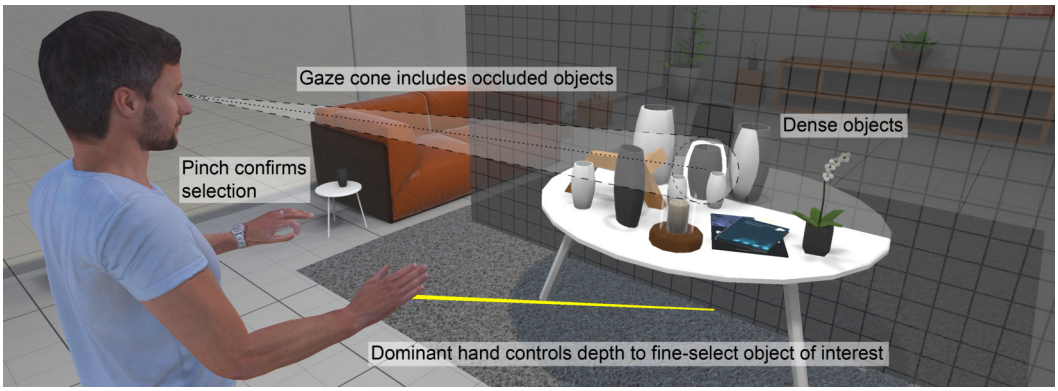


Fig. 1. Example of using Gaze&Wall, a novel interaction technique designed to enable precise object selection in dense environments. Users initially focus their gaze on the desired object. To tackle occlusion, users can indicate their desired depth by pointing with their dominant hand (DH) towards the ground. They finalize the selection with a pinch gesture with the non-dominant hand (NDH).

Raypointing, the status-quo pointing technique for virtual reality, is challenging with many occluded and overlapping objects. In this work, we investigate how eye-tracking input can assist the gestural raypointing in the disambiguation of targets in densely populated scenes. We explore the concept of Gaze+Plane, where the intersection between the user's gaze and a hand-controlled plane facilitates 3D position specification. In particular, two techniques are investigated: Gaze&Wall, which employs an indirect plane positioned in depth using a hand ray, and Gaze&Racket, featuring a hand-held and rotatable plane. In the first experiment, we reveal the speed-error trade-offs between Gaze+Plane techniques. In a second study, we compared the best techniques to newly designed gesture-only techniques, finding that Gaze&Wall is less error-prone and significantly faster. Our research has relevance for spatial interaction, specifically on advanced techniques for complex 3D tasks.

Authors' Contact Information: Uta Wagner, Aarhus University, Aarhus, Denmark, uta.wagner@cs.au.dk; Matthias Albrecht, University of Konstanz, Konstanz, Germany, matthias.albrecht@uni-konstanz.de; Andreas Asferg Jacobsen, Aarhus University, Aarhus, Denmark, 201905139@post.au.dk; Haopeng Wang, Lancaster University, Lancaster, United Kingdom, h.wang73@lancaster.ac.uk; Hans Gellersen, Lancaster University, Lancaster, United Kingdom and Aarhus University, Aarhus, Denmark, hwg@cs.au.dk; Ken Pfeuffer, Aarhus University, Aarhus, Denmark, ken@cs.au.dk.

ACM 2573-0142/2024/12-ART534
https://doi.org/10.1145/3698134

CCS Concepts: • **Human-centered computing → Mixed / augmented reality**; **Pointing**.

Additional Key Words and Phrases: eye-tracking, gaze interaction, occlusion, object selection, disambiguation, complex 3D tasks

ACM Reference Format:
Uta Wagner, Matthias Albrecht, Andreas Asferg Jacobsen, Haopeng Wang, Hans Gellersen, and Ken Pfeuffer. 2024. Gaze, Wall, and Racket: Combining Gaze and Hand-Controlled Plane for 3D Selection in Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 8, ISS, Article 534 (December 2024), 25 pages. https://doi.org/10.1145/3698134

1 Introduction

Interaction in crowded virtual environments can become overwhelming, especially in applications like immersive analytics [15], biomedical visualization [43], molecular modelling [17], and 3D design [17], where numerous virtual objects closely coexist, potentially causing occlusions and overlaps. One of the major challenges herein lies in the specification of a position in all three dimensions to select any interactive element within the visualisation.

The 3D interface standard, raypointing, struggles to handle densely populated environments. Pointing via the bare hand, a hand-held controller, or by eye-gaze enables object selection within a ray's range and those it intersects, but it relies on extra user input to specify depth and circumvent occlusions [2]. This challenge amplifies when dealing with interface systems without multi-purpose controllers, instead based on freehand gestures [38]. Gestures are natural for 3D interaction but lack precision and expressiveness compared to controllers equipped with a large input vocabulary and physical delimiters. Yet, if barehanded communication represents the future of spatial computing (e.g., Microsoft Hololens 2, Meta Quest 3, Apple Vision Pro), it demands explorations across the spectrum of 3D manipulations – and in particular for hard occlusion tasks that have been reserved to controllers so far. Then, visual analysts can navigate 3D visualizations in mobile contexts without immediate controller access, facilitating seamless interaction with a single head-worn device.

Interaction techniques can be classified into *separation* of the task into lower-dimensional transformations, and *integral* task mappings that merge multiple inputs to a multiple degrees-of-freedom interaction [22]. Consider the RayCursor technique [4], which utilizes a controller's touchpad to establish the ray's depth. Since the same hand manages both raypointing and depth control, users tend to complete the task one after the other. Conversely, Conductor [52], for instance, employs a dual-pointing approach, where a ray and a plane's intersection define a 3D point. The technique allocates the ray and plane control sub-tasks to both hands for parallel engagement. This affords integral 3D specification, and can result in users being faster and more efficient to perform selections in densely populated scenes. Despite the performance benefits, two-handed pointing can raise complexity, as of the increased coordination needs and physical effort over time. Exploration of alternatives that can lower the cognitive, temporal, and manual demands on such a complex task is important to render 3D operations easier to use.

VR head-mounted devices with integrated eye-tracking are valuable for gestural interfaces, as gaze input lowers the physical and temporal demands associated with gestures [44]. State-of-the-art HMDs utilize multimodal eye and hand tracking using a Gaze + Pinch interaction model [31], which simplifies hand engagement to a pinch gesture for confirming gaze selections. Here, we investigate the incorporation of the gaze modality into the more complex 3D specification tasks, with the aim to lower the effort and enhance efficiency. The dual-pointing approach can in principle be adopted to gaze and single-handed rays, but it changes the nature of the interaction. Eye movements are fast and effortless, often preceding the acquisition of the target before manually engaging in the task [41]. Using gaze as the primary ray and one hand for plane control affords interactions on par with

two-handed techniques – with the effort of one hand. However, as our eyes move fundamentally different to our hands, the challenge lies in fine-tuning the nuances of the multimodal inputs.

In this paper, we investigate Gaze+Plane-based techniques for 3D position specification in VR. Gaze+Plane denotes a principle where the natural gaze direction of the user is employed as the primary ray, coupled with a hand-controlled plane, to specify a 3D point when employed in concert. Building upon the Gaze+Plane principle, we have devised two specific techniques, each characterized by a unique hand-controlled plane design:

- Gaze&Wall: The gaze ray intersects with a wall, in the form of a 2D surface that is always perpendicular to the user. Its depth is manipulated by pointing with the hand at the ground where the wall is to be located. This technique dedicates the plane fully to complement the gaze ray with depth specification.
- Gaze&Racket: Gaze intersects with a plane whose control is based on a racket metaphor. Inspired by Conductor [52], the user holds a virtual plane in their hand, granting them the ability to flexibly orient it and 'slice' through the 3D scene. This technique affords the flexibility to adjust the point of intersection by using any potential plane transformation possible by hand.

In Figure 1, we demonstrate an example of Gaze&Wall interaction for a virtual modelling scenario. A user facing a 3D model with a variety of objects, and wants to select a specific object of interest. They start by looking at the target, but it's obscured by other objects. To resolve this, the user engages their dominant hand to point at the ground, indicating the depth of the desired target. This translates the Wall to the corresponding depth, and selects a single object. A pinch with the non-dominant hand finalizes the selection.

As prior work focused on controller-based techniques, we designed the concept *Ray +Plane* which combines Ray&Wall and Ray&Racket as two gestural techniques to which we compare the Gaze+Plane techniques. Here the dominant hand is assigned to point with the ray, and the non-dominant hand controls the plane, as inspired by the Conductor concept [52]. By dual-pointing with both hands, the user can specify a 3D point by intersecting the ray and plane. Selection confirmation is performed by a pinch gesture of the non-dominant hand.

Across two user studies, we investigate the efficiency of the techniques for object selection in densely populated virtual environments. The first study aimed to get a first understanding of the various design parameters of the Gaze+Plane techniques, such as the visual design of the planes (Racket vs Wall), and the order of how the eye and hand modalities can be employed by the user. The order defines how our system provides visual feedback, which can affect the performance. In gaze-first mode, the UI pre-selects and highlights nearby targets along the gaze. In plane-first mode, targets on the plane are highlighted. From this study, we obtained multiple insights into our final designs, improved the visual perception of the planes, selected a "gaze first, plane second" principle, and addressed selection slip errors. Given this, we conducted our main experimental comparison of two optimized gaze-based techniques to two hand-based interaction techniques. The study results offer extensive insights into temporal, spatial, and task-load aspects. Notably, Gaze&Wall emerged as the top-performing technique, delivering substantial time and error reductions compared to all other methods. Additionally, both Wall-based techniques resulted in reduced physical motion involving the head and both hands.

In summary, the contributions of this paper include the following points. First, the design and implementation of Gaze&Wall and Gaze&Racket, as two multimodal eye-hand techniques for 3D occlusion selection, for the benefit of eliminating an entire pointing sub-operation from the 3D specification task by integrating eye-tracking. Second, a first experiment that revealed strengths and limitations with respect to the order of eye/hand modalities, slipping errors, and plane design, useful and used to design more advanced interaction techniques. Third, a second experiment that

compares two optimised techniques from the first study to two newly developed gesture-only interaction techniques adopted from the prior art; revealing Gaze&Wall is the currently best performing technique for 3D specification in eye- and hand-tracked VR systems.

2 Related Work

Our work is at the intersection between the fields of occlusion-management techniques designed for hand-controlled UIs, and gaze-based human-computer interaction.

2.1 Hand-controlled Occlusion Management

Occlusion, where objects are hidden from view, is a common occurrence in spatial object selection tasks [2]. Navigating for a clear view is a workaround, but in other cases, it can be inconvenient for frequent selections over a prolonged period of time. One can incorporate multiple viewports, x-ray and cutaway views for a see-through effect [5], and transformations of objects and scenes for easier reach [3, 13, 14]. Small improvements can be gained through ray position and direction refinement via ray optimization, non-linear input mappings, and filtering methods [4, 16, 49].

A major class of occlusion management are interaction techniques that offer users explicit UI commands to cover the input range of degrees of freedom (DOF) for 3D specification. These can be framed relative to Jacob's theory on integrality and separability of the task structure [22]. DOF separation breaks down sub-tasks into sequential steps, allowing to precisely operate one dimension at a time, while DOF integration enables simultaneous control across multiple degrees of freedom for fast task completion. For instance, Depth Ray and Lock Ray allow for disambiguation when multiple targets are within the initial ray, using a cursor on the ray manipulated, e.g., by moving the controller forward or backwards in mid-air [18]. RayCursor extends the concept by employing the touchpad of a controller to specify cursor depth [4]. Upon activating the selection trigger, the object intersected by the ray and closest to the 3D cursor is selected. These examples indicate DOF separation, as it is difficult to simultaneously employ one hand for both raypointing and cursor dragging.

Only a little work exists on using hand gestures for occlusion management. While simple gestural input is intuitive to perform, it does not afford the complexity of controller-based input that features combinations of buttons and joystick movements, such as the selection techniques described in [51], making it more challenging to design for occlusion management. Recent efforts by Delamare et al. [12] proposed the MultiFingerBubble as a gestural 3D Bubble cursor. The technique enhances hand pointing with a disambiguation mode where each finger is assigned to a target in the pointing area, allowing for the fine selection of up to 5 targets. Shi et al. [38] investigated occlusion techniques for freehand gestures, proposing extra modes for raypointing. HandDepthCursor uses a "back" and "forward" gesture of the non-dominant hand. The second technique, HandConeGrid, moves all candidates of a ray-based area selection of the non-dominant hand to a 2D grid when pinching the dominant hand, allowing for easier access. In their evaluation, they showed their techniques outperform the MultiFingerBubble with regard to user performance and effort. Our interaction techniques are distinct, as we focus on a different class of techniques that do not separate the task into several interaction steps using special gestures or different target views.

In contrast, techniques based on DOF integration are often realised through bimanual interfaces, specifically helpful in compound selection/positioning tasks that afford strategies to use both hands simultaneously [8]. For instance, the iSith technique [48] employs a dual-pointing approach, with two hands controlling two rays, and the 3D point is determined by the shortest distance between these rays. This computed 3D point is subsequently used for proximity testing against the scene as an object-snapping mechanism to the nearest target. As well, Conductor [52] is an intersection-based technique that expands the concept to *Ray + Plane* held in the hands, allowing

for an easier way to specify depth. In their user study, the users were more efficient using this technique in comparison to RayCursor, showing the benefits of the integral technique design. Lastly, GazeRayCursor [10] extends RayCursor by using the intersection between controller and gaze ray for selection, leading to superior performance than RayCursor. Our work shares the use of eyes and hands to accomplish occlusion tasks, but is novel in the investigation of 1) gestures with distinct UI challenges, 2) two-handed techniques to balance the division of labour across hands, and 3) novel Wall and Racket-based techniques.

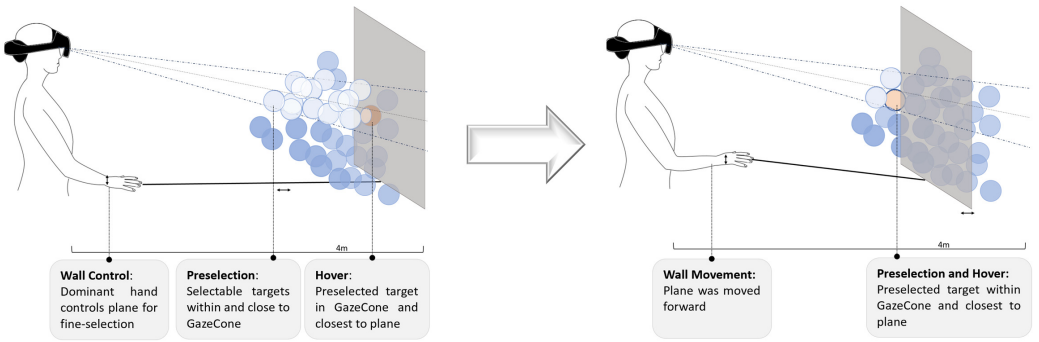
2.2 Gaze-based Human-Computer Interaction

There are several advantages to engaging the eyes in VR – Tanriverdi and Jacob [42] for instance list natural interaction, high interactivity, fast speed, and robust eye-tracking – and research in this space is rapidly growing in recent time [36]. There are many new options to consider in the design space for eye-tracking as input medium in VR and AR [20, 29], and recent developments have spawned innovative ways that address the Midas-Touch problem [21] and provide new interface systems, such as via smart coupling between eye and head inputs [41], expressive menu interfaces [1, 33], information retrieval in AR [25, 34] and depth-based inference of objects [20, 39].

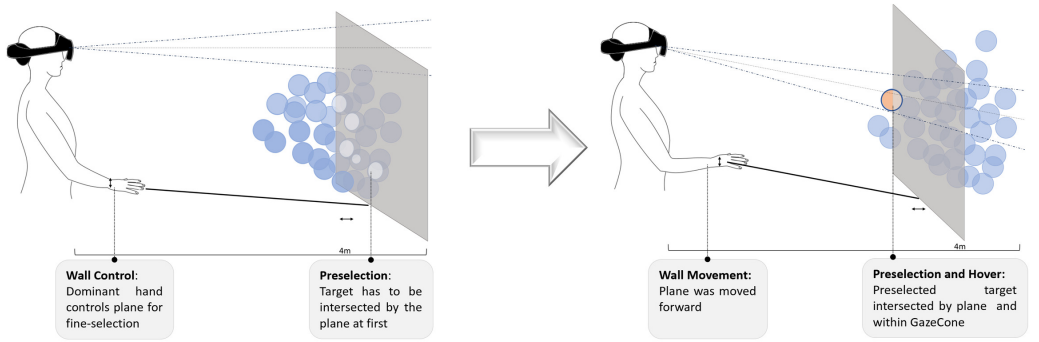
A related line of research is the study of interfaces that combine eye-tracking with hand-controlled input in a multimodal way. This is of relevance as, for example, the Gaze + Pinch-based interaction technique as introduced by Pfeuffer et al. [31] has been demonstrated to be more efficient than gesture-only interactions in selection and menu tasks [26, 44] and is becoming widely available for virtual reality consumers (e.g., Microsoft Holo Lens 2, Apple Vision Pro). The default model is focused on raypointing-based gaze, inheriting the same limitations as manual raypointing for occluded virtual scenes. Several related disambiguation methods have been investigated for eye-tracking. For instance, Outline Pursuits [40] allows for disambiguating between raypointing-selected targets by utilising distinct smooth pursuit eye movement patterns for each target. Vergence Selection [39] analyses the correlation between eye and target depth motion to select targets precisely. Several researchers have explored how eye-tracking can be used to interact with menus, which involves a similar dual-task structure of pointing at the menu and then opening a menu item. There are many ways based on DOF separation to merge eye and hand inputs, such as to switch from raypointer to eye-based input to access individual menu elements [26, 34, 35]. Conversely, to use gaze pointing to activate the menu and hand input for item control [23, 32, 37]. For instance Shi et al. [37] have explored eye-hand techniques to accomplish the task of region selection in 3D, including two eye-hand techniques based on Gaze+Pinch and Gaze-Hand Alignment where gaze points at the initial point and a hand gesture defines a rectangle. Yu et al. [50] assessed object manipulation techniques for interior design via controller and gaze. Contrary to our present study, the studies in the two papers did not reveal the benefits of multimodal techniques over manual baselines, indicating that the efficiency depends highly on the task at hand, necessitating a focused investigation for the underexplored occlusion task.

3 Gaze + Plane Interaction Techniques

Our Gaze+Plane proposal represents a specialisation of the "Ray + Plane" intersection-based 3D specification. The eyes take over the first sub-task of ray pointing. As the eyes naturally point to targets of interest, this can eliminate a whole sub-task from the overall interaction. In this context, we propose two plane metaphors to combine with the gaze ray: *Wall* and *Racket*. Both plane types implement the 1€ Filter [9] for the hand position and hand direction to stabilize the underlying hand tracking data, ensuring smoother and more reliable interactions (Wall: $f_{c_{min}} = 0.01, \beta = 1$, Racket: $f_{c_{min}} = 0.01, \beta = 50$). We also used a 1€ Filter ($f_{c_{min}} = 0.5, \beta = 50$) to smooth the gaze direction and prevent gaze jittering. The filter for gaze differs from the hand filter because the



(a) Gaze&Wall



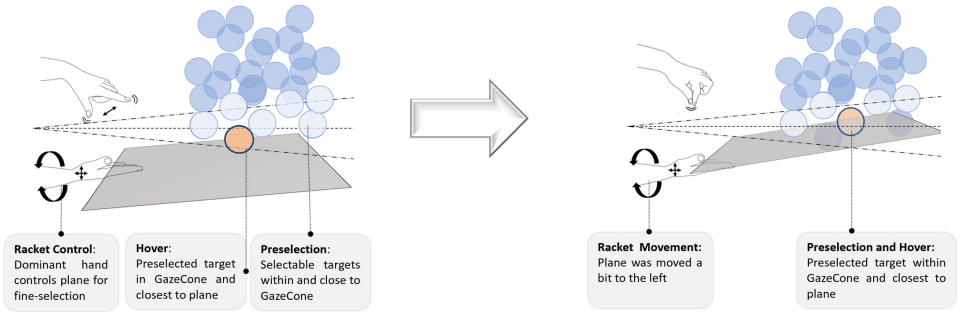
(b) Wall&Gaze

Fig. 2. Shows both modalities with Wall Metaphor based on the Gaze+Plane principle and outlines two stages for each: gazing and closest to the plane (vice versa for Wall&Gaze). The Wall Metaphor enables the user to explicitly define depth.

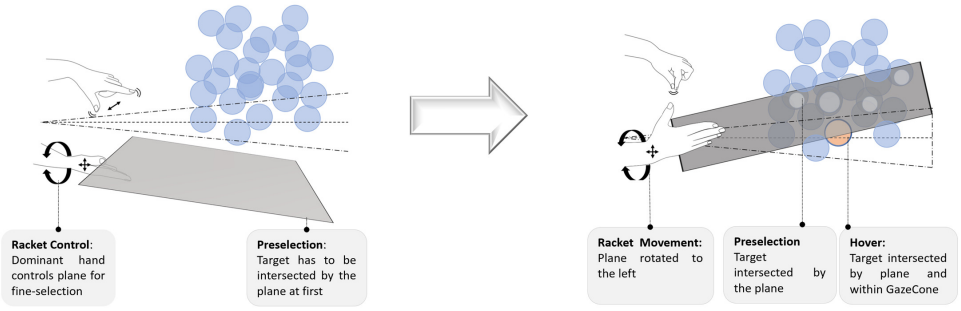
dynamics of the two modalities are different. The gaze filter was iteratively adjusted through pilot experiments. . For both planes we decided to design them in gray with a transparency of 82.6%, allowing users to see through it.

3.1 Gaze&Wall (Figure 2a, Figure 5a)

The first technique employs eye-gaze in unity with a *Wall* metaphor, where a virtual plane is displayed within the 3D virtual scene. The primary purpose of the wall metaphor is to represent depth in conjunction with raypointing. While raypointing is well-suited for specifying the X and Y directions, it is limited in defining depth (Z). The Wall allows users to define depth explicitly, creating a clear separation of concerns. The wall is straight, with its centre oriented towards the user. The transparency serves as visual feedback, indicating which targets are in front of or behind the wall. Interaction involves using the wall to set the depth level of selection and the ray to specify a point within that depth. We have experimented with various methods to position the wall, including



(a) Gaze&Racket.



(b) Racket&Gaze.

Fig. 3. Shows both modalities with Racket Metaphor based on the Gaze+Plane principle and outlines two stages for each: gazing and closest to the plane (vice versa for Racket&Gaze). This Racket Metaphor is inspired by Conductor[52] and provides a flexible way to intersect in dense 3D environments as the plane is aligned to the hand.

using a ray to point at a desired location on the virtual scene's ground floor. This approach provides an intuitive experience similar to teleportation or other raypointing-based features, and the wall's large visual feedback is ambient, visible across almost the entire field of view. We have set the size of the Wall to a width of 6 meters and a height of 3.6 meters, ensuring that it generously covers the target space in our experiments. For ergonomic reasons, the ground-pointing ray originates from the user's wrist and points 45 degrees downward toward the middle finger and 15 degrees toward the thumb. To reduce the hand jittering, we applied a 1€ Filter with $f_{c_{min}} = 0.01$ and $\beta = 1$ to both wrist position and direction. We applied a linear interpolation to the plane position between consecutive frames by $5 * Time.deltaTime$ in corporate with the issue of varied angular motor size with ray-cast pointing on the plane [28] and thus made Wall able to provide an accurate depth support.

3.2 Gaze&Racket (Figure 3a, Figure 5c)

We designed a *Racket* Metaphor for plane control to assist the gaze-based raypointing. Inspired by the controller-based technique Conductor [52], this approach involves users moving and rotating their dominant hand to control the orientation and direction of a handheld plane (referred to as the "racket"). Unlike the Wall metaphor, the plane is not necessarily orthogonal to the user's view, and it is not primarily associated with depth. Instead, the racket provides a flexible way for users to intersect the 3D space in unique ways to facilitate target selection, especially in highly populated environments potentially enabling quick and precise target selection.. We also see an advantage in the fact that lateral (X and Y) and depth (Z) movements are possible simultaneously making the interaction feel more natural and intuitive for the user. We have set the size of the Racket to 10.15 x 8 meters, ensuring that it generously covers the target space in our experiments.

3.3 Design Considerations

3.3.1 One vs. Two-handed Interaction. One-handed interaction is the common standard, but two-handed interfaces can come in handy for integral manipulation of higher-dimensional tasks [8, 21].

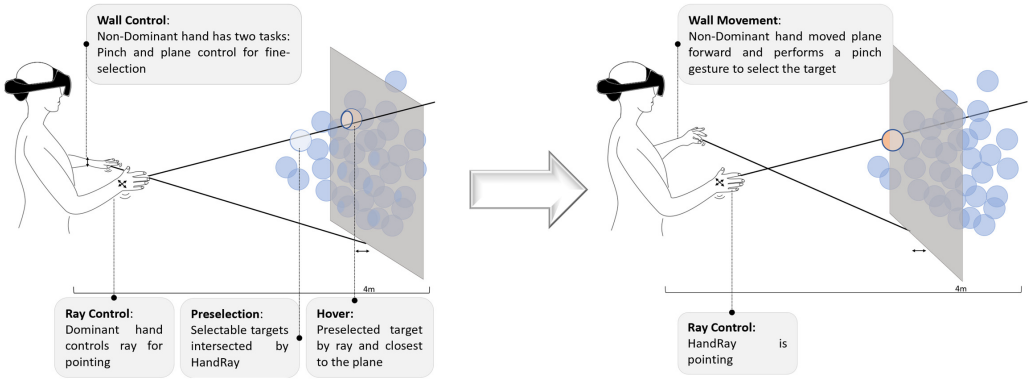
In a potential unimanual mode, both plane control and confirmation gestures are performed with the same hand, offering the advantage of requiring only one hand for interaction. However, simultaneous execution of pinch and pointing can impair accuracy due to the Heisenberg effect [6] as there's a slight shift in raypointing for the plane. Integrating eye gaze into two-handed tasks can substantially reduce physical effort and render tasks one-handed[30]. In principle, it's possible to design a one-handed technique where the eyes specifies the ray, and the hand performs both plane control and confirmation gestures. However, using both hands is preferred.

This approach aligns with the Conductor concept, but with a twist by allocating plane control to the dominant hand and relying on pinch-based confirmation with the non-dominant hand. This choice is rooted in the understanding that the dominant hand typically possesses superior motor control precision [19].

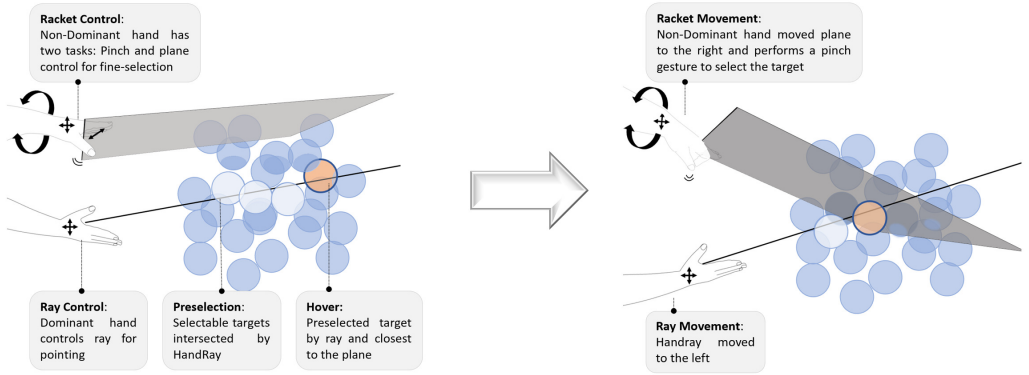
3.3.2 Modality Order. For both Racket and Wall variations, there is a fundamental choice in the temporal division of labour between the eyes and hands. Which modality should have priority in defining the initial target candidates, and which modality is suited for the disambiguation phase? This is important, as the UI highlights the candidates through visual feedback. The right visual feedback can have a potential effect on performance that we aim to investigate. In this context, two modes are possible:

- In a *Plane-First Mode* (Figure 2b, Figure 3b), the users move the plane for a coarse target preselection. Targets that are within 5 cm of the plane are preselected. The target on the plane closest to the gaze ray is selected. When there is only one target on the plane, the target is selected without the need for employing gaze. We indicate this modality order with the *second* word Gaze in the name of the technique: Wall&Gaze and Racket&Gaze.
- In a *Gaze-First Mode* (Figure 2a, Figure 3a), the users look at the target for a coarse preselection. Due to the imprecise eye-tracking, eye gaze can not always pinpoint the target. Hence, all targets within a 1.5-degree radius around the gaze ray are preselected. The target within the gaze cone closest to the plane is selected. When there is only one target within the gaze cone, the target is selected without the need for the plane. We indicate this modality order with the *first* word Gaze in the name of the technique: Gaze&Wall and Gaze&Racket.

3.3.3 Visual Feedback. Visual feedback is important to indicate which targets are within preselection, and in the next stage to indicate which target will be finally selected among the preselections. We experimented with different visual feedback mechanisms and found that highlighting



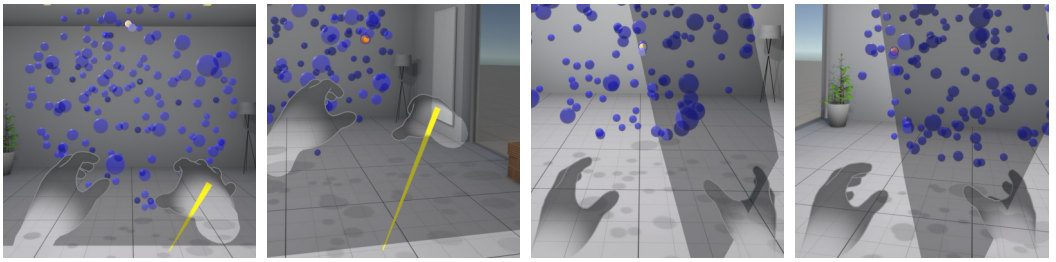
(a) Ray&Wall



(b) Ray&Racket

Fig. 4. Shows both modalities with (a) Wall and (b) Racket Metaphor based on the Ray + Plane principle. These are the two hand-based techniques to the Gaze+Plane techniques. Each is shown in two stages: pointing and closest to plane selection.

all pre-selected candidates can be distracting. Feedback for all pre-selections in the *plane-first mode* faced considerable visual noise due to the dynamic highlighting of objects intersected by the hand-controlled plane. As a result, we opted not to use feedback. Instead, the natural physical intersection of the plane to transparent objects can be used. In the *gaze-first* mode, there was no natural physical intersection, requiring feedback. This was acceptable since there were only a few pre-selected targets within the gaze cone, minimizing the impact on visual flickering. As a result, with *gaze-first* all targets pre-selected within the gaze cone become slightly transparent and light blue. For both *plane-first* and *gaze-first* orders, a blue outline around a target indicates that the target is within the gaze cone, closest to the plane, and thus can be selected. Only one target can be selected and have the blue outline at a time.



(a) Gaze&Wall (b) Wall&Gaze (c) Gaze&Racket (d) Racket&Gaze

Fig. 5. Overview of all four techniques used in study 1 from a user perspective. (a) Gaze&Wall: selects the target within the GazeCone and closest to the plane. (b) Wall&Gaze: selects the target intersected by the plane and closest to the gaze ray. (c) Gaze&Racket: select the target within the GazeCone and closest to the plane. (d) Racket&Gaze: selects the target intersected by the plane and closest to the gaze ray

4 User Study 1: Modality order and Plane Type

The first study investigates *(RQ1) How do the two plane metaphors (Wall, Racket) affect the user's performance?* to determine on which plane users are more accurate and faster and which is also less physically demanding for depth movement. and *(RQ2) How does modality order (gaze first, plane first) affect the user's performance?* as it is intersting to find out with which modality order users are more accurate and faster, and which order is more intuitive and natural for them. The study used a within-participant design with two independent variables. We utilized a 4×2 design of the variables *TECHNIQUE* and *DENSITY*. The order of conditions (combination of *TECHNIQUE* and *DENSITY*) was counter-balanced using a balanced Latin square. We compare the following interaction techniques: Wall&Gaze, Gaze&Wall, Racket&Gaze, and Gaze&Racket (Figure 5). Density describes the object population within the cone area. We test each technique with the following densities: D_{Low}=75, D_{High}=150 targets.

We chose two item densities for more variety in the task and according to [52],[26], [4], [40], [45]. The low density represents a task with little occlusion, and the high density was chosen to experience greater target occlusion. The study design for study 1 was: 12 *participants* $\times$ 4 *techniques* $\times$ 2 *densities* $\times$ 30 *trials* = 2880 data points in total.

We recruited 12 participants (4 female and 8 male), aged $22 - 37$ ($M = 27.00$, $SD = 4.55$). 10 were right-handed, 2 were left-handed, 2 wore glasses, and 1 wore contact lenses. Participants rated their knowledge on a scale of 1 to 5, with ratings indicating low to moderate experience in XR/VR/AR ($M = 3.00$, $SD = 1.41$), 3D hand gestures ($M = 2.17$, $SD = 1.19$), and eye-hand interaction ($M = 2.08$, $SD = 0.90$).

The techniques and study software were implemented using the OVR toolkit in Unity3D (Version 2020.3.43f1) on the Meta Quest Pro (106°x95.57° field of view, 1800x1920 pixels per eye), which supports hand and eye tracking. The accuracy of the eye tracking on Meta Quest Pro is around 1.652° [45]. The study participants were standing in a large, quiet room during the study.

4.1 Task

We adopt an object-selection task in VR similar to closely related work [4, 52], where users are tasked to select one target among many distractor targets. As all techniques are gaze-based, and we aim to assess an occlusion task, we place targets within a cone that begins in front of the user and stretches away from the user. Targets in the form of spheres were randomly placed within the

cone shape with the apex at the user's eye level. The cone had a length of 4 meters and a radius of 0.5 meters at a distance of 1 meter from the apex. Targets could only appear between 1 to 4 meters away from the apex and had a minimum required distance of 13 cm between their center positions to prevent overlapping targets. For each trial, a fixed amount of targets, depending on the density, are pseudo-randomly spawned according to the described placement constraints within the cone. A single target (diameter 8 cm) is randomly selected as the target of interest in opaque orange that needs to be selected. Opaque light orange communicates that the target of interest is currently within the gaze cone (only for the gaze-first techniques).

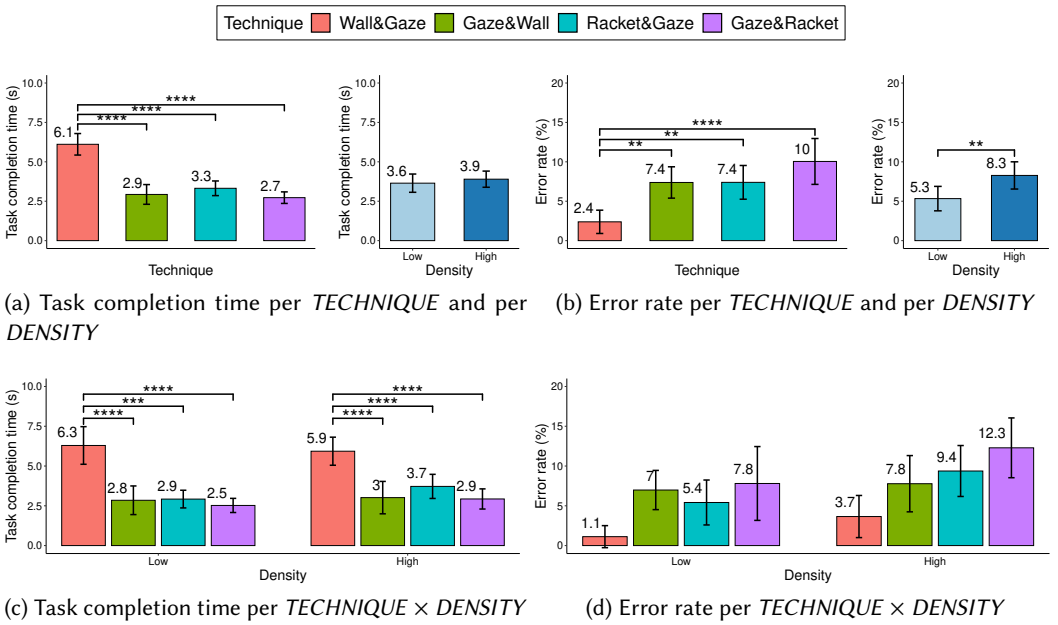
4.2 Procedure

Initially, participants were introduced to the study and asked to complete consent and demographic forms. Furthermore, the functioning of each interaction technique was explained before each run. Calibration of the eye tracker took place before the start of each new technique. Following this, a 1-minute training run with only 10 trials was conducted. This served the purpose of familiarizing users with the technique and preventing any learning effects. Once all instructions were provided to the participants, the study was started. At the end of each technique, a questionnaire was completed, allowing participants a 2-3 minute break between the techniques. After testing all four techniques, participants ranked them. The entire duration of the study was approximately 50 minutes.

We measured *Task Completion Time (TCT)*, as the time taken to successfully complete a task from when the target appeared until the trial was finished via pinch. Second, *Error Rate (ER)* is the number of trials in which the correct target was not chosen or was not selected within 30 seconds, divided by the total number of trials per condition. Third, *Hand Movement (HM)*, as the cumulative difference in palm position between frames, can be used to determine how much the user had to move their hand, which can correlate with physical fatigue [50]. We divide the total length by the TCT to normalize different trial lengths. We used the *NASA TLX questionnaire* (Task Load Index) [11] to measure the subjective task load experienced by the study participants on a 7-point scale, followed by open-ended questions on eye/hand fatigue. A ranking of techniques was filled out at the end of the study. Additionally, we have offered users the chance to provide feedback on the technique through a comment field. We categorized the open-ended questions and analyzed them according to the frequency of responses. The questionnaire is divided into six sub-scales: (1) Mental Demand, (2) Physical Demand, (3) Temporal Demand, (4) Performance, (5) Effort, and (6) Frustration.

4.3 Results

We first removed all timeouts from the data, which refers to cases where no target was confirmed within 30 seconds. These timeouts were due to temporary poor performance of the headset, making it impossible for users to select the target, for example, when hand tracking didn't work. In total, we eliminated 25 timeouts, comprising 0.9% of the trials (8 for Wall&Gaze, 2 for Gaze&Wall, 12 for Racket&Gaze, 3 for Gaze&Racket). Secondly, for the analysis of task completion time, a total of 70 outliers were excluded if a trial time exceeded the $Mean + 3 \times SD$ threshold. Concerning task completion times, a total of 2.4% of trials were removed (13 for Wall&Gaze, 18 for Gaze&Wall, 21 for Racket&Gaze, 18 for Gaze&Racket). Subsequently, we have tested normality tests on the quantitative variables and applied data transformations (ART [46] and Box-Cox [7]) when dealing with factors that exhibited non-normal distributions. We conducted a repeated measures ANOVA for the quantitative data (with Greenhouse-Geisser corrections if sphericity was violated), followed by estimated marginal means post-hoc pairwise comparisons with Bonferroni corrections. In the case of the non-normal distributed data of the questionnaires, we used a Friedman test with



(a) Task completion time per *TECHNIQUE* and per *DENSITY*

(b) Error rate per *TECHNIQUE* and per *DENSITY*

(c) Task completion time per *TECHNIQUE* × *DENSITY*

(d) Error rate per *TECHNIQUE* × *DENSITY*

Fig. 6. Mean task completion time and error rate of study 1, including significant post-hoc tests. Error bars denote 95% confidence intervals.

post-hoc Conover tests (Bonferroni corrected). In all the following figures, we denote statistical significance with * for p < .05, ** for p < .01, *** for p < .001, and **** for p < .0001.

4.3.1 User Performance and Hand Movement (Figure 6, Figure 7). Regarding TCT (F^3_{33} = 30.61, p<.0001, η^2_G = 0.475), we find that users were significantly slower with Wall&Gaze (6.1s) than with the other three techniques (p<.0001, Gaze&Wall: 2.9s, Racket&Gaze: 3.3s, Gaze&Racket: 2.7s). In case of a *TECHNIQUE* × *DENSITY* interaction effect (F^3_{33} = 4.07, p=.015, η^2_G = 0.037), users were significantly slower (p<.001) with Wall&Gaze for D_{Low} (6.3s) than with all the other techniques (p<.001, Gaze&Wall: 2.8s, Racket&Gaze: 2.9s, Gaze&Racket: 2.5s). We observe that Wall&Gaze is also slower for D_{High} (5.9s) compared to the other techniques (p<.0001, Gaze&Wall: 3s, Racket&Gaze: 3.7s, Gaze&Racket: 2.9s) With regards to ER (F^3_{33} = 9.33, p<.001, η^2_G = 0.280), ER was significantly lower with Wall&Gaze (2.4%) compared to the other three techniques (Gaze&Wall: 7.4%, p=.007, Racket&Gaze: 7.4%, p=.004, Gaze&Racket: 10%, p<.0001). As well, users exhibited significantly higher error rates with the gaze-first mode techniques (p<.001). A main effect for density (F^1_{11} = 16.29, p<.001, η^2_G = 0.106) confirmed the expected higher ER for D_{High} (8.3%) compared to D_{Low} (5.3%, p=.002). Analysis of hand movement ($F^{1.8}_{20.1}$ = 6.19, p=.009, η^2_G = 0.156) showed significantly less hand movement with Gaze&Wall (0.021) compared to two other techniques (Racket&Gaze: 0.034, p<.001, Gaze&Racket: 0.033, p=0.043).

4.3.2 Usability Questionnaire (Figure 8, Figure 9). Both gaze-first mode techniques were most preferred (Gaze&Wall: 50%, Gaze&Racket: 50%). Regarding user ratings, we find significant effects for physical demand ($\chi^2(3)$ = 8.22, p=.042, W = 0.228), effort ($\chi^2(3)$ = 12.27, p=.007, W = 0.341), and hand fatigue ($\chi^2(3)$ = 15.03, p=.002, W = 0.418). Wall&Gaze was perceived as more physical demanding (Mdn = 5) than Gaze&Racket (Mdn = 2.5, p=.009). Participants felt they had to put in more effort with Wall&Gaze (Mdn = 5) to be successful compared to

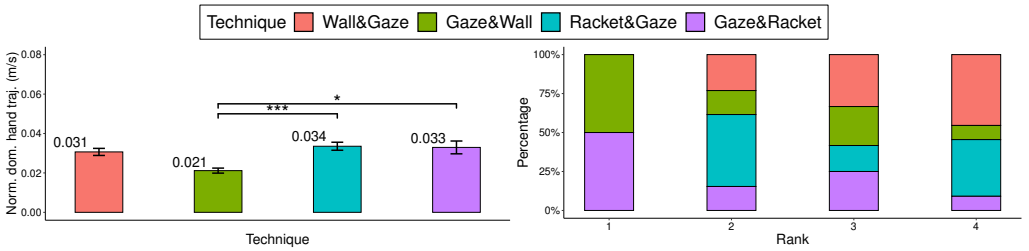


Fig. 7. Normalized dominant hand trajectory per technique in study 1, including significant post-hoc tests. Error bars denote 95% confidence intervals.

Fig. 8. Percentages of rankings for each technique *TECHNIQUE* in study 1.

Gaze&Wall (Mdn = 3, p=.002) and Racket&Gaze (Mdn = 3.5, p=.013). Participants also felt more hand fatigue with Wall&Gaze (Mdn = 4.5) compared to Gaze&Wall (Mdn = 3, p=.002), Racket&Gaze (Mdn = 3.5, p=.025) and Gaze&Racket (Mdn = 3, p=.015).

4.3.3 User Feedback. Users found that they could be more precise with Wall&Gaze when the targets were close (P7: *'... Racket&Gaze sometimes made it harder to be precise than with the Wall&Gaze, but it will most likely depend on where the objects are ...'*). Otherwise, they perceived the technique as strenuous because the target was easily overshot despite gentle hand movements (P6: *' Difficult as slight hand movements would shift the target to some other balls, leading to many errors ´*). Physical and mental demand was reported, e.g., *'There is too much strain on my wrists required much more focus* (P8). Three participants found the technique easier and more precise (*'I found Wall&Gaze more accurate compared to the other techniques'*) (P4, P5, P12 similar).

Five users noted limitations of Racket&Gaze, especially in cases of high occlusion, where it was challenging to select targets precisely (e.g., P5: *' When occlusion was high, it was much harder to differentiate them and the plane did not help much'*).

Gaze&Wall was regarded positively (P2: *'...certainly more comfortable to use gaze...'*, P8: *'faster and more intuitive'*). Users found the use of the plane complementary to the eyes: *'The plane on that axis is nice because it makes up for the lack of depth in the eye-gaze detection.'* (P1). However, it was occasionally reported as imprecise (P1), especially when attempting the trial without even using the plane, leading to more errors, particularly in high occlusion.

Gaze&Racket was a favourite among users, and seven users stated it is simple and straightforward, appreciating the additional freedom it offered from hand constraints (P3: *'It is faster, and gives freedom for the hand'*). However, users found it challenging when occlusion was high (P5: *'..., it was much harder to differentiate them'*). This was partly due to the range of motion of the hand plane, which, in cases of very high occlusion, led to excessive hand movements. Occasionally, users became frustrated because hand rotation seemed somewhat arbitrary (P12: *'Rotating the hand plane with my hand seemed a bit like a hit or miss'*).

4.4 Discussion

The plane-first mode techniques were slower than the gaze-first mode techniques. This was likely due to the need for the hand to be moved carefully and precisely to prevent excessive movement of the plane and overshooting of targets. Although this modality priority allowed for more precise selection, carrying it out over an extended period led to fatigue and mental strain. The error rate was higher with Gaze&Racket as gaze first users work very quickly and could accomplish tasks without the plane. Longer time to focus gaze also lessened overshooting. Also gaze-first mode

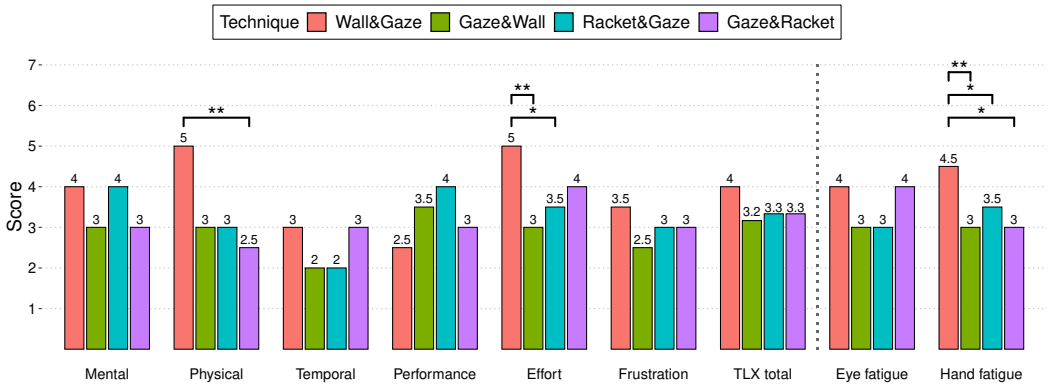


Fig. 9. Median NASA TLX scores and eye/hand fatigue scores per *TECHNIQUE* in study 1, including significant post-hoc tests.

led to errors. The hand performed the pinch gesture while the eyes were already looking for the next target. Regarding the *Racket Metaphor*, users showed greater flexibility but carried the risk of reduced precision and required more extensive hand and body movement, primarily in depth. Specifying only depth using the *Wall* made the technique, especially in gaze-first mode, less physically demanding. This and general 3D visualisation difficulties amplified issues with depth perception. Other insights included that despite training, many users often did not perform a full pinch gesture or their hands left the tracking area, increasing errors, and the Heisenberg problem where jittery hand pointing and pinch led to intersecting with a wrong target.

Overall, the results highlight a trade-off between speed and error across the techniques, with a clear trend towards gaze-first mode techniques. This means that considering the temporal aspect, it becomes evident that both gaze-based techniques enable faster selection, even significantly faster than with Wall&Gaze. Furthermore, Gaze&Wall and Gaze&Racket were dominantly preferred by users because they found them to be more comfortable, faster, and more intuitive due to the use of gaze control. This is the rationale behind our decision to advance with these techniques to the next study.

5 User Study 2: Comparative study

Based on the insights gained from the first study, we refined the gaze-first mode techniques for a second study. The second study investigates two questions: *(RQ1) How does the Wall Metaphor visual feedback fare compare to Racket Metaphor visual feedback?* To determine how the visual feedback of each metaphor contributes to a more precise selection. and *(RQ2) How do Gaze&Wall and Gaze&Racket compare to Ray&Racket and Ray&Wall regarding user performance?* Since it's interesting to determine whether users select faster and more accurately with techniques, in combination with gaze, compared to just using hands.. Our goal was to comprehend gaze-based techniques, focusing on speed trade-offs and efficiency, particularly in densely populated environments. The tracking accuracy of controllers is technically vastly superior to the current state of camera hand tracking. Nevertheless, we think controller-less interaction techniques offer strong benefits that makes them worth investigating (see Section 1). As such, we aimed to assess how gaze-based techniques distinguish themselves in comparison to the controller-based techniques from the literature, which we adapted into gesture-based techniques for a more accurate, fair, and meaningful comparison. This means we expand on the previous study in two ways: (1) Baselines,

we design gesture-only based techniques as baselines (Figure 4a). (2) We focus on the most efficient modality sequence from study 1. As these techniques require a combination of gaze and hand modalities, we tested two different modality sequences within the concept of Gaze+Plane. To ensure consistency in modality sequences, we adapted the "gaze first, plane second" principle for gesture-based techniques to "ray first, plane second". For gesture-based techniques, this order is not relevant as they are uni-modal.

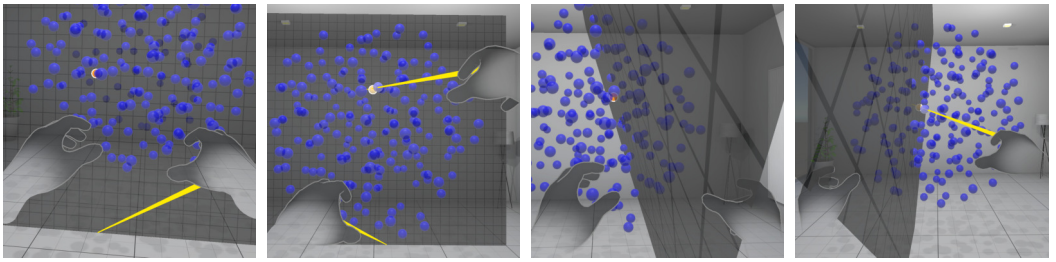
We employed a 4×2 design of the variables *TECHNIQUE* and *DENSITY*. The order of conditions (combination of *TECHNIQUE* and *DENSITY*) was counter-balanced using a balanced Latin square. We compare the following interaction techniques: Gaze&Wall, Gaze&Racket, Ray&Wall, and Ray&Racket (Figure 10). Each technique is tested with the following densities: D_{Low}=100, D_{High}=200 (Figure 11). The low density was chosen as targets do not overlap, and the highest was chosen to create greater target occlusion. The study design for study 2 was: 12 *participants* $\times$ 4 *techniques* $\times$ 2 *densities* $\times$ 30 *trials* = 2880 data points in total.

We recruited another 12 participants (3 female, 8 male, and 1 non-binary) via email and word of mouth from and outside the local university and community. On a scale between 1 and 5, participants rated their knowledge of XR/VR/AR as little to moderate ($M = 3, SD = 1.21$), 3D hand gestures ($M = 2.75, SD = 1.42$), and eye-hand interaction ($M = 2.33, SD = 1.50$). The ages of the participants ranged from 22 to 39 ($M = 26.50, SD = 4.50$). 10 were right-handed and 2 left-handed, 2 wore glasses and 1 wore contact lenses.

5.1 Changes from Study 1

We use the same apparatus as Study 1. Following the results, observations, error rates, and user feedback from Study 1, we aim to optimize and adjust the following changes to further improve the techniques.

- Error Reduction: we introduce a target-locking mechanism to reduce slipping errors. If the pinch strength as reported by the hand tracking module is above 0.3, we lock the currently gaze-preselected targets. Above 0.4, we lock the current hovered target and above 0.5 also the plane. This better captures a participant's intent to confirm a target and prevents errors when either the gaze or plane slips off the correct target.
- Visual Perception: We improved the plane design to better perceive its relative depth to the user by adding a grid pattern to the plane. This also helps to identify the orientation of the racket, especially when the edges are outside of the user's field of view. We also made the plane a bit darker, making it clearer what targets are behind the plane and what targets are cut by the plane. We changed its width to 4 m and its height to 3.5 m to better visualize its position in the virtual room. Finally, we modified room lighting, making shadows appear directly under the target cloud on the floor. We think this may additionally improve depth perception.
- Visual Feedback: We disabled target hover when it is behind the active plane in the direction of a user's head for finer selection using the plane as a depth filter. We also reduced the hover dwell time to 200 milliseconds for more responsive target selection. We changed the color of the target outline from blue to white to see targets clearer in the more dense task of study 2. We disabled the pre-selection feedback for gaze-based techniques to reduce visual clutter. For the gesture-based techniques, opaque light orange indicates that the target of interest is cut by the ray.
- In the first study, there was a relatively high error rate across techniques, indicating the task complexity. To alleviate this issue, the second study has users sitting, which allows them to rest their arms, reduces fatigue and hand tremors, and renders hand tracking more robust. This



(a) Gaze&Wall (improved) (b) Ray&Wall (c) Gaze&Racket (improved) (d) Ray&Racket

Fig. 10. Screenshots of the techniques in use for study 2 from a user's perspective. Gaze&Wall (a) via gaze cone, and Ray&Wall (b) via hand-ray, select the target candidates, and disambiguation happens via proximity to the plane. Gaze&Racket (c) and Ray&Racket (d) use rays equally, but with the racket metaphor.

allows for the reduction of potential bias from errors and gives a clearer picture of the relative differences between the technique designs.

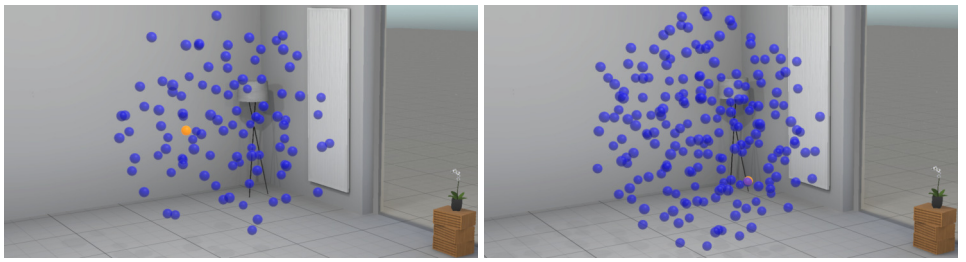
5.2 Baseline: Gestural Ray + Plane Interaction Techniques

For evaluation of the Gaze+Plane techniques, we designed two hand-based techniques. The natural baseline would be a gesture-only technique instead of a controller, given the scope of the paper on gestural interfaces. As there is no gesture baseline from prior work, we adopted two concepts. First, we opted for the ray-and-pinch technique, where users point via hand-ray and pinch to confirm, as widely adopted in the current status quo of HMDs (Meta Quest, Hololens 2) for hand-based interactions in hand-tracking UIs. Other gestures like thumb-bend may be potentially interesting, but they haven't been extensively tested and aren't used in current HMDs. E.g., thumb-bending may still be subject to the Heisenberg effect despite their unfamiliarity, as when you move your thumb, the hand can be affected, but is of interest in future iterations. Second, we adopt the best-performed controller technique, the Semi-Automatic (SA) Conductor [52] for 3D specification.

The main idea is that both techniques have a plane controlled by the non-dominant hand, intersected with a ray controlled by the dominant hand. This is similar to the gaze-based techniques, but replacing the gaze-ray with the hand-ray. Second, confirmation is performed by a pinch gesture of the non-dominant hand, for two reasons.

It's beneficial to have the DH assigned only to ray pointing and not pinch, as this task involves high precision and typically comes after the NDH sets the frame of reference (Z) for the DH. Following that, the NDH performs primarily a 1DOF depth specification task (Z), whereas the DH performs a 2DOF pointing task. Moving pinch to the NDH balances workload across hands. Visual feedback for the gesture-based techniques is similar to the gaze-based techniques. Targets cut by the ray are slightly transparent and light blue. A blue outline indicates that the target cut by the ray is closest to the plane and thus can be selected by pinching.

5.2.1 Ray&Wall (Figure 4a, Figure 10b). Ray&Wall is an interaction technique that utilises the user's non-dominant hand to determine depth in 3D space with a horizontal plane. The user's non-dominant hand is used for controlling depth (Z) by moving the wall, equal to Gaze&Wall. The wall is manipulated by a hand-ray, that originates from the user's wrist and points 45 degrees downward toward the middle finger and 15 degrees toward the thumb (like Gaze&Wall).



(a) Low density $D_{Low} = 100$ (b) High density $D_{High} = 200$

Fig. 11. Target densities used in study 2. The higher density leads to more occlusion.

The user's dominant hand is used to determine the direction with a ray and thus defines the X and Y coordinates (see Figure 4b). Its direction is controlled by simply pointing outwards with the hand, based on a standard absolute virtual pointer metaphor , that originates from the user's wrist and points 20 degrees downward toward the middle finger. We smooth both rays for pointing and manipulating with a 1€ Filter. Like Zhang et al. [52], we apply the parameters $f_{c_{min}} = 0.1$ and $\beta = 50$.

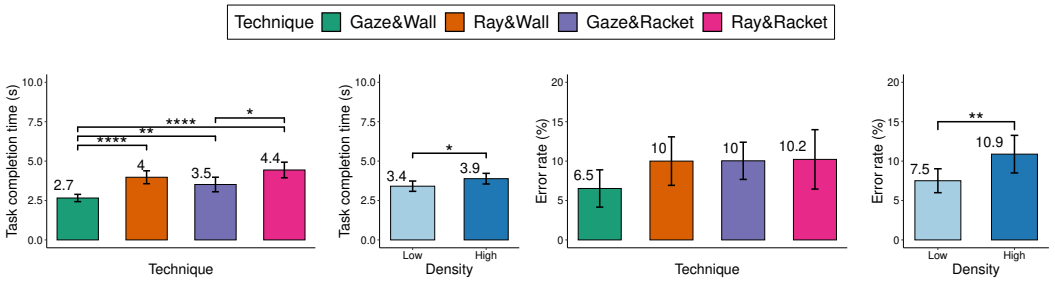
5.2.2 Ray&Racket (Figure 4b, Figure 10d). While SA Conductor was built with controllers, we implemented our hand-based techniques with hand tracking and selections were confirmed with a pinch gesture at the non-dominant hand for comparing with Gaze&Racket. The techniques select targets independently from the ray-plane intersection point and targets can be selected with only the hand ray on the dominant hand. When multiple targets appear on the ray, the plane on the non-dominant hand acts as a depth indicator and the closest target to the plane is selected. However, it is only possible to select targets that are either sliced by the plane including a tolerance margin of 1 cm or that are before the plane, towards the user's head. For comparing with Gaze&Racket and Gaze&Wall, we implemented Ray&Racket and Ray&Wall with Racket and Wall as the plane on the non-dominant hand, respectively.

5.3 Task and Procedure

For the occlusion tasks, we tested two different target densities (Figure 11). The low density D_{Low} contained 100 targets, while the high density D_{High} included 200 targets. The high target density was aimed at inducing occlusion effects. Each density included 30 trials. Targets of interest were specifically located in the rear half of the sphere. The participants were instructed to select the target object, which was partly heavily obscured due to the object density, as quickly and accurately as possible. All targets were pseudo-randomly placed within a 2-meter diameter sphere. The sphere center had a distance of 3 meters from the user.

5.4 Results

We use the same methods to analyse the data as in Study , including checking assumptions for statistical tests and corrections.. We eliminated 5 timeouts, comprising 0.2% of the trials (1 for Ray&Wall, 2 for Gaze&Racket, 2 for Ray&Racket). For the analysis of TCT, 62 outliers were excluded if a trial time exceeded the $Mean+3\times SD$ threshold. Concerning TCT, a total of 2.2% of trials were removed (13 for Gaze&Wall, 16 for Ray&Wall, 17 for Gaze&Racket, 16 for Ray&Racket).



(a) Task completion time per *TECHNIQUE* and per *DENSITY*

(b) Error rate per *TECHNIQUE* and per *DENSITY*

Fig. 12. Study 2: Mean task completion time and error rate with significant post-hoc tests. Error bars denote 95% confidence intervals.

5.4.1 User Performance (Figure 12). Regarding TCT, we found a significant main effect of the factor *TECHNIQUE* ($F^{2.1}_{23.1} = 17.63$, p<.0001, $\eta^2_G = 0.391$) and *DENSITY* ($F^{1}_{11} = 8.35$, p=.015, $\eta^2_G = 0.079$). Users were significantly faster with Gaze&Wall (2.7s) than the other three techniques (Ray&Wall: 4.0s, p<.0001, Gaze&Racket: 3.5s, p=.004, Ray&Racket: 4.4s, p<.0001) and Gaze&Racket (3.5s) is faster compared to Ray&Racket (4.4s, p=.026). The users were also significantly faster with D_{Low} (3.4s) compared to D_{High} (3.9s, p=.015). In case of errors ($F^{1}_{11} = 11.17$, p=.007, $\eta^2_G = 0.065$), users made significantly fewer errors with D_{Low} (7.5%) compared to D_{High} (10.9%, p=.007).

5.4.2 Head and Hand Movement (Figure 13). Regarding head movement ($F^{2.0}_{21.5} = 7.29$, p=.004, $\eta^2_G = 0.201$), participants moved their head more with Gaze&Racket (0.010) compared to two other techniques (Gaze&Wall: 0.006, p<.001, Ray&Wall: 0.007, p=.008) and more with Ray&Racket (0.008) compared to Gaze&Wall (0.006, p=.010). With regards to hand movement ($F^{3}_{33} = 9.63$, p<.001, $\eta^2_G = 0.331$), Participants used their dominant hand more with Gaze&Racket (0.036) compared to all other techniques (p<.0001, Gaze&Wall: 0.015, Ray&Wall: 0.020, Ray&Racket: 0.018). Lastly, for non-dominant hand motion ($F^{3}_{33} = 6.97$, p<.001, $\eta^2_G = 0.224$) participants moved more with Gaze&Racket (0.039) compared to Ray&Racket (0.023, p=.021) and more with Ray&Racket (0.023) compared to two other techniques (p<.001, Gaze&Wall: 0.017, Ray&Wall: 0.015).

5.4.3 Usability Questionnaire (Figure 13d). Gaze&Racket was most preferred (66.6%) together with Gaze&Wall (16.6%) and Ray&Racket (16.6%). Regarding the usability questions, no significant difference was found except for eye fatigue ($\chi^2(3) = 17.39$, p<.001, $W = 0.483$) and hand fatigue ($\chi^2(3) = 9.73$, p=.021, $W = 0.270$). Ray&Racket was perceived as less fatiguing for the eyes ($Mdn = 2$) compared to Gaze&Racket ($Mdn = 3$, p=.042). Ray&Wall was perceived as more fatiguing for the hands ($Mdn = 5$) compared to Gaze&Wall ($Mdn = 4$, p=.038).

5.4.4 User Feedback. Gaze&Racket was favoured by eight users and overall positively received. E.g., P2 states *'It felt least straining to use while also doing what I wanted to to most of the time.'*. Users noted that compared to the gesture-based variant, hand fatigue decreased, while eye fatigue remained minimal (P4: *'Both gaze methods were much better in terms of hand fatigue, with barely any eye fatigue'*). The freedom of hand movement was perceived as intuitive and natural, e.g. P9 says: *'I did not have to move my hands so much with the gaze. I prefer Gaze&Racket because I could rotate the plane to filter the balls that match the specific swarm of balls in that situation.'*). But users also reported that the freedom of movement led to more errors and increased hand effort, which

ultimately became frustrating (P1: *'...because using my gaze for choosing targets was jumpy and felt fast-fast-paced, and I felt like I made more frustrating errors because I could move the plane in so many ways.'*).

For two people Gaze&Wall was favoured because it was more intuitive, simpler, and less tiring when using gaze, including reduced hand fatigue (P12: *'The Gaze&Wall was easier to control. Pointing with the eyes demanded less effort,...'*). Although some users found the orthogonal plane and its associated "restricted" movement on the z-axis easy to control (P1: *'...the movement of the plane was one-dimensional, and not as confusing'*), many users noted that the plane was more sensitive to hand movements. They had to concentrate on not moving their hand abruptly, as otherwise, they couldn't achieve the necessary precision. This subsequently led to frustration and fatigue (P7: *'The Gaze&Wall was more sensitive to my hand movements, ..., was frustrating and fatiguing'*).

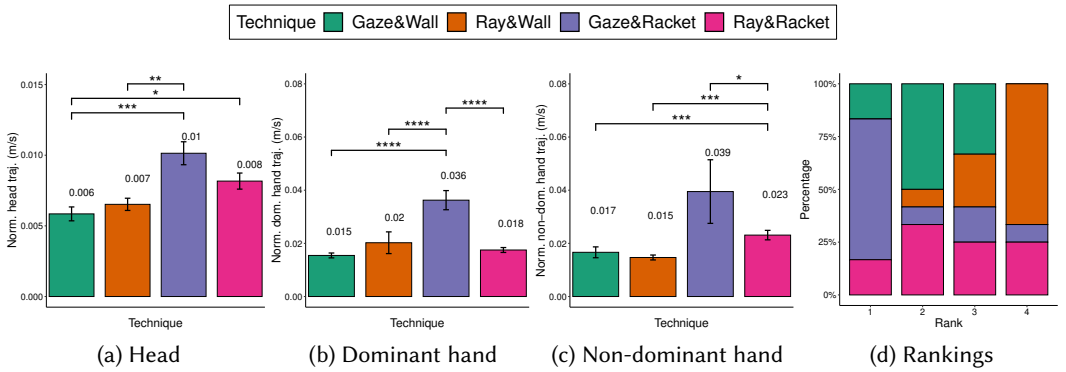
Two users rated Ray&Racket as popular because they had better control with the "Ray" compared to the gaze-based techniques, resulting in more successful outcomes (P1: *'... was the easiest, I felt more in control when using my hand for choosing the target instead of my gaze(felt jumpy)'*). Even though they were more precise with the "Ray" in their dominant hand, users found the gesture-based technique somewhat peculiar because they had to control the plane with the same hand they used for the pinch gesture to confirm actions (P3: *'using hands seems to select from such a distance feels weird'*). Some were frustrated that the pinch gesture didn't work when they had to rotate their hand to reach the target. Beyond a certain rotation, the hand tracking could no longer recognize the pinch gesture as intended (P9: *'... was a bit annoying because rotating the plane with my left hand also affected the pinching (the camera could not see my fingers'*). Furthermore, performing the pinch gesture always involved some hand movement, which contributed to unintentional movements of the plane. As P12 pointed out, controlling the plane in the non-dominant hand was physically demanding, and performing the pinch gesture simultaneously was difficult to control (P12: *'Pinching and Ray&Racket was very hard to control, it moved the plane too much when I was pinching'*).

Eight users rated Ray&Wall as the least favourite, and there were several reasons for this. Firstly, users found it challenging and awkward to select objects from a certain distance using both hands. It was also considered complicated because the non-dominant hand was responsible for both controlling the plane and confirming actions with the pinch gesture (P7: *'...left hand has two functions, which was sometimes confusing and/or difficult to perform'*). Furthermore, it was difficult to position the plane smoothly in depth as each pinch gesture triggered hand movement (P10: *'Plane mapping to the floor was harder to get a feel for and kept demanding my attention...'*). Over extended periods, this technique proved physically demanding and very tiring for users (P4: *'...was physically demanding'*).

5.5 Discussion

The result of the study showed that gaze-based techniques, especially Gaze&Wall, perform faster and more efficiently than raycasting techniques. We assume that the following primary factors contribute to this outcome. Firstly, we replaced the demanding pointing task of the dominant hand with the faster gaze input. Secondly, we delegated control of the plane to the dominant hand for a more precise manipulation of the plane. In the case of bare-hand techniques, users had to focus on coordinating both hands, with the non-dominant hand assigned two tasks. However, the Heisenberg issue made with dual pointing and pinching it particularly challenging for the non-dominant hand. In both Ray&Racket and Ray&Wall, this rotation often resulted in the hand moving out of the tracking area or the system misinterpreting the user's pinch gesture, introducing unintended selections and increasing user demand and frustration.

Gaze&Wall offers, among other things, the advantage of allowing the depth plane to be solely controlled and determined by the hand, while gaze is responsible only for rapid selection on the x-



(a) Head (b) Dominant hand (c) Non-dominant hand (d) Rankings

Fig. 13. (a-c) Normalized head and hand trajectory per *TECHNIQUE* in study 2, including significant post-hoc tests. Error bars denote 95% confidence intervals. (d) Percentages of rankings for each technique in study 2.

and y-axis. Of course, further research is needed to, for example, reduce error rates, as Gaze&Wall was not less error-prone than the other techniques in Study 2.

6 Overall Discussion

Across two studies, we have focused on understanding the Gaze+Plane interaction and optimizing their capability of 3D specification, specifically examining the gaze modality as a substitute to manual raypointing. Additionally, we have compared gaze-based and bare-hand techniques, evaluating their potential advantages, particularly in speed trade-offs in dense environments. Our first experiment assessed input parameters and fine-tuning possibilities for the Gaze + Plane techniques. The techniques that used a plane-first mode were slower in comparison to the gaze-first mode techniques, showing that a gaze-then-hand sequence is more efficient, while the careful and precise hand movements to avoid excessive plane movement led to increased fatigue and mental strain. The second study was our main experiment – comparing multimodal to unimodal techniques – with the result that gaze-based techniques, especially Gaze&Wall, outperformed manual techniques. This finding suggests that gaze-based methods have the potential to provide faster and more efficient interactions, showcasing their advantages in scenarios where speed and accuracy are crucial. The results from both studies contribute to our understanding of strengths and limitations of different gaze-based interaction techniques and offer insights into their potential applications and improvements.

In particular, across both our studies, we find that there is a discrepancy between user performance and preferences. In the first study, users noted that gaze-precedence was considered faster and more intuitive, which we speculate that it led to the higher rankings. In the second study, we find that Gaze&Racket was preferred although the performance of Gaze&Wall was better. Here we think it is related to users liking the free racket movement with their hand, finding it intuitive and natural, even though Gaze&Wall had faster task completion times. This shows that users preferred natural use, however in situations where performance is favoured, we speculate that Gaze&Wall is more suitable.

An essential consideration in the evaluation of these techniques is that not all tasks involve occluded targets. It is crucial to emphasize that interaction techniques must accommodate both occluded and non-occluded scenarios. In cases where occlusion is not a factor, hand-based input, such as hand ray, may offer advantages, particularly for precise selection of small targets. This

versatility ensures that users can efficiently interact in various contexts in VR. The Gaze+Plane approach bears a strong resemblance to Gaze + Pinch interaction [31], especially if the target is not occluded. One can consider it as *'Gaze+Plane ≈ Gaze + Pinch + Plane'*, where the plane is an optional component to use on-demand that complements the default UI. For example, in a computer-aided design application as envisioned in Figure 1, the user can instance a supplementary plane to the available hand to accomplish 3D specification.

While our techniques have an effective division of labour across the eyes and the hand, they are not one-handed as we aimed to minimise error rate by avoidance of the Heisenberg problem [47]. As we set out with the premise that gaze can substitute a hand's pointing action, it is theoretically possible to achieve unimanual techniques. For example, with controllers, the Heisenberg problem does not really exist, and coupling with eye-tracking allows in principle, one-handed operation. Eye-hand UI systems can support one-handed control of conflicting tasks such as navigation and interface translation [33], and exploring how the eyes and hands map to various degree-of-freedom tasks and reference frames can give further insight into this matter. Other venues to investigate may include the development of advanced algorithms for target selection, more robust hand-tracking technology, or alternative pointing confirmation methods that come with their own pro and cons [24, 27].

While our study offers valuable insights into 3D interaction techniques, it is essential to acknowledge several limitations for the interpretation of our findings. First, we did not specifically address the interaction with small targets, a factor known to influence the efficiency of gaze-based interactions. Second, we note that our occlusion tasks did not consistently generate fully occluded targets, which introduces variability in the data. Third, our work demonstrates one instance of the wall or racket concept, other design parameters are left for future work. For example, the wall size can involve trade-offs in visibility and reachability of targets, and with Racket, alternative designs could ease the interaction with a wider range of angles. Fourth, we developed gestural techniques adopted from the prior art in controllers with a specific division of labour across hands. We assigned pinch confirmation of the task to the non-dominant hand. Other handedness and input mappings are still to be tested, and could affect the outcome of the user study. Applying pinch confirmation to the dominant hand could boost performance as the dominant hand offers more precision, but can also be detrimental as the hand may be overloaded with high degrees-of-freedom tasks (pointing+confirmation). Fifth, our second study had sitting participants for more convenient task repetitions and to lower error rates without affecting relative differences among the techniques, but future studies are required to fully understand potential bias.

Sixth, we may not be able to make direct comparisons with studies using a 21-point scale by utilizing a seven-point scale for each element. However, it still provides a reliable basis for internal comparison between the conditions. Seventh, it would have been interesting to conduct a power analysis to determine if this was the appropriate sample size for their study or to acknowledge and discuss the small sample as a limitation of their results. Eighth, although we believe that every participant faced the same challenge in completing the random task, we found during our investigation that it might have been better to employ a fixed distribution of tasks in random order. This approach would have allowed each participant the opportunity to perform each distribution once. Lastly, this paper's scope precluded controller-based techniques (e.g., [4, 10, 52]), for an in-depth exploration of the gestural UI. Future research can expand the scope for a comprehensive perspective, especially considering the input vocabulary and physical delimiters unique with controllers. Despite these limitations, our study represents an essential step towards advancing the field of 3D disambiguation techniques and provides a foundation for future research endeavours.

7 Conclusion

In this work, we explored how the use of eye-tracking input can enhance gestural raypointing for the selection of 3D points, particularly in densely populated environments. We developed new techniques, Gaze&Wall and Gaze&Racket, based on the Gaze+Plane principle, as well as two additional new techniques, Ray&Racket and Ray&Wall, following the dual-pointing approach. In a point-and-select task, we initially examined the speed-accuracy trade-offs of the Gaze+Plane techniques themselves. Subsequently, we compared the best-performing techniques with the newly developed bare-hand raypointing techniques.

With our research, we provide valuable insights into multimodal gaze and hand-based techniques, exploring temporal, spatial and task-related aspects. This is because 1) Gaze&Wall, in comparison to all other methods, leads to a reduction in errors and a significant decrease in the required time, providing itself as a powerful technique, and 2) wall-based techniques inherently result in reduced physical movement of the head and hand movements.

Our research is relevant for spatial interaction, specifically focusing on advanced techniques for complex 3D tasks. We provide insights and approaches that can enhance the effectiveness and efficiency of interactions in three-dimensional environments. Furthermore, our findings contribute to the broader field of human-computer interaction by highlighting the need to harness the unique capabilities of both eyes and hands. We thereby aim to reduce user effort and enhance efficiency, ultimately facilitating more natural and immersive interactions in complex 3D environments, which can prove highly beneficial in areas such as 3D modelling, and simulations.

Acknowledgments

This work has received funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grants no. 101021229 GEMINI, and no. 740548 CIO).

References

[1] Sunggeun Ahn, Stephanie Santosa, Mark Parent, Daniel Wigdor, Tovi Grossman, and Marcello Giordano. 2021. StickyPie: A Gaze-Based, Scale-Invariant Marking Menu Optimized for AR/VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 739, 16 pages. https://doi.org/10.1145/3411764.3445297

[2] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136.

[3] Felipe Bacim, Mahdi Nabiyouni, and Doug A Bowman. 2014. Slice-n-Swipe: A free-hand gesture user interface for 3D point cloud annotation. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 185–186.

[4] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300331

[5] Christoph Bichlmeier, Sandro Michael Heining, Marco Feuerstein, and Nassir Navab. 2009. The virtual mirror: a new interaction paradigm for augmented reality environments. *IEEE Transactions on Medical Imaging* 28, 9 (2009), 1498–1510.

[6] Doug Bowman, Chadwick Wingrave, Joshua Campbell, and Vinh Ly. 2001. Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments. (2001).

[7] G. E. P. Box and D. R. Cox. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* 26, 2 (1964), 211–252. http://www.jstor.org/stable/2984418

[8] W. Buxton and B. Myers. 1986. A Study in Two-Handed Input. *SIGCHI Bull.* 17, 4 (apr 1986), 321–326. https://doi.org/10.1145/22339.22390

[9] Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1 € Filter: A Simple Speed-Based Low-Pass Filter for Noisy Input in Interactive Systems *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2527–2530. https://doi.org/10.1145/2207676.2208639

[10] 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* (Christchurch, New Zealand) *(VRST '23)*. Association for Computing Machinery, New York, NY, USA, Article 19, 11 pages. https://doi.org/10.1145/3611659.3615693

[11] L. Colligan, H. W. Potts, C. T. Finn, and R. A. Sinkin. 2015. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record, Vol. 84. International journal of medical informatics, 469–476. https://doi.org/10.1016/j.ijmedinf.2015.03.003

[12] William Delamare, Maxime Daniel, and Khalad Hasan. 2022. MultiFingerBubble: A 3D Bubble Cursor Variation for Dense Environments. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 453, 6 pages. https://doi.org/10.1145/3491101.3519692

[13] Niklas Elmqvist. 2005. BalloonProbe: Reducing occlusion in 3D using interactive space distortion. In *Proceedings of the ACM symposium on Virtual reality software and technology*. 134–137.

[14] Niklas Elmqvist and M Eduard Tudoreanu. 2007. Occlusion Management in Immersive and Desktop 3D Virtual Environments: Theory and Evaluation. *Int. J. Virtual Real.* 6, 2 (2007), 21–32.

[15] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, Tim Dwyer, Jens Grubert, Jason H. Haga, Nurit Kirshenbaum, Dylan Kobayashi, Tica Lin, Monsurat Olaosebikan, Fabian Pointecker, David Saffo, Nazmus Saquib, Dieter Schmalstieg, Danielle Albers Szafir, Matt Whitlock, and Yalong Yang. 2021. Grand Challenges in Immersive Analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 459, 17 pages. https://doi.org/10.1145/3411764.3446866

[16] Alex Olwal Steven Feiner. 2003. The flexible pointer: An interaction technique for selection in augmented and virtual reality. In *Proc. UIST*, Vol. 3. 81–82.

[17] Jonathon B Ferrell, Joseph P Campbell, Dillon R McCarthy, Kyle T McKay, Magenta Hensinger, Ramya Srinivasan, Xiaochuan Zhao, Alexander Wurthmann, Jianing Li, and Severin T Schneebeli. 2019. Chemical exploration with virtual reality in organic teaching laboratories. *Journal of Chemical Education* 96, 9 (2019), 1961–1966.

[18] Tovi Grossman and Ravin Balakrishnan. 2006. The Design and Evaluation of Selection Techniques for 3D Volumetric Displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (Montreux, Switzerland) *(UIST '06)*. Association for Computing Machinery, New York, NY, USA, 3–12. https://doi.org/10.1145/1166253.1166257

[19] Yves Guiard. 1987. Asymmetric Division of Labor in Human Skilled Bimanual Action. *Journal of Motor Behavior* 19, 4, 486–517. https://doi.org/10.1080/00222895.1987.10735426

[20] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300855

[21] Robert J. K. 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* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[22] Robert J. K. Jacob, Linda E. Sibert, Daniel C. McFarlane, and M. Preston Mullen. 1994. Integrality and Separability of Input Devices. *ACM Trans. Comput.-Hum. Interact.* 1, 1 (mar 1994), 3–26. https://doi.org/10.1145/174630.174631

[23] 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173655

[24] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-Air Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 63–71. https://doi.org/10.1145/2807442.2807489

[25] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. 2020. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 930–939. https://doi.org/10.1109/VR46266.2020.00113

[26] 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. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 145 (may 2022), 18 pages. https://doi.org/10.1145/3530886

[27] 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*. Association for Computing Machinery, New York, NY, USA, Chapter 15, 7. https://doi.org/10.1145/3448018.3457998

[28] Mathieu Nancel, Olivier Chapuis, Emmanuel Pietriga, Xing-Dong Yang, Pourang P. Irani, and Michel Beaudouin-Lafon. 2013. High-Precision Pointing on Large Wall Displays Using Small Handheld Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 831–840. https://doi.org/10.1145/2470654.2470773

[29] Ken Pfeuffer, Yasmeen Abdrabou, Augusto Esteves, Radiah Rivu, Yomna Abdelrahman, Stefanie Meitner, Amr Saadi, and Florian Alt. 2021. ARtention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics* 95 (2021), 1–12. https://doi.org/10.1016/j.cag.2021.01.001

[30] Ken Pfeuffer, Jason Alexander, and Hans Gellersen. 2016. Partially-Indirect Bimanual Input with Gaze, Pen, and Touch for Pan, Zoom, and Ink Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2845–2856. https://doi.org/10.1145/2858036.2858201

[31] 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* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[32] Ken Pfeuffer, Lukas Mecke, Sarah Delgado Rodriguez, Mariam Hassib, Hannah Maier, and Florian Alt. 2020. Empirical Evaluation of Gaze-Enhanced Menus in Virtual Reality. In *26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) *(VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 20, 11 pages. https://doi.org/10.1145/3385956.3418962

[33] Ken Pfeuffer, Jan Obernolte, Felix Dietz, Ville Mäkelä, Ludwig Sidenmark, Pavel Manakhov, Minna Pakanen, and Florian Alt. 2023. PalmGazer: Unimanual Eye-hand Menus in Augmented Reality. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction (SUI '23)*. Association for Computing Machinery, New York, NY, USA, 10 pages. https://doi.org/10.1145/3607822.3614523

[34] Robin Piening, Ken Pfeuffer, Augusto Esteves, Tim Mittermeier, Sarah Prange, Philippe Schröder, and Florian Alt. 2021. Looking for info: Evaluation of gaze based information retrieval in augmented reality. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part I 18*. Springer, 544–565.

[35] Thammathip Piumsomboon, Gun Lee, Robert W Lindeman, and Mark Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE symposium on 3D user interfaces (3DUI)*. IEEE, 36–39.

[36] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-Worn Extended Reality. *ACM Comput. Surv.* 55, 3, Article 53 (mar 2022), 39 pages. https://doi.org/10.1145/3491207

[37] Rongkai Shi, Yushi Wei, Xueying Qin, Pan Hui, and Hai-Ning Liang. 2023. Exploring Gaze-Assisted and Hand-Based Region Selection in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 7, ETRA, Article 160 (may 2023), 19 pages. https://doi.org/10.1145/3591129

[38] Rongkai Shi, Jialin Zhang, Yong Yue, Lingyun Yu, and Hai-Ning Liang. 2023. Exploration of Bare-Hand Mid-Air Pointing Selection Techniques for Dense Virtual Reality Environments. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 109, 7 pages. https://doi.org/10.1145/3544549.3585615

[39] 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 257, 15 pages. https://doi.org/10.1145/3544548.3580685

[40] Ludwig Sidenmark, Christopher Clarke, Xuesong Zhang, Jenny Phu, and Hans Gellersen. 2020. Outline Pursuits: Gaze-Assisted Selection of Occluded Objects in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376438

[41] Ludwig Sidenmark and Hans Gellersen. 2019. 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* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

[42] Vildan Tanriverdi and Robert J. K. Jacob. 2000. Interacting with Eye Movements in Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 265–272. https://doi.org/10.1145/332040.332443

[43] Mythreye Venkatesan, Harini Mohan, Justin R Ryan, Christian M Schürch, Garry P Nolan, David H Frakes, and Ahmet F Coskun. 2021. Virtual and augmented reality for biomedical applications. *Cell reports medicine* 2, 7 (2021).

[44] Uta Wagner, Mathias N. Lystbæk, Pavel Manakhov, Jens Emil 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, 15 pages. https://doi.org/10.1145/3544548.3581423

[45] 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* (Nantes, France) *(IMX '23)*. Association for Computing Machinery, New York, NY, USA, 216–221. https://doi.org/10.1145/3573381.3596467

[46] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.

[47] Dennis Wolf, Jan Gugenheimer, Marco Combosch, and Enrico Rukzio. 2020. Understanding the Heisenberg Effect of Spatial Interaction: A Selection Induced Error for Spatially Tracked Input Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376876

[48] Hans Peter Wyss, Roland Blach, and Matthias Bues. 2006. iSith-Intersection-based spatial interaction for two hands. In *3D User Interfaces (3DUI'06)*. IEEE, 59–61.

[49] Difeng Yu, Hai-Ning Liang, Xueshi Lu, Kaixuan Fan, and Barrett Ens. 2019. Modeling endpoint distribution of pointing selection tasks in virtual reality environments. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.

[50] Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. Gaze-Supported 3D Object Manipulation in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 734, 13 pages. https://doi.org/10.1145/3411764.3445343

[51] Difeng Yu, Qiushi Zhou, Joshua Newn, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Fully-Occluded Target Selection in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3402–3413. https://doi.org/10.1109/TVCG.2020.3023606

[52] Futian Zhang, Keiko Katsuragawa, and Edward Lank. 2022. Conductor: Intersection-Based Bimanual Pointing in Augmented and Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 6, ISS, Article 560 (nov 2022), 15 pages. https://doi.org/10.1145/3567713

Received 2024-02-22; accepted 2024-05-30