

HeadDepth: Gaze Raycasting with Head Pitch for Depth Control

Haopeng Wang*
Lancaster University

Florian Weidner†
University of Glasgow

Yasmeeen Abdrabou‡
Lancaster University

Ken Pfeuffer§
Aarhus University

Hans Gellersen¶
Lancaster University
Aarhus University

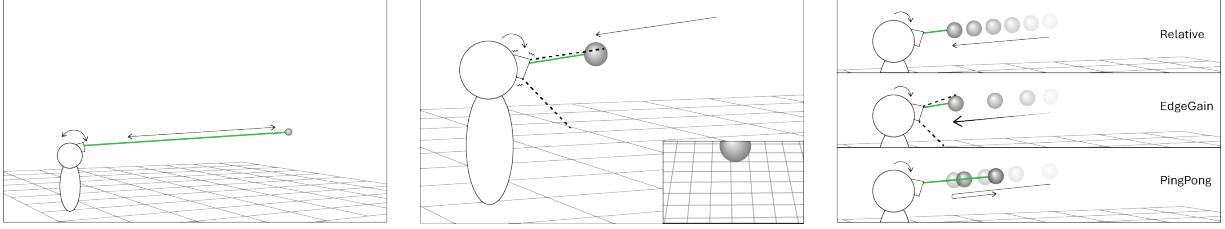


Figure 1: We propose HeadDepth for 3D positioning by gaze raycasting with head pitch for depth control (left). Head pitch is intuitive to use but not straightforward to map for control as it affects viewing angles (middle). We compare three strategies for mapping pitch to depth in a comfortable viewing range: *RELATIVE* velocity-based mapping, *EDGEgain* with gain dependent on eye-in-head angle, and *PINGPONG* as an absolute back-and-forth mapping that repeats at different pitch angles.

ABSTRACT

Gaze is fast and intuitive for raycasting, but lacks a way to control depth for input in 3D. We propose HeadDepth to augment gaze with vertical head rotation (pitch) to control depth along the line of sight, and investigate three pitch-to-depth mappings: *Relative* maps pitch velocity to changes in depth; *EdgeGain* adds dynamic gain dependent at eccentric gaze angles; and *PingPong* provides an absolute back-and-forth mapping that repeats at different pitch angles. HeadDepth is not as fast as controller-based RayCursor but has the advantage of being hands-free. In a user study, all three variants proved effective for 3D positioning; PingPong required least effort and EdgeGain was least affected by differences in tasks. Eye-head coordination and task had a significant effect, as head movements in support of gaze can affect depth control synergistically or antagonistically. Our results have significance beyond HeadDepth as they generalize to any interaction where head rotational input concurs with gaze fixation on visual feedback. Interaction designs may benefit from dynamic adjustment to eye-in-head angle to prevent discomfort and maintain objects within the user’s field of view.

1 INTRODUCTION

Gaze is widely used for input in 3D environments, as it naturally lends itself to fast pointing in any direction [45, 3, 15]. This makes gaze efficient for pointing at objects over any distance but it lacks a means to actively control depth. We propose HeadDepth as a hands-free multimodal method that integrates gaze with depth control by head movement. Depth control by the head enables object movement along the line of sight and in integration with gaze object positioning anywhere in 3D (Figure 1, left).

Head pitch relates intuitively to depth, as we tilt our heads down for many of our near-space interactions and up for interactions over larger distances. However, integration with gaze for 3D input is not straightforward as eye and head movements are coupled. Gaze shifts are routinely supported by head rotation, which can interfere

with the use of head movement for control [12]. Relying on head pitch to move an object along the line of sight implicitly affects the eye-in-head viewing angle and user’s view (Figure 1, middle). The eye-in-head viewing range is limited, more so upward than downward [19], and gaze is less comfortable at eccentric angles [40]. This poses a design problem of mapping pitch to depth in a manner that aligns with comfortable viewing, which we investigate through the design of three mapping strategies and the study of different effects of concurrent eye and head movement on 3D positioning.

Figure 1 (right) illustrates the three mappings we investigate. *RELATIVE* uses a mapping of pitch angular velocity to changes in depth, to enable larger movements in depth with less head rotation. *EDGEgain* adds dynamic gain to increase speed at eccentric gaze angles and keep gaze within a comfortable viewing range. *PINGPONG*, in contrast, uses an absolute mapping of pitch rotation to depth position that repeats in a back-and-forth manner, so that any depth position can be reached within comfortable eye-in-head range, irrespective of how far up or down users tilt their head to gaze at a target. We compared the three mappings in a study (N=16), with controller-based 3D positioning [2] for reference. The study was designed to investigate the influence of small versus large target amplitudes and eccentricities, in particular on visual comfort; and the influence of eye-head coordination and task on HeadDepth performance, as head movements performed with gaze can be synergistic or antagonistic in their implicit effect on depth control.

HeadDepth proved effective for 3D positioning across all three mappings. It is naturally not as efficient as controller-based input but enables hands-free input. *PINGPONG* required least head movement and *RELATIVE* most. *EDGEgain* was least impacted by eye-head interaction effects. Synergistic tasks required less time and effort than antagonistic ones, and performance was worst on tasks where gaze shifts interfered with keeping depth unchanged. These results are significant for the design of 3D input techniques that integrate input from gaze and head. They also have wider significance as they generalise to use of head rotational input that is concurrent with gaze attention to visual feedback.

2 RELATED WORK

We build on background of gaze input in 3D, head movement for depth control, and eye and head coordination for multimodal input.

2.1 Gaze Input in 3D Environments

Gaze is natural for pointing and faster than hand or head for raycasting in 3D [45, 3, 15]. It can support the selection of objects over

*e-mail: h.wang73@lancaster.ac.uk

†e-mail: florian.weidner@glasgow.ac.uk

‡e-mail: y.abdrabou@lancaster.ac.uk

§e-mail: ken@cs.au.dk

¶e-mail: h.gellersen@lancaster.ac.uk

any distance [33, 41], but it lacks a method for active depth control. The eyes naturally converge to focus on objects. Vergence eye movement (and similarly differential eye-head movement [23]) can be used to discriminate or switch between objects that are nearer or farther in the line of sight [24, 55, 38]. However, vergence occurs only in reaction to a stimulus and cannot afford deliberate user-controlled depth input. Problems such as the selection of occluded, dense, or moving objects instead require a separate interaction step or complementary modality [39, 52]. Prior work has used controller or hand input to control a ray or plane to manipulate the depth at which gaze intersects [8, 46], or to control depth directly by hand movement (but limited to near space) [47].

In this work, we augment gaze with head movement for depth control, as that provides a hands-free method for 3D positioning. This has no obvious baseline, as prior gaze extensions for 3D input have been limited. As a point of reference, we instead compare HeadDepth with RayCursor, a fully controller-based method for raycasting with depth control [2].

2.2 Depth Control with Head Movement

Head movement controls the viewport in head-mounted displays but is also used for pointing [25, 48] and other gestural input [53]. Depth control by head has been considered for applications that require hands-free operation, for example, in surgery and robotic control [31, 35]. A range of works have considered translational head movement to control depth by leaning, stepping or “bobbing” back and forth, as this intuitively manipulates distance [30, 22, 20, 54]. Rotational movement provides a larger input range, specifically when seated. Head roll (tilt left and right) has been used to control the depth movement of a robotic arm as it avoids looking down or up from the task [49]. However, roll introduces directional ambiguity, making it difficult for users to intuitively determine in which coordinate system depth is modified, as it could be both the world-coordinate system or the head-coordinate system. We consider head pitch (tilt up and down) more naturally suited for depth control, as it does not disrupt directions, and many of our “heads down” interactions are closer up (e.g., in manual range) while “heads up” interaction is typically over larger distances.

Head pitch is supported by neck extension up to 55° , and flexion of about 45° [21]. Mapping of rotation to control can be absolute or relative. With absolute mapping, there is a fixed output value for any input, for example, for viewport control [18, 36] and head pointing [28]. Relative mapping, in contrast, adjusts the output by an offset and has, for example, been used for head pointing with dynamic gain [48, 10] and refinement of gaze input by head movement [15, 42, 13]. We consider absolute and relative mapping, with designs adapted for depth control integrated with gaze.

2.3 Coordinated Eye and Head Input

In 3D user interfaces, head input has been widely considered as a proxy for gaze [27], and the modalities have been contrasted in speed versus control they afford [3, 34]. A range of work has contributed techniques that integrate input from eye and head, with gaze as primary pointing modality and head movement to confirm [41, 43], refine [15, 42, 13] and manipulate [29, 43] gaze selections. These techniques utilise head rotational input over small ranges, whereas depth presents a large range to control.

A principal challenge for integrating the modalities is that eye and head are naturally coupled for directing visual attention [40]. We naturally move our head to support eye saccades, which can interfere with head movement for control and cause unintended input [12]. There are specific aspects of eye-head coordination to consider for extension of gaze pointing with head pitch for complementary control. Head rotation during a gaze fixation is compensated by stabilising eye movement (VOR) but affects gaze eccentricity, i.e. the eye-in-head angle. The eyes can rotate around 45°

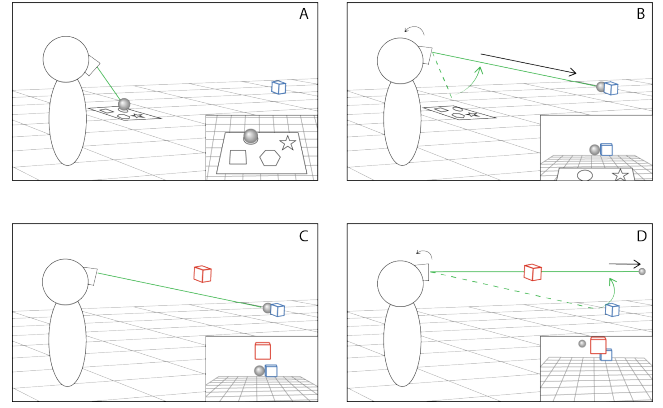


Figure 2: In this example, pitching the head up increases depth and vice versa. The user selects a sphere with the gaze from a tablet (A) and places it next to a blue cube at a distance (B). Aligning the gaze to the target position naturally pitches the head up, coincidentally moving the sphere towards the cube—a synergistic situation. When the user places the sphere next to the red cube (C→D), looking upward moves the sphere (again) away, hindering interaction—an antagonistic situation.

down relative to the head but only around 30° up, declining with age [19], but most gaze fixations occur naturally within $10\text{--}15^\circ$ from the head centre [43]. This leaves room for additional movement to eccentric angles, in other work used for explicit gaze control [9], and in this work to allow for head pitch relative to the line of sight. The interplay of gaze and head pitch is complex and warrants a more detailed analysis of the problem space as the basis for the design of control mappings.

3 PROBLEM ANALYSIS

Eye and head movements are coupled in 3D interaction [37, 40]. In considering head pitch for depth control along the gaze ray, we need to carefully consider interactions between the eye and the head. We identified the following two core problems for the design of depth mappings.

3.1 Synergistic and Antagonistic Movement Types

In 3D environments, we rotate eyes, head, and torso to bring targets into a comfortable viewing angle [40]. Head movement naturally supports eye saccades and can implicitly trigger depth changes if used for depth control.

Consider the example illustrated in Figure 2. We assume that in this example, pitching the head up moves the target away (and vice versa; cf. Section 2.2). Selecting a sphere from the hand-held tablet, the user may wish to place it next to a distant cube in mid-air (A→B). Looking towards the target location naturally involves pitching the head upward, implicitly triggering depth input and moving the sphere farther away. In this example, this behaviour aids the user because the natural head pitch moves the sphere closer to the target location. We call these situations *synergistic*. However, depending on the implementation and target location, it may hinder interaction, leading to *antagonistic* situations: In Figure 2 C→D, the user may want to move the sphere next to the red cube. When shifting the gaze to the target location, the head rotates (slightly) up. This moves the sphere away—the opposite of what the user wants, making depth control more difficult. A third situation is when the target location does not require the user to change the depth. The user can complete the task by moving only the eyes. For example, moving the sphere slightly lower from the red cube’s left side can be done with a simple eye gaze shift. We aim to design our map-

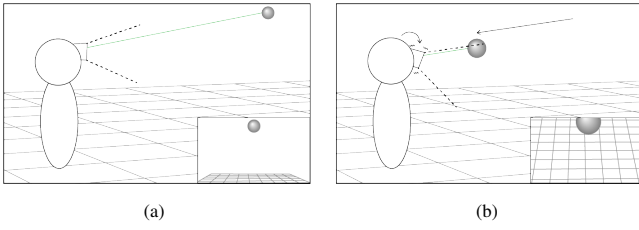


Figure 3: The user looks at the sphere in the upper field of view (3a). They rotate their head down to bring the sphere closer, leading to an extreme eye-in-head angle or even moving the sphere out of the eye tracker’s range (3b).

ping strategies to exploit (or at least be robust to) these synergistic, antagonistic and no-depth-change conditions.

3.2 Eccentric Viewing Angles during Gaze Fixation

To maintain a comfortable viewing experience, the head rotates during gaze pointing to keep the eye-in-head angle¹ within about 20 degrees, which is the comfortable eye-in-head angle for viewing [40]. Consequently, the range of head rotation that can be used for comfortable interaction is limited by the comfortable eye-in-head angle. This may become exacerbated if the eye-in-head angle is already large at the start of a depth change. Figure 3 illustrates such a scenario. The eye-in-head angle is already large after acquiring a target up far with gaze (Figure 3a). The user may want to bring the target closer by pitching the head downwards (cf. Figure 3b). Doing so will increase the eye-in-head angle, cause eye strain [40], decrease eye-tracking precision [11], or even move the sphere out of view. This problem becomes especially problematic for targets at eccentric viewing angles and large target amplitudes. Thus, when designing and evaluating head pitch for depth control, it is important to consider that the input range is limited by visual comfort. This strongly influences the pitch-to-depth mapping strategies.

3.3 Design and Study Objectives

Based on our problem analysis, we formulated the following objectives for our work:

- O1 Develop hands-free depth mapping strategies, providing competitive speed, accuracy, visual comfort, and user experience.
- O2 Investigate the influence of synergistic, antagonistic, and no-depth-change situations on performance, especially speed and accuracy.
- O3 Investigate the influence of target eccentricity/amplitude on performance, especially visual comfort.

4 DESIGN OF DEPTH MAPPINGS

To enhance gaze raycast, we designed depth mapping strategies that use head pitch rotation to control depth along the gaze ray. While eye gaze indicates the direction, the users can pitch their heads up and down to change depth (e.g., of a cursor or “grabbed” object) along the line of sight.

We designed three depth mapping strategies: RELATIVE, EDGE-GAIN, and PINGPONG. RELATIVE and EDGE-GAIN are implemented based on relative mapping, while PINGPONG is based on absolute mapping (cf. Section 2.2).

4.1 RELATIVE

We designed RELATIVE as a speed-sensitive mapping: Fast pitching allows quick travel across the depth range, while slow pitching preserves precision. With this, the user can travel a large distance

in depth with a small amount of pitch rotation to maintain a comfortable eye-in-head angle while being precise when they slow their head down for refinement.

RELATIVE calculates the new depth every frame by scaling the amount of head pitch rotation since the last frame with a gain factor:

$$depth_{new} = depth_{current} + G_{rel} \cdot \Delta pitch \quad (1)$$

where G_{rel} is the speed-sensitive gain function for scaling the change in head pitch rotation $\Delta pitch$.

We use the same speed-based linear interpolation function for G_{rel} as RayCursor for manual depth control [2]. In this model, a slow pitch gives a small gain factor for scaling; fast pitching outputs a large gain factor:

$$G_{rel} = \begin{cases} G_{max}, & \text{if } v_{pitch} \geq v_{max} \\ G_{min}, & \text{if } v_{pitch} < v_{min} \\ G_{min} + \frac{G_{max} - G_{min}}{v_{max} - v_{min}} (v_{pitch} - v_{min}), & \text{otherwise} \end{cases} \quad (2)$$

where v_{pitch} is head pitch speed, obtained as the magnitude derivative of the head pitch angle between two frames. G_{max} and G_{min} are the maximum and minimum gain factors that can be applied to the depth change. v_{max} and v_{min} are the corresponding speed values that the function reaches for G_{max} and G_{min} . Equation (2) outputs a maximum gain if the head moves faster than v_{max} (and vice versa). Otherwise, the gain is linearly interpolated between these two limits, with higher v_{pitch} resulting in a higher gain (and vice versa).

RELATIVE mitigates problems of visual comfort (e.g., during antagonistic situations as well as with eccentric starting angles) with two general strategies: First, it supports clutching for larger depth traversal through repeatedly pitching their head quickly in the desired direction, then slowly realigning their eye and head gaze. Second, users can use fast rotations to adjust the depth quickly towards the desired location—arriving before the eye-in-head angle becomes uncomfortable.

4.2 EDGE-GAIN

Our second mapping strategy, EDGE-GAIN, also uses relative mapping and is designed to avoid uncomfortable eye-in-head angles. It considers the current eye-in-head angle in the gain function to do this. When the eye-in-head angle exceeds a predefined angle, EDGE-GAIN scales up the gain, avoiding excessive angles. The closer the eye-in-head angle gets to the comfort threshold, the stronger the gain is scaled up. We implemented this by adding a F_{scale} to the gain function G_{rel} :

$$depth_{new} = depth_{current} + F_{scale} \cdot G_{rel} \cdot \Delta pitch. \quad (3)$$

$F_{scale} = 1$ if the eye-in-head angle is below a predefined eye-in-head angle. Above this angle, F_{scale} increases linearly as long as the eye-in-head angle increases. This behaviour is implemented with:

$$F_{scale} = \begin{cases} 1 + \frac{F_{thr} - 1}{\theta_{thr} - \frac{\theta_{thr}}{F_{thr}}} (\theta - \frac{\theta_{thr}}{F_{thr}}), & \text{if } \theta \geq \frac{\theta_{thr}}{F_{thr}} \text{ \& } \theta > \theta_{pre} \\ 1, & \text{else} \end{cases} \quad (4)$$

with

- θ is the current eye-in-head angle.
- θ_{thr} is the eye-in-head-angle’s comfort threshold.
- F_{thr} controls the strength of the scaling (tunable parameter).
- $\frac{\theta_{thr}}{F_{thr}}$ defines when scaling starts.
- θ_{pre} is the eye-in-head angle in the previous frame.

θ_{thr} , the eye-in-head angle’s comfort threshold, is not a fixed value but depends on the viewing direction [40]. It is lower and narrower

¹The angle between eye-forward and head-forward.

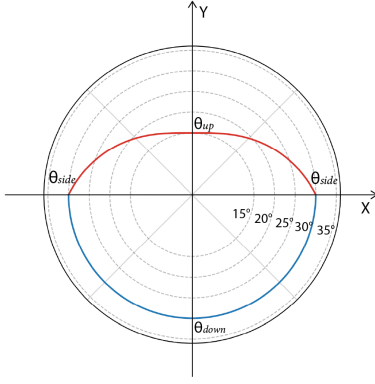


Figure 4: Visualization of the comfortable eye-in-head angle θ_{thr} in FOV, following Sidenmark et al. [40]. Gaze has a lower comfort threshold in the upper half compared to the lower half.

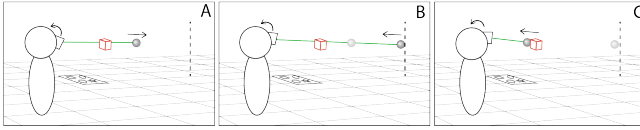


Figure 5: Illustration of PINGPONG. The user selects a sphere with the gaze from a tablet and places it next to the red cube. She could not pitch further down to bring the sphere closer due to visual comfort. Instead, she pitches up and moves the sphere away (A) until it reaches a limit (B). On the limit, the direction is reversed. The head tilts up, and the sphere moves in the opposite direction (C).

in the upper field of view than in the lower field of view, as illustrated in Figure 4. The effect of this asymmetric and viewing-direction-dependent value is that F_{scale} increases earlier when looking up than when looking down. θ_{thr} is defined as follows:

$$\theta_{thr} = \begin{cases} (1 - (1 - \frac{\theta_{up}}{\theta_{side}}) \cdot \sin(\tan^{-1}(\frac{y}{x}))) \cdot \theta_{side}, & y \geq 0 \\ (1 - (1 - \frac{\theta_{down}}{\theta_{side}}) \cdot \sin(\tan^{-1}(\frac{y}{x}))) \cdot \theta_{side}, & y < 0 \end{cases} \quad (5)$$

with

- x and y representing the eye-in-head angle in degrees ($x+$ is right and $y+$ is up).
- θ_{up} , θ_{down} are the y -axis' comfort angles (constants).
- θ_{side} is the x -axis' comfort angle (constant).

EDGEgain adjusts sensitivity based on the head pitch speed and eye-in-head angle to keep depth control comfortable and efficient (RELATIVE only uses the speed of head pitch). Like RELATIVE, users can clutch.

4.3 PINGPONG

PINGPONG is based on an absolute mapping, which means that the relationship between input and output is fixed and that clutching is neither required nor possible. The assumption is that this supports learnability and the users' recall of depth values, as the same head pitch angle consistently produces the same depth value. In a standard absolute mapping, each depth value would map to one head pitch angle. While this might be beneficial as users may recall angles and depth positions, this would prevent access to certain regions. For example, if the lowest pitch angle (looking far down) is mapped to the closest depth value, selecting the same close depth value with the maximum head pitch (looking far up) is impossible. To solve this problem and retain the absolute mapping, we designed PINGPONG. PINGPONG maps multiple head pitch angles to

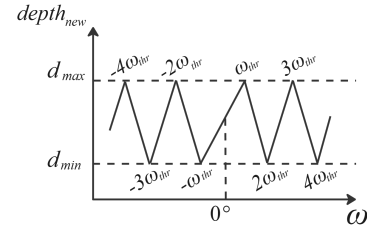


Figure 6: The transfer functions of PINGPONG. Pitching the head upward (ω increases) from the forward position ($\omega = 0^\circ$) moves the object farther away until it reaches the maximum depth limit (d_{max}) at a head pitch angle of ω_{thr} . If the head continues tilting upward beyond this point, the object comes closer until the change reverses at the minimum depth limit (d_{min}) at a head pitch angle of $2\omega_{thr}$.

the same depth value, illustrated in Figure 5. If a user keeps pitching their head upwards (A→B), depth increases before reaching the maximum, where it reverses direction (B→C; and vice versa).

Figure 6 shows the transfer function of PINGPONG. It is composed of a series of linear interpolations defined by the depth limits d_{min} and d_{max} , the current head pitch angle ω , and a head pitch threshold ω_{thr} (the head pitch angle at which the user reaches the depth limits). Outside $[-\omega_{thr}; \omega_{thr}]$, the gain is twice as high as inside this range to prevent excessive eye-in-head angles. The lower gain inside this range eases interaction in the centre of the FOV: When translating objects at eye-level, where many interactions happen, users can conveniently navigate across the full depth range, relying on fixed pitch-to-depth mappings, without uncomfortable eye-in-head angles and the need for clutching.

When designing the transfer function with ω_{thr} , d_{min} , and d_{max} , we had to ensure that the linear interpolations solved the original problem of the naive absolute mapping (only having access to limited areas) while providing comfortable eye-in-head angles and adequate sensitivity. Every depth position can be reached from several specific head angles, and the design ensures that at least one of these is within comfortable eye-in-head range. This is the case even in antagonistic situations or when the interaction starts from eccentric angles. Sensitivity for fine adjustments is higher, though, in the central range, where most interactions usually happen.

4.4 Summary

Using head pitch for depth control of gaze raycast is complex. The relationship between eye, head, and desired target location leads to edge cases (e.g., antagonistic, Section 3.1, and off-centre start location Section 3.2), and these can lead to uncomfortable eye-in-head angles, hindering interaction. Strategies mapping head pitch to depth must ensure that these angles are avoided (or minimized) while not impairing performance. We did this with RELATIVE, EDGEgain, and PINGPONG.

5 EVALUATION

We conducted a within-subject study to evaluate the mapping strategies regarding efficiency, effort, and visual comfort.

5.1 Task

The task was a sequential 3D docking task. Targets were semi-translucent spheres which were positioned at the corners of a cube [1, 47] (cf. Figure 7). Participants were instructed to move a small opaque green sphere (cursor) from one corner to another, following specific sequences. Figure 8 illustrates the sequences (in two directions). One individual corner-to-corner movement is a trial. Four trials make up a sequence. In a sequence, the start is the same as the end position. We included all corner-to-corner movements except

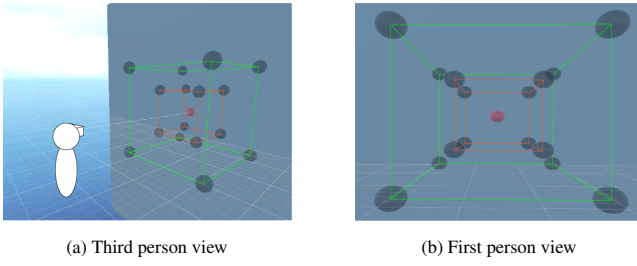


Figure 7: Targets (grey spheres) were positioned at the corners of a small and large cube (connecting lines only for illustration). A red sphere spawned at the cube’s centre, acting as a start button. During the task, only one sphere (red or grey) was presented to participants at a time.

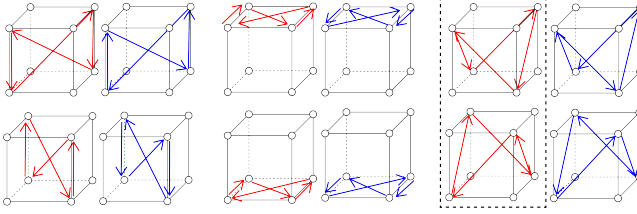


Figure 8: Sequences used in task illustrated in two directions (red and blue). Participants started from a random sequence node, and the rest followed the order shown by arrows. Sequences in the dashed box were also used for practice.

the purely horizontal movements, as they do not involve head pitch. Target positions were confirmed by pulling the controller trigger to minimize the effect of the confirmation method.

The task started at the cube’s centre, and participants had to align eye- and head-forward vector (and controller-forward vector if using the controller-based technique) with the red sphere and confirm with a button press on the controller. Next, one grey target sphere appeared at a random corner (the red sphere disappeared, and no other sphere was visible). The participant had to move the cursor into this new sphere and confirm with a button click. This click indicates the start of a sequence (cf. Figure 8). After a sequence, the cursor disappeared. The red sphere was in the centre again for the next task. Movement from and to the red sphere was not analysed.

When the cursor was inside the target sphere, the target sphere turned from grey to blue. A yellow fading line guided the user to the next target sphere. Pulling the trigger provided visual, auditory, and haptic feedback. If the cursor was inside the sphere, it turned green; otherwise, it turned red. Regardless, the next one was shown.

5.2 Experimental Design

5.2.1 Independent Variables

MappingStrategies (O1): RELATIVE, EDGE-GAIN, and PINGPONG as well as RAYCURSOR as reference (Baloup et al. [2]). RAYCURSOR is controller-based and requires manual input. As a standard technique, it serves to contextualise our contributions within a practical and established interaction paradigm, and also provides a reference for natural eye and head behaviour in 3D positioning. More details in Section 5.4.2.

MovementType (O2): Trials are grouped into subsets: AllSequences (all trials), Synergistic, Antagonistic, and NoDepthChange (shown in Figure 9).

CubeSize (O3): We tested with two cube sizes in the study to account for small and large amplitudes/eccentricity: The small cube has an edge length of 1m (targets 15° from the centre). The

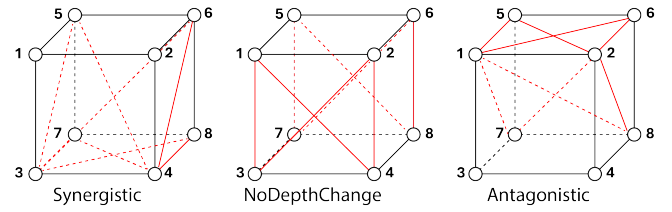


Figure 9: Synergistic, NoDepthChange, and Antagonistic trial sets (in red) grouped for the analysis of movement type.

large cube has an edge length of 2m (targets 35° from the centre). Both cubes are placed 3m in front of the user at eye level. We chose the large cube’s size so that all corners are barely visible when facing straight ahead, leading theoretically to the most extreme eye-in-head angles if the head stays still. The small cube size covered the central area of the field of view where many interactions happen.

5.2.2 Measures

The performance of mapping strategies was measured via movement time (speed), error rate (accuracy), cumulative head movement (effort), and maximum eye-in-head angle (visual comfort) for each trial. Movement time is the interval between confirming the previous target position and the current target position. Errors were counted when the cursor was outside the target 200 ms before confirmation to filter out late-trigger issues [14] and minimise the Heisenberg effect [51]. Cumulative head movement was the accumulated difference of the head vector measured via the HMD-forward vector between consecutive frames from the start to the end of a trial. Maximum eye-in-head angle was obtained from the maximum angular difference between head and eye vector per trial.

In our study, the reference mapping strategy, RAYCURSOR, inherently exerts minimal influence on head movement and maximum eye-in-head angle as it does not require eye and head input. Consequently, to analyse these two dependent variables, we have chosen to exclude RAYCURSOR from direct comparisons. We include its results in our reporting to ensure completeness and transparency.

For subjective metrics, the workload was measured with the RAW NASA-TLX questionnaire [5]. We also measured Ease (“The technique was easy to use”) and Comfort (“The technique was comfortable to use”) on a 7-Likert scale. Hand, neck, and eye fatigue were independently measured with the BORG CR10 scale [4].

5.3 Procedure

Upon arrival, participants were briefed on the study, completed a consent form, filled out a demographics questionnaire, and received a short presentation of the task.

Each participant completed four sessions (RELATIVE, EDGE-GAIN, PINGPONG, RAYCURSOR). Session order was counterbalanced using a balanced Latin square. The experimenter introduced and demonstrated the mapping strategy at the start of each session. Next, participants were seated in a fixed chair, fitted with a VR headset, calibrated eye-tracking and instructed to hold the controller in their dominant hand. They then started practising to familiarise themselves with the technique. The practice task was a shortened version of the experimental task, with two sequences (cf. Figure 8) for each cube size. If needed, they were allowed one additional round. After practice, participants proceeded with the task and were instructed to complete it as quickly and accurately as possible.

Each session included 24 sequences (12 for the large and 12 for the small cube). The order of sequences and start location was random. After each session, participants filled out post-session questionnaires, provided any comments, and took a short break before moving on to the next session. After completing four sessions, par-

Table 1: Target visual and world sizes for small and large cubes.

Target	Small Cube		Large Cube	
	Visual Size	World Size	Visual Size	World Size
Front Targets (1–4)	5.5°	0.25m	7°	0.3m
Back Targets (5–8)	4°	0.25m	4°	0.3m

Participants were asked to rank the mapping strategies and comment on their preferences. The study took around 90 minutes.

5.4 Implementation Details

5.4.1 Experiment Environment

Target sizes remained constant throughout the study. To focus on mapping strategies, we set the size of the back targets to 4° of visual angle, which is robust to gaze error [11], minimizing the impact of eye-tracking accuracy. The front targets were the same physical size as the back targets to ensure consistent perception of depth cues. Detailed target sizes are shown in Table 1. The red sphere was generated at the centre of the task cubes, 3 meters from the participants. It was placed at eye level and had the same world size as the target spheres. The green cursor controlled was 0.02m in diameter, equivalent to 0.38° visual degrees when placed at a distance of 3m away. For uniform contrast and depth cue, we placed a grey plane sized 22.5m × 60m (W × H), 6 meters away, fixed and centred at ground level. The directional light in the scene was off to prevent distractions.

5.4.2 Mapping Strategies

We experiment with interactions out of the reach of hands and within indoor spaces and thus used a depth range of 1 to 5 meters [26]. We applied 1€ Filters [7] with sampling frequency of 90Hz to head velocity ($f_{cmin} = 1$, $\beta = 0$) [48], gaze position and direction ($f_{cmin} = 0.05$, $\beta = 10$) [18] for input filtering in RELATIVE, EDGE GAIN, and PINGPONG. Pitching upwards increases the depth (and vice versa, cf. Section 2.2). No depth output of these mapping strategies exceeds 1 to 5 meters from gaze origin.

RELATIVE was implemented as described in Section 4.1, with: $v_{min} = 0.2$ rad/s, the threshold below which head movement is considered refinement [48]. $v_{max} = 0.6$ rad/s, set to three times v_{min} , consistent with the ratio between low and high gain states in RAYCURSOR [2]. $G_{max} = 0.4$ m/°, covering the full depth range (1 to 5 meters) with 10° of pitch rotation (half of the comfortable eye-in-head angle [40]). $G_{min} = \frac{0.4}{3}$ m/°, aligned with the ratio between v_{min} and v_{max} .

EDGE GAIN was implemented as described in Section 4.2, with $\theta_{up} = 15^\circ$, a conservative upper eye-in-head angle boundary for earlier activation of edge scaling. $\theta_{down} = 30^\circ$ and $\theta_{side} = 30^\circ$, liberal values to prevent excessive speed boosts in other directions. $F_{thr} = 3$, matching the three times ratio applied in RELATIVE. Other parameters were identical with RELATIVE.

PINGPONG was implemented as described in Section 4.3. The translation sensitivity of the outer linear functions was set to the high gain of RELATIVE ($\theta_{thr} = 10^\circ$). This ensured any position within the depth range could be reached with a maximum eye-in-head angle of 15°.

RAYCURSOR was implemented to be identical to the original work [2] (including the mapping, filtering, and corresponding parameters) without target assistance. The cursor distance from the controller was set to a minimum of zero with no maximum.

Table 2: Summary of repeated measures ANOVA on the dependent variables. Effect sizes are reported as η_p^2 for the error rate and η_g^2 for all other metrics. Significant results are highlighted in grey.

Variable	Effect	ANOVA		
		F value	p	Effect Size
Normalized Movement Time ^g	M x T x S	F (3.55, 53.22) = 1.337	.270	.006
	M x T	F (4.03, 60.46) = 3.468	.013	.023
	M x S	F (3, 45) = 1.776	.165	.008
	T x S	F (1.83, 27.42) = 2.022	.155	.003
	M	F (2.03, 30.46) = 24.388	<.001	.314
	T	F (1.97, 29.53) = 13.736	<.001	.038
	S	F (1, 15) = 64.285	<.001	.156
Error Rate ^p	M x T x S	F (9, 465) = 1.126	.342	.021
	M x T	F (9, 465) = 1.236	.271	.023
	M x S	F (3, 465) = 1.420	.236	.009
	T x S	F (3, 465) = 1.075	.359	.007
	M	F (3, 465) = 12.141	<.001	.073
	T	F (3, 465) = 2.138	.095	.014
	S	F (1, 465) = 6.393	.012	.014
Normalized Head Movement ^g	M x T x S	F (2.79, 41.79) = 0.531	.651	.001
	M x T	F (2.93, 43.94) = 1.046	.381	.002
	M x S	F (2, 30) = 3.393	.047	.006
	T x S	F (1.64, 24.63) = 2.491	.112	.002
	M	F (2, 30) = 3.703	.037	.037
	T	F (1.38, 20.77) = 8.118	.005	.022
	S	F (1, 15) = 265.850	<.001	.322
Max Eye-in-Head Angle ^g	M x T x S	F (2.60, 38.98) = 1.650	.199	.004
	M x T	F (2.55, 38.29) = 2.992	.05	.012
	M x S	F (1.42, 21.27) = 5.518	.019	.027
	T x S	F (1.27, 19.09) = 7.807	.008	.017
	M	F (2, 30) = 3.229	.054	.035
	T	F (1.25, 18.69) = .690	.448	.003
	S	F (1, 15) = 254.209	<.001	.805

M = MappingStrategy; T = MovementType; S = CubeSize

^p = η_p^2 ; ^g = η_g^2

5.4.3 Apparatus

We used an HTC VIVE Pro Eye VR headset for the study, with 110° diagonal FOV, 2880×1600 pixels resolution, and 90 Hz refresh rate. The study task was presented in a VR environment developed in Unity 2022.3.15 with the OpenVR XR Plugin 1.2.1 on a computer with an Intel Core i7-12700KF CPU, 32 GB RAM, and an NVIDIA GeForce RTX 4070 GPU.

5.5 Participants

16 participants (8 self-identified as male, 8 as female) aged 19–48 (M = 28.69, SD = 7.90) were recruited from our local university via flyers and word-of-mouth. 13 reported normal vision, 2 wore glasses, and 1 wore contact lenses. 6 reported no, 6 monthly, 2 weekly, and 2 daily playing video games. 2 reported no, 13 occasional, and 1 weekly experience with VR. 6 reported no, 9 occasional, and 1 weekly experience with an eye-tracking device. The university ethics committee approved the study. Participants received 10 GBP as compensation.

6 RESULT

We have 96 data points per technique, 384 per participant, totalling 6144. Outliers (151, 2.46%) were removed if the value deviated by more than 3 standard deviations from the grand mean, leaving 5993 data points. Movement time and cumulative head movement from 2D or 3D diagonal trials were normalized by $\sqrt{2}$ and $\sqrt{3}$.

We conducted repeated measures ANOVAs ($\alpha = .05$). When sphericity was violated (Mauchly’s test), Greenhouse-Geisser corrections were applied. Normality was validated using Shapiro-Wilk tests, histograms, and QQ plots. Bonferroni-corrected post-hoc tests were used as needed. If data was non-normal, we applied Aligned Rank Transform [50] with accompanying post-hoc tests. Subjective data were analyzed using Friedman tests and Bonferroni-corrected Wilcoxon signed-rank post-hoc tests. Table 2) contains all ANOVA results and test statistics. Raw data is available in the supplemental material.

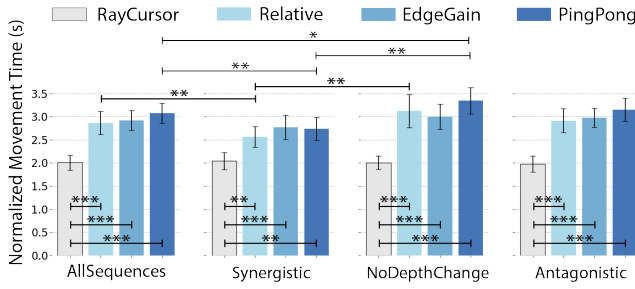


Figure 10: Normalized movement time of each mapping strategy at each level of movement type. Error bars represent the 95% CI.

6.1 Normalized Movement Time

Figure 10 illustrates normalized movement time by MappingStrategy and MovementType. A 3-way ANOVA did not find a significant 3-way interaction. We found a significant 2-way interaction between MappingStrategy $\times$ MovementType. All combinations of movement properties with RAYCURSOR were faster than other factor combinations (all $p \leq .002$). For both PINGPONG and RELATIVE, Synergistic movements were significantly faster than NoDepthChange and AllSequences (all $p \leq .031$). For PINGPONG, we found that NoDepthChange movements were also slower than AllSequences ($p = .025$). There were no significant differences between other factor combinations (all $p \geq .918$).

We found a significant main effect for MappingStrategy. While no significant difference was shown between RELATIVE ($M = 2.87$, 95% CI [2.73, 3.01]), EDGE GAIN ($M = 2.92$, 95% CI [2.80, 3.04]), and PINGPONG ($M = 3.08$, 95% CI [2.95, 3.21]; all $p \geq .347$), RAYCURSOR ($M = 2.01$, 95% CI [1.93, 2.09]) was faster than all three (all $p < .001$), independent of cube size or movement type. This is partially explained by the interaction effect of MappingStrategy $\times$ MovementType. However, the main effect of RAYCURSOR is strong as it consistently outperforms our mapping strategies.

We also found a significant main effect of MovementType. Synergistic movements ($M = 2.53$, 95% CI [2.40, 2.66]) were faster than Antagonistic ($M = 2.75$, 95% CI [2.62, 2.89]), NoDepthChange ($M = 2.87$, 95% CI [2.71, 3.03]), and AllSequences ($M = 2.72$, 95% CI [2.59, 2.84]; all $p \leq .048$). NoDepthChange movements were also slower than AllSequences (all $p \leq .020$). The interaction effect of MappingStrategy $\times$ MovementType can partially explain this main effect. We found a significant effect of CubeSize. Participants were faster when positioning within small cubes (2.45, 95% CI [2.36, 2.55]) compared to large cubes (2.98, 95% CI [2.89, 3.08]; $p < .001$).

6.2 Error Rate

We transformed the error rate with ART as it was not normally distributed. We conducted a 3-way ANOVA on the error rate. We did not find any significant interaction.

We found a significant effect of MappingStrategy. Post-hoc tests showed no significant difference between RELATIVE ($M = 3.17\%$, 95% CI [2.38%, 3.95%]), EDGE GAIN ($M = 3.73\%$, 95% CI [2.73%, 4.74%]), and PINGPONG ($M = 4.45\%$, 95% CI [3.13%, 5.77%]; all $p \geq .682$). RAYCURSOR ($M = 1.11\%$, 95% CI [0.66%, 1.56%]) was less error-prone than the other three (all $p < .001$).

In addition, we found a significant main effect of CubeSize. Small cubes (3.40%, 95% CI [2.77%, 4.03%]) were more error-prone than large cubes (2.83%, 95% CI [2.09%, 3.56%]; $p = .012$).

We found no other main or interaction effects ($p > .271$).

6.3 Normalized Head Movement

Figure 11 illustrates normalized head movement. We conducted a 3-way ANOVA on normalized head movement ($^{\circ}$) with only REL-

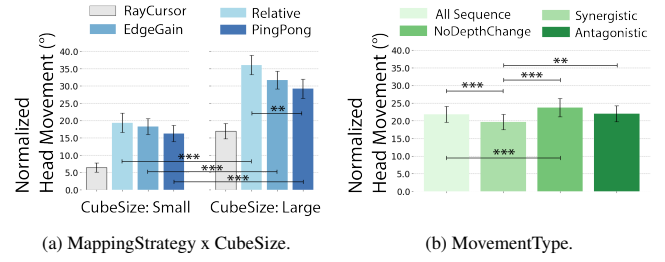


Figure 11: Normalized head movement of (a) each mapping strategy for each level of cube size and (b) each level of movement type. Error bars represent the 95% CI. Statistical analysis was only performed between RELATIVE, EDGE GAIN, and PINGPONG.

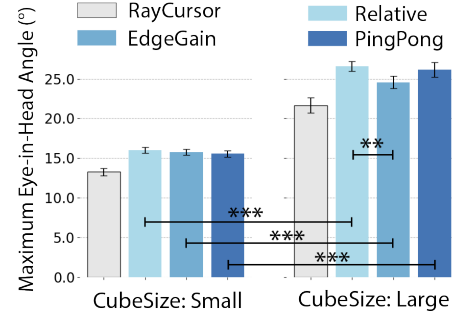


Figure 12: Max eye-in-head angle of each mapping strategy for each level of cube size. Error bars represent the 95% CI. Statistical analysis was only performed between RELATIVE, EDGE GAIN, and PINGPONG.

ATIVE, EDGE GAIN, and PINGPONG included (as outlined in Section 5.2.2). We did not find a significant 3-way interaction.

We found a significant 2-way interaction between MappingStrategy $\times$ CubeSize (Figure 11a). Post hoc analysis showed participants needed to rotate their heads more with RELATIVE than PINGPONG for the large cube ($p = .010$). In addition, all factor combinations with the small cube required significantly less head movement than those with the large cube (all $p < .001$).

We found a significant main effect of MappingStrategy. Post hoc tests revealed significant differences between all pairwise comparisons (all $p \leq .014$), with RELATIVE ($M = 27.72$, 95% CI [25.28, 30.16]) requiring the most, followed by EDGE GAIN ($M = 25.00$, 95% CI [22.91, 27.09]), with PINGPONG ($M = 22.75$, 95% CI [20.62, 24.87]) requiring the least amount of head movement. Note that the interaction effect partially explains this effect.

Furthermore, we found a significant main effect of MovementType (Figure 11b). Synergistic movements (19.65, 95% CI [17.51, 21.79]) required less head movement than AllSequences (21.79, 95% CI [19.54, 24.03]), NoDepthChange (23.72, 95% CI [21.10, 26.34]), and Antagonistic (22.00, 95% CI [19.76, 24.25]; all $p \leq .001$). NoDepthChange movements required a significantly higher amount of head movement than AllSequences ($p < .001$).

Finally, we found a significant main effect for CubeSize. Small cubes (15.10, 95% CI [13.81, 16.38]) required less head movement than large cubes (28.48, 95% CI [26.93, 30.04]; $p < .001$; partially explained by the interaction effect).

6.4 Max Eye-in-Head Angle

Figure 12 illustrates the maximum eye-in-head angle by mapping strategy and cube size. A three-way ANOVA was conducted on the max eye-in-head angle ($^{\circ}$) with only RELATIVE, EDGE GAIN, and PINGPONG included (cf. Section 5.2.2). We did not find a

Mapping Strategy	Overall TLX	Ease	Comfort	Hand Fatigue	Eye Fatigue	Neck Fatigue
RAYCURSOR	M = 5.86, 95% CI [4.39, 7.33]	M = 6.00, 95% CI [5.49, 6.51]	M = 5.44, 95% CI [4.90, 5.97]	M = 4.12, 95% CI [2.89, 5.36]	M = 1.81, 95% CI [0.88, 2.75]	M = 1.69, 95% CI [0.85, 2.52]
RELATIVE	M = 10.27, 95% CI [8.95, 11.58]	M = 3.88, 95% CI [3.26, 4.49]	M = 3.88, 95% CI [3.28, 4.47]	M = 0.81, 95% CI [0.40, 1.22]	M = 6.12, 95% CI [4.93, 7.32]	M = 5.00, 95% CI [3.70, 6.30]
EDGEgain	M = 10.78, 95% CI [9.07, 12.49]	M = 4.38, 95% CI [3.64, 5.11]	M = 4.12, 95% CI [3.29, 4.96]	M = 1.25, 95% CI [0.59, 1.91]	M = 6.12, 95% CI [4.71, 7.54]	M = 5.94, 95% CI [4.35, 7.52]
PINGPONG	M = 10.70, 95% CI [8.96, 12.43]	M = 3.56, 95% CI [2.76, 4.36]	M = 3.31, 95% CI [2.50, 4.13]	M = 1.12, 95% CI [0.59, 1.66]	M = 6.31, 95% CI [5.18, 7.44]	M = 5.06, 95% CI [3.68, 6.44]

Table 3: Mean values and 95% CIs of subjective metrics for each mapping strategy.

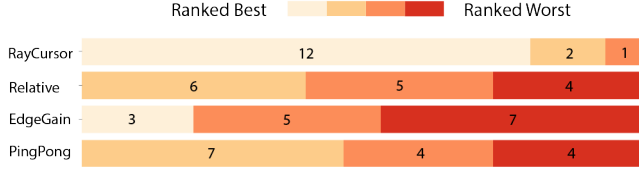


Figure 13: Preference ranking for each mapping strategy (N = 15).

significant 3-way interaction.

We found a significant two-way interaction between MappingStrategy $\times$ CubeSize. Pairwise comparison showed EDGEgain (24.56, 95% CI [23.78, 25.34]) required less eye-in-head angle than RELATIVE (26.58, 95% CI [25.97, 27.20]) for the large cube ($p = .002$).

We also found a significant two-way interaction between MovementType $\times$ CubeSize. However, we found no significant difference between movement types in the post hoc pairwise comparison.

We did not find a significant main effect on MappingStrategy (RELATIVE: M = 21.30, 95% CI [20.32, 22.29]; EDGEgain: M = 20.16, 95% CI [19.28, 21.04]; PINGPONG: M = 20.84, 95% CI [19.79, 21.89]). RAYCURSOR had an overall max eye-in-head angle of M = 17.47, 95% CI [16.57, 18.36].

We found a significant main effect of CubeSize. Small cubes (15.15, 95% CI [14.91, 15.40]) required less extreme eye-in-head angle than large cubes (24.73, 95% CI [24.26, 25.20]); $p < .001$.

6.5 Subjective Metrics

Table 3 lists subjective results. RAYCURSOR was consistently rated better than RELATIVE, EDGEgain, and PINGPONG (all $p \leq .026$). This includes task load (mental demand, performance, effort, frustration, and overall), neck fatigue, eye fatigue, ease, and comfort (all $p \leq .002$, $W \geq .318$). The only exception is comfort, where we did not detect a significant difference between RAYCURSOR and EDGEgain ($p = .142$). RAYCURSOR led to higher reported hand fatigue than the other three (all $p \leq .001$). We found no other significant difference (all $p \geq .142$).

Figure 13 shows post-study preference ranking (N = 15, as one participant's data was missing). Most participants (N = 12) ranked RAYCURSOR as the best. EDGEgain showed mixed preference. It was ranked best by 3 participants, although 7 participants ranked it lowest. RELATIVE and PINGPONG were appreciated by some (ranked second best by 6 and 7 respectively) but were not ranked first by any participant.

7 DISCUSSION

We investigated three head-pitch-to-depth mapping strategies (RELATIVE, EDGEgain, PINGPONG) and related them to the controller-based RAYCURSOR. Our results indicate that all HeadDepth mappings perform well for depth control. There were no outstanding differences between the mappings. As hand movement is known to have higher throughput than head movement [6], it was expected that the controller baseline would perform better in objective performance measures and subjective participant feedback. Given the nature of the techniques, RAYCURSOR has low eye and neck fatigue but high hand fatigue, whereas HeadDepth shows opposite results — suggesting further ergonomics optimizations. The

performance trade-off for gaining hands-free control was surprisingly low (around 1s slower, about 2.5% higher error rate, maximum eye-in-head angle difference less than around 5°). Only normalized head movement showed a stark difference up to 25°, which is no surprise as HeadDepth requires more head movement.

We aimed to develop hands-free competitive mapping strategies (O1), and our results suggest that head pitch for depth control is a viable alternative without major performance degradation when the hands are occupied (e.g., during manual work such as maintenance or repair jobs) or cannot be used otherwise (e.g., for people with injuries or disabilities). The benefit of HeadDepth is that it can be easily enabled in off-the-shelf AR/VR devices with eyetracking, such as the Meta Quest Pro or Pimax Crystal Light. With that, HeadDepth is a readily available solution for depth control for contexts in which the hands are busy and for accessibility.

Regarding our objective to investigate the influence of different movement types (O2), as expected, synergistic movements generally performed better than antagonistic movements. In detail, we found that the EDGEgain technique was the least affected by movement type in terms of movement time, while PINGPONG resulted in the least head movement and RELATIVE the most. This suggests that EDGEgain might be an interesting candidate to pursue further, despite PINGPONG leading to less head movement, as it might offer a consistent user experience without very challenging and uncomfortable (but also without very easy and convenient) situations. Interestingly, situations with NoDepthChange often perform the worst in movement time and head movement. Intuitively, participants should have been able to rely on gaze rotations in these situations without moving their head at all, as all targets were in view. However, participants still tended to move their heads in accordance with gaze direction, affecting pitch and thus depth control. This is explained by the close coordination of eye and head in gaze shifts, and natural support of eye saccades by head movement. Previous studies have suggested that VR may induce more head movement than naturally used for gaze [32], and also reported individual differences in head movement tendency [17].

Cube size significantly impacted all objective measures (O3). As per Fitts' Law, this was expected: Larger cubes required a larger travel distance between targets, which degraded movement time and increased cumulative head movement. Larger cubes also required more eccentric eye-in-head angles as targets were closer to the edge of the field of view. Smaller cubes led to more errors, which can be partially explained by the slightly smaller target size (Fitts' Law). Overall, larger target eccentricity and amplitude between targets negatively influenced the results, but these effects remained within acceptable limits.

7.1 Designing with and for HeadDepth

We outlined the mappings' configuration for our setup and task in Section 5.4. Developers and designers must balance comfort and available depth range when customizing them (and interaction techniques relying on them).

For simple relative mappings (as in RELATIVE), designers must balance how quickly users reach maximum gain. To cover larger depth ranges than our 1–5m, designers can either increase sensitivity, allowing large depth shifts with small head movements or keep sensitivity low and rely on clutching. The former is suitable for coarse tasks but challenging for precise adjustments. The latter en-

ables precise adjustments but requires more effort for larger depth traversals, e.g., via clutching. The optimal solution and customization depend on the use case.

A similar trade-off applies for mappings that dynamically scale up the gain factor. In EDGEgain, such a scale factor modulates gain. If gain-scaling starts at a smaller eye-in-head angle (i.e., earlier), the head pitch angular range where depth changes unscaled—and by that, with high precision—becomes smaller. Consequently, users change depth more often with a scaled gain. This offers less precision but allows them to make larger depth adjustments. Adjusting these two “zones” gives designers room to design for different depth ranges required by different use cases.

Customizing absolute mappings for different depth ranges requires adjusting the (linear) interpolation(s). In PINGPONG, this can be done by modifying the slope of the individual linear interpolations (they can have different slopes) and adding/removing repetitions. Extending this behaviour with cubic or bicubic interpolations could make sensitivity depend on head pitch and, in addition, head yaw. This is relevant as the range of head pitch changes with head yaw [44]. However, this might further complicate maintaining comfortable eye-in-head angles.

7.2 Implications for Transfer Functions

Our work implies that integrating the eye-in-head angle into the design may benefit many, if not all, head-rotation-based interaction techniques (as long as head gaze does not simply act as a proxy for eye gaze). In addition, our work implies that it can also be beneficial to consider the relationship between head rotation and eye-in-head angle when designing transfer functions for other interactions. Imagine using a controller to move a grabbed object in 3D, where a relative transfer function adjusts how the object moves. Here, gain during translation can be very high because the hand is good for fast and precise control. However, a gain that is too high could lead to objects moving too quickly towards uncomfortable eye-in-head angles or out of view. This would hinder interaction by requiring additional head movements. Integrating the eye-in-head angle into the transfer function to reduce the gain could support interaction.

7.3 Using HeadDepth for Interaction

Currently, HeadDepth is an evaluated means to an end for depth control of eye gaze. Designed as a target indication method [16], it still requires a confirmation mechanism, such as dwelling on a button press, for a complete interaction. Nonetheless, it holds potential for broader applications that involve control for depth or any continuous value. With gaze input, or just need visual feedback.

HeadDepth may support 3D selection in virtual environments by easing disambiguation during controller raycast selection of targets in dense or (partially) occluded environments. Integrating HeadDepth can offer an easy way to move the cursor behind a distracting or occluding target. Here, having a controller reduces the challenge of uncomfortable eye-in-head angles while integrating HeadDepth for depth control and dwell for confirmation avoids reliance on buttons and, by that, the Heisenberg effect [51].

While we evaluated HeadDepth in VR, the mappings promise value beyond that. For example, to control smart appliances such as light bulbs, thermostats, or speakers. Such devices could be activated by gaze, and HeadDepth can change values such as brightness, volume, or temperature. Although voice control is an alternative, design for eye-in-head comfort remains important as users need to maintain gaze on the UI to receive visual feedback.

Furthermore, HeadDepth can complement existing interaction techniques. Gaze+Pinch [33] offers easy and efficient 2D selection and confirmation. However, 3D mid-air positioning relies on hand translation, which can be tiresome (Gorilla Arm). Here, HeadDepth can complement Gaze+Pinch by allowing for controlling depth without hand translation.

7.4 Limitations

Participants reported visual distraction by the cursor, and we observed attention drawn from the target to follow the cursor with their eyes. This may have impacted performance, however, in the same manner across conditions.

We used a depth range of 1-5m. We expect effects observed to generalize to other ranges common in AR/VR applications. However, moving content close to the eyes in AR/VR can cause visual discomfort as it strains vergence and accommodation, and eye-head movement to control depth might add to discomfort. We combined amplitude and eccentricity via cube size in our study. This design enabled us to investigate the effect of movement type (driven by gaze shifts at different eccentricities) but prohibited insight into separate effects of amplitude and eccentricity.

Our study involved only 16 healthy, relatively young participants with a university background. While this is appropriate for early-stage evaluation, the limited and homogeneous sample restricts generalizability. Sensitivity analysis via G*Power showed we could reliably detect effects of $\eta^2 = .03$ or larger. Smaller effects would require a larger sample. These findings offer useful initial insights but should not be considered fully robust.

Finally, the head is effective for refining 2D eye gaze direction [42, 13]; using it simultaneously for depth control can introduce challenges, potentially undermining the precision and comfort it can offer during 2D gaze interactions.

8 CONCLUSION

This work is inspired by the ease and speed at which gaze can be used for 2D raycasting and provides an in-depth investigation into the extension of gaze raycast with head pitch for depth control. This includes: a problem analysis of interactions between eye and head rotational movements to consider in contexts where head movement controls input for gaze other than as a proxy; The design and evaluation of pitch-to-depth control mappings that demonstrate relative, absolute, and eye-in-head-angle adjusted solutions to control depth within a comfortable view; and empirical insights into synergistic and antagonistic effects of using head pitch for depth control, fully integrated with gaze.

We draw several key conclusions: (1) Integrating head-based depth control with gaze results in a robust, effective, and hands-free 3D positioning method. This is significant for control without input devices, contexts in which the hands are busy, and generally for accessibility. (2) The interaction model provides great simplicity from a user's perspective. It reduces a relatively complex task (control of movement in 3D) to a control movement in 1D along the line of sight. (3) The design behind the scenes is complex, though, as eye and head movements continuously interact: Gaze shifts affect head movement for control, and head movements affect eye-in-head viewing angles. Our results underscore the value of integrating head and gaze movements for more natural and effective 3D interactions.

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).

DIGITAL APPENDIX

We provide open-source code, a demo Unity project, a video figure, and the raw data for further experimentation: <https://doi.org/10.5281/zenodo.16679236>

REFERENCES

- [1] R. Balakrishnan and G. Kurtenbach. Exploring bimanual camera control and object manipulation in 3D graphics interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pp. 56–62. Association for Computing Machinery, New York, NY, USA, May 1999. (accessed 2024-07-04). doi: 10.1145/302979.302991 4
- [2] M. Baloup, T. Pietrzak, and G. Casiez. Raycursor: A 3d pointing facilitation technique based on raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300331 1, 2, 3, 5, 6
- [3] 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, pp. 1:1–1:9. ACM, New York, NY, USA, 2018. doi: 10.1145/3206343.3206349 1, 2
- [4] G. Borg. Psychophysical bases of perceived exertion. *Medicine and science in sports and exercise*, 14(5):377–381, 1982. 5
- [5] J. C. Byers, A. Bittner, 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. 5
- [6] S. K. Card, J. D. Mackinlay, and G. G. Robertson. A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst.*, 9(2):99–122, Apr. 1991. doi: 10.1145/123078.128726 8
- [7] G. Casiez, N. Roussel, and D. Vogel. 1 € filter: A simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2527–2530. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208639 6
- [8] D. L. Chen, M. Giordano, H. Benko, T. Grossman, and S. Santosa. 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*, VRST '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3611659.3615693 2
- [9] M. Choi, D. Sakamoto, and T. Ono. Kuiper belt: Utilizing the “out-of-natural angle” region in the eye-gaze interaction for virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517725 2
- [10] C.-L. Deng, L. Sun, C. Zhou, and S.-G. Kuai. Dual-gain mode of head-gaze interaction improves the efficiency of object positioning in a 3d virtual environment. *International Journal of Human-Computer Interaction*, 0(0):1–16, 2023. doi: 10.1080/10447318.2023.2223861 2
- [11] B. J. Hou, Y. Abdrabou, F. Weidner, and H. Gellersen. Unveiling variations: A comparative study of vr headsets regarding eye tracking volume, gaze accuracy, and precision. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 650–655, 2024. doi: 10.1109/VRW62533.2024.00127 3, 6
- [12] B. J. Hou, J. Newn, L. Sidenmark, A. Ahmad 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 1, 2
- [13] B. J. Hou, J. Newn, L. Sidenmark, A. A. Khan, and H. Gellersen. Gazeswitch: Automatic eye-head mode switching for optimised hands-free pointing. *Proc. ACM Hum.-Comput. Interact.*, 8(ETRA), may 2024. doi: 10.1145/3655601 2, 9
- [14] M. Kumar, J. Klingner, R. Puranik, T. Winograd, and A. Paepcke. Improving the accuracy of gaze input for interaction. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ETRA '08, p. 65–68. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1344471.1344488 5
- [15] 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*, pp. 1–14. ACM, Montreal QC Canada, Apr. 2018. doi: 10.1145/3173574.3173655 1, 2
- [16] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017. 9
- [17] H. S. Lee, F. Weidner, and H. Gellersen. Patterns in motion: On head- and non-head movers in vr during viewport. In *2024 IEEE Virtual Reality (VR)*, 2024. doi: 10.1109/VRW62533.2024.00125 8
- [18] H. S. Lee, F. Weidner, L. Sidenmark, and H. Gellersen. Snap, pursuit and gain: Virtual reality viewport control by gaze. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642838 2, 6
- [19] W. J. Lee, J. H. Kim, Y. U. Shin, S. Hwang, and H. W. Lim. Differences in eye movement range based on age and gaze direction. *Eye*, 33(7):1145–1151, 2019. doi: 10.1038/s41433-019-0376-4 1, 2
- [20] X. Liu, L. Wang, W. Ke, and S.-K. Im. Object manipulation based on the head manipulation space in VR. *International Journal of Human-Computer Studies*, 192:103346, Dec. 2024. doi: 10.1016/j.ijhcs.2024.103346 2
- [21] E. LoPresti, D. M. Brienza, J. Angelo, L. Gilbertson, and J. Sakai. Neck range of motion and use of computer head controls. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies*, Assets '00, p. 121–128. Association for Computing Machinery, New York, NY, USA, 2000. doi: 10.1145/354324.354352 2
- [22] X. Lu, D. Yu, H.-N. Liang, X. Feng, and W. Xu. DepthText: Leveraging Head Movements towards the Depth Dimension for Hands-free Text Entry in Mobile Virtual Reality Systems. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1060–1061. IEEE. doi: 10.1109/VR.2019.8797901 2
- [23] D. Mardanbegi, C. Clarke, and H. Gellersen. Monocular gaze depth estimation using the vestibulo-ocular reflex. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, ETRA '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3314111.3319822 2
- [24] D. Mardanbegi, T. Langlotz, and H. Gellersen. Resolving target ambiguity in 3d gaze interaction through vor depth estimation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300842 2
- [25] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. 38(1):21–45. doi: 10.1111/cgf.13390 2
- [26] Microsoft. Eye-gaze-based interaction - mixed reality, Mar 2023. [Accessed: 2024-07-12]. 6
- [27] M. R. Mine. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept*, 1995. 2
- [28] M. Nancel, O. Chapuis, E. Pietriga, X.-D. Yang, P. P. Irani, and M. Beaudouin-Lafon. High-precision pointing on large wall displays using small handheld devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 831–840. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2470654.2470773 2
- [29] T. Nukarinen, J. Kangas, O. Špakov, P. Isokoski, D. Akkil, J. Rantala, and R. Raisamo. Evaluation of HeadTurn: An Interaction Technique Using the Gaze and Head Turns. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, pp. 1–8. ACM. doi: 10.1145/2971485.2971490 2
- [30] J. Oh, K. Ando, S. Iizuka, L. Guinot, F. Kato, and H. Iwata. 3d head pointer: A manipulation method that enables the spatial localization for a wearable robot arm by head bobbing. In *2020 23rd International Symposium on Measurement and Control in Robotics (ISMCR)*, pp. 1–6. IEEE. doi: 10.1109/ISMCR51255.2020.9263775 2
- [31] K.-B. Park, S. H. Choi, J. Y. Lee, Y. Ghasemi, M. Mohammed, and H. Jeong. Hands-free human-robot interaction using multimodal gestures and deep learning in wearable mixed reality. 9:55448–55464. Conference Name: IEEE Access. doi: 10.1109/ACCESS.2021.3071364 2
- [32] K. Pfeil, E. M. Taranta, A. Kulshreshth, P. Wisniewski, and J. J. LaViola. A comparison of eye-head coordination between virtual and physical realities. In *Proceedings of the 15th ACM Symposium on Applied Perception*, SAP '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3225153.3225157 8

- [33] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 99–108. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132180 2, 9
- [34] 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
- [35] S. Sadri, S. A. Kohen, C. Elvezio, S. H. Sun, A. Grinshpoon, G. J. Loeb, N. Basu, and S. K. Feiner. Manipulating 3d anatomic models in augmented reality: Comparing a hands-free approach and a manual approach. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 93–102. IEEE. doi: 10.1109/ISMAR.2019.00-21 2
- [36] S. P. Sargunam, K. R. Moghadam, M. Suhail, and E. D. Ragan. Guided head rotation and amplified head rotation: Evaluating semi-natural travel and viewing techniques in virtual reality. In *2017 IEEE Virtual Reality (VR)*, pp. 19–28. IEEE. doi: 10.1109/VR.2017.7892227 2
- [37] L. Sidenmark. *Coordinated Eye and Head Movements for Gaze Interaction in 3D Environments*. PhD thesis, Lancaster University, 2023. doi: 10.17635/lancaster/thesis/1884 2
- [38] L. Sidenmark, C. Clarke, J. Newn, M. N. Lystbæk, K. Pfeuffer, and H. Gellersen. 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*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3580685 2
- [39] L. Sidenmark, C. Clarke, X. Zhang, J. Phu, and H. Gellersen. Outline Pursuits: Gaze-assisted Selection of Occluded Objects in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–13. Association for Computing Machinery. doi: 10.1145/3313831.3376438 2
- [40] 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, 3, 4, 6
- [41] 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
- [42] 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 2, 9
- [43] L. Sidenmark, D. Potts, B. Bapisch, and H. Gellersen. Radi-eye: Hands-free radial interfaces for 3d interaction using gaze-activated head-crossing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445697 2
- [44] R. A. Swinkels and I. E. Swinkels-Meewisse. Normal values for cervical range of motion. *Spine*, 39(5):362–367, 2014. doi: 10.1097/BRS.000000000000158 9
- [45] 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
- [46] U. Wagner, M. Albrecht, H. Wang, K. Pfeuffer, and H. Gellersen. Gaze, wall, and racket: Combining gaze and hand-controlled plane for 3d selection in virtual reality. pp. 1–24, June 2024. ISS '24 ACM Interactive Surfaces and Spaces, ACM ISS 2024; Conference date: 27-10-2024 Through 30-10-2024. 2
- [47] U. Wagner, A. Jacobsen, T. Feuchtner, H. Gellersen, and K. Pfeuffer. Eye-hand movement of objects in near space. pp. 1–13, Aug. 2024. UIST '24 - The 37th Annual ACM Symposium on User Interface Software and Technology, UIST '24; Conference date: 13-10-2024 Through 16-10-2024. 2, 4
- [48] H. Wang, L. Sidenmark, F. Weidner, J. Newn, and H. Gellersen. Headshift: Head pointing with dynamic control-display gain. *ACM Trans. Comput.-Hum. Interact.*, aug 2024. Just Accepted. doi: 10.1145/3689434 2, 6
- [49] M. R. Williams and R. F. Kirsch. Evaluation of head orientation and neck muscle EMG signals as three-dimensional command sources. 12(1):25. doi: 10.1186/s12984-015-0016-6 2
- [50] 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 6
- [51] D. Wolf, J. Gugenheimer, M. Combosch, and E. Rukzio. 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*, CHI '20, p. 1–10. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376876 5, 9
- [52] X. Xu, Y. He, Y. Ge, and Z. Zheng. Eyeexpand: A low-burden and accurate 3d object selection method with gaze and raycasting. In *Computer Graphics Forum*, p. e70144. Wiley Online Library, 2025. doi: 10.1111/cgf.70144 2
- [53] Y. Yan, C. Yu, X. Yi, and Y. Shi. Headgesture: Hands-free input approach leveraging head movements for hmd devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4), dec 2018. doi: 10.1145/3287076 2
- [54] D. Yu, H.-N. Liang, X. Lu, T. Zhang, and W. Xu. DepthMove: Leveraging Head Motions in the Depth Dimension to Interact with Virtual Reality Head-Worn Displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 103–114. doi: 10.1109/ISMAR.2019.00-20 2
- [55] C. Zhang, T. Chen, E. Shaffer, and E. Soltanaghai. Focusflow: 3d gaze-depth interaction in virtual reality leveraging active visual depth manipulation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3613904.3642589 2