

MagicPitch: 3D Object Manipulation with Gaze, Pinch, and Head Pitch

Haopeng Wang*
Lancaster University

Pavel Manakhov†
Lancaster University

Thorbjørn Mikkelsen‡
Aarhus University

Florian Weidner§
University of Glasgow

Ken Pfeuffer||
Aarhus University

Hans Gellersen||
Lancaster University
Aarhus University

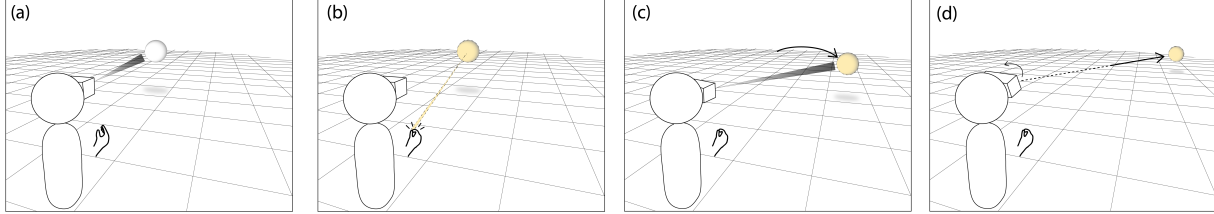


Figure 1: Object manipulation using MAGICPITCH. The user looks at the object (a) and then pinches to move and rotate it as with Gaze+Pinch (b). Shifting gaze away redirects the object (c). Tilting the head up or down translates the object forward or backward along the head-to-object axis (d). The yellow ray indicates the hand is actively moving the object, and the dark shading indicates the gaze ray (both are shown for illustration only).

ABSTRACT

Gaze has been integrated to assist 3D object manipulation with the hand for fast and effortless selection and radial movement, yet depth translation still largely depends on hand input. In this work, we investigate using gaze for radial redirection and head pitch for depth control together to complement 6DoF hand manipulation by introducing MAGICPITCH – a trimodal technique that integrates gaze, pinch, and head pitch. We conducted a 3D docking task (N=24) to evaluate MAGICPITCH alongside its variant, which scales the pitch-depth effect via hand movement to improve stability. Our results show that redistributing depth control to head pitch significantly reduces hand effort with a trade-off to slightly increased neck motion, highlighting the potential of head–hand cooperation for fatigue-aware multimodal XR interaction. Overall, the findings demonstrate the effectiveness of trimodal integration of gaze, hand, and head pitch input for 3D object manipulation and suggest a general principle of complementing 6DoF hand input with gaze and head input to offload manual effort.

Index Terms: Virtual reality, human-computer interaction, multimodal interaction, input modalities, 3D object manipulation

1 INTRODUCTION

Hand input is highly expressive for object manipulation in 3D space, but limited in reach. GAZE+PINCH uses gaze raycasting to extend the hand’s reach and enables users to acquire objects over any distance, by looking at the target while performing a pinch gesture [22, 23]. However, manipulation after acquisition remains limited by the hand’s movement range. Once an object has been pinched, it can only be moved within limits, and any larger movement necessitates clutching [12]. This is not ergonomically op-

timal in use cases such as repeated room-scale object placement, where large arm extensions are tiresome, or seated VR and public-transport interactions, where arm movement is constrained.

In this work, we propose MAGICPITCH to overcome this limitation by integrating coordinated gaze and head pitch with hand input. MAGICPITCH extends GAZE+PINCH by using gaze not only for acquisition but also for translation of objects across the 3D environment. However, as gaze does not easily allow for depth control, MAGICPITCH further integrates head pitch to move objects closer or farther in 3D space. With this technique, users can perform 6DoF manipulation at a relatively large scale while keeping the hand within a comfortable manipulation zone for reduced effort and input space.

MAGICPITCH is distinct in integrating input from the eye, head, and hand. It is built on two ideas from prior art. The first is MAGIC, a long-established technique for gaze-assisted manual input [49]. MAGIC exploits the fact that users naturally look toward the target of a pointing action and uses gaze to rapidly translate the pointer in cascaded input, where the hand only needs to refine the pointer position. The technique originates from 2D pointing but has been adopted to support object translation in 3D, albeit limited to movement of objects fixed in depth [48], or within near space for manual depth adjustment [39]. The second idea is depth control by head pitch, inspired by recent work that extended gaze with head movement for hands-free 3D positioning [42]. MAGICPITCH adopts head pitch but implements it differently to align with eye-hand coordination and unifies three modalities for 3D object manipulation.

We evaluated two variants of MAGICPITCH — one naive and one that prioritises hand over head input — in a user study using a 3D docking task (N=24) and compared them to MAGIC and GAZE+PINCH. Our results show that redistributing depth control to head pitch via MAGICPITCH reduces overall hand movement (at the cost of neck movement), compresses the necessary hand motor space, and is preferred over baselines. Task completion time is comparable, given sufficient training. This introduces a trade-off between head and hand effort, highlighting an important design consideration for balancing fatigue in multimodal XR interaction.

In sum, we contribute an empirical investigation of integrating eye, head, and hand simultaneously for 6DoF object manipulation in 3D. These results are significant beyond MAGICPITCH itself as an interaction technique, as they show that trimodal interaction is possible and that users can coordinate the eyes, head, and hand for

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

†e-mail: p.manakhov@lancaster.ac.uk

‡e-mail: thormik@cs.au.dk

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

¶e-mail: ken@cs.au.dk

||e-mail: h.gellersen@lancaster.ac.uk

3D interactions beyond pointing and selection [21]. The design and integration of trimodal input, along with the handling of input interference, provide insights into 3D input techniques. The principle of using eye and head input to form a complete 3D approach to offset and offload the effort of hand input in 3D is generalizable.

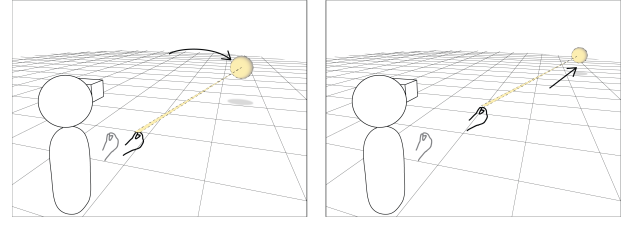
2 RELATED WORK

Gaze lends itself to assist manual input in 3D environments, as it extends reach [36], and is faster and less effort than the hands for raycasting [33, 40]. GAZE+PINCH has become the most widely adopted gaze-assisted technique, enabling fast acquisition of objects over any distance, by looking at the target and performing a pinch gesture [22, 23]. In GAZE+PINCH, the role of gaze is limited to assist selection, after which manipulation is solely by the hands. Several works have explored the extension of gaze-assistance object translation across surfaces [34, 35] and in 3D space [10, 11, 39, 48]. These approaches have their roots in the MAGIC concept of cascading input from gaze for coarse positioning to manual input for fine-grained control [49] but vary in how they integrate input from the two. In MAGICPITCH, we use a conservative approach of passing control to gaze only for larger saccadic movements, while control is returned to the hand as soon as a fixation is detected.

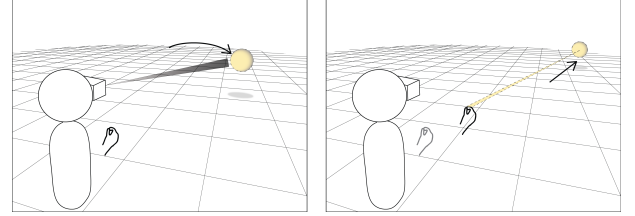
Based on MAGIC, gaze can support the movement of objects across the field of view (FoV), but not in distance from the user. Techniques exist that support inference and selection of different levels of focal depth from eye movement [16, 27, 50], but these are responsive to visual stimulus rather than capable of actively moving objects in depth. Methods that combine gaze and hand for 3D object manipulation thus fully rely on the hands for depth control. This is physically demanding as moving objects towards or away from the user requires arm extension and repeated clutching with a significant amount of hand movement [10, 15], undermining the concept’s core promise of minimal hand effort [22]. Depth can also be controlled through hand gestures [31], manual scaling of the environment [7] or bimanual strategies [1, 52], but these require separate actions that are not compatible with our aim to seamlessly extend GAZE+PINCH. For MAGICPITCH, we thus integrate head input to assist the hands in moving objects in depth.

Head movement affords stable control over large input ranges [9, 41]. In the context of HMD use, this range spans a comfortable 44° of flexion and 33° of extension [14]. It has also demonstrated reduced user effort and ergonomic advantages [32, 46], and has been widely combined with hand input as a means of reducing interaction effort [13, 18, 25]. Previous works have considered head pitch (up-and-down tilting) [42], head roll (lateral tilting) [43], and back-and-forth movement of the head to control depth [20, 47, 51]. Our work takes inspiration from HeadDepth, a recent work integrating head pitch with gaze raycasting to provide a hands-free method for 3D position control [42]. The work highlighted the natural mapping of pitch to depth, as we tend to tilt our heads down to look at objects that are nearer, and up when we look into the distance. However, gaze shifts make implicit use of head movement, to maintain the eyes in a comfortable eye-in-head position [28], which can interfere with depth control [42]. In MAGICPITCH, we avoid such interference by using head pitch only in hand mode and not during gaze-based translation of objects. Unlike for gaze, the input of head and hand is not cascaded but combined for depth control.

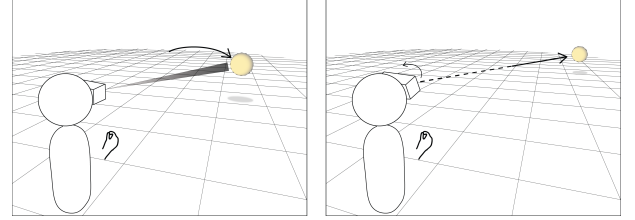
Our work is distinct in integrating input from three modalities. In prior works, eye, head, and hand have been compared as alternatives for pointing in 3D [2, 13, 24, 33], whereas we propose to combine them according to their relative strengths. Recent related work on integrating three modalities for pointing proposed the use of gaze for coarse, head for fine control, with hand only for confirmation of input [21]. In contrast, we integrate gaze and head for movement over larger ranges to preserve the precision and expressivity of the hand for object manipulation in 6DOF.



(a) User moving an object to the right and forward with GAZE+PINCH. The lateral and depth translation is performed by moving the hand to the right and forward, respectively.



(b) User moving an object to the right and forward with MAGIC. The lateral translation is performed by shifting gaze to the right. Depth translation is achieved by physically moving the hand forward.



(c) User moving an object to the right and forward with MAGICPITCH (and with MAGICMODPITCH). The lateral translation is performed by shifting gaze to the right. Depth translation is achieved by pitching the head upward. The dashed line shows the depth translation axis for head pitch, which is the direction of the object in the user’s view.

Figure 2: Example interaction sequence for translating an object to the right and forward in a large scale using GAZE+PINCH (Fig. 2a), MAGIC (Fig. 2b), and MAGICPITCH (Fig. 2c). MAGICMODPITCH behaves the same as MAGICPITCH, except the head-pitch contribution to depth could be scaled down by hand movement. The illustrated interactions happen after an object is selected by pinching (Fig. 1b). Note that in both MAGIC and MAGICPITCH, the user retains all the hand controls for GAZE+PINCH. The faint hand indicates the hand position at the time the pinch gesture was performed.

3 TECHNIQUE DESIGN AND IMPLEMENTATION

This section explains the design and build-up of MAGICPITCH and the integration across gaze, pinch, and head pitch. The design and implementation follow a cumulative approach, with techniques built on one another. Starting from the base GAZE+PINCH, we extend it with the MAGIC principle (MAGIC), then integrating head pitch for depth (MAGICPITCH), and finally a variation for managing the conflict between head and hand input (MAGICMODPITCH). Fig. 2 illustrates how different techniques support the movement of objects after being selected via pinching.

Following common GAZE+PINCH practice, we use the thumb tip as the hand input for both translation and rotation [4, 17, 23]; for brevity, we refer to the thumb tip as the “hand” hereafter. Across all of the techniques, hand rotation fully determines the object’s rotation. Translation, however, is handled differently. In GAZE+PINCH, it is determined solely by hand translation. MAGIC adds the ability to teleport the object in eye-centred radial coordinates (XY) using gaze. Both variants of MAGICPITCH also employ

Table 1: Comparison of interaction techniques by input modality and controlled DOF during manipulation.

Technique	Hand			Gaze	Head Pitch
	Trans. XY	Trans. Z	Rot.	Trans. XY	Trans. Z
GAZE+PINCH	✓	✓	✓		
MAGIC	✓	✓	✓	✓	
MAGICPITCH	✓	✓	✓	✓	✓
MAGICMODPITCH	✓	✓	✓	✓	✓*

XY = translation along the eye-centred radial surface.

Z = translation along the head-to-object axis.

* = Head-pitch contribution to Z is scaled by hand movement.

head pitch to control translation along the head-to-object axis (Z). Tab. 1 summarises the differences between the techniques in input modality and controlled degrees of freedom (DoF). Note that in all cases, translation is expressed in radial polar coordinates.

3.1 GAZE+PINCH

With GAZE+PINCH [23], the user first looks at the object with their eyes, performs a pinch gesture to acquire it (as illustrated in Fig. 1a-b). While holding the pinch, the user can move and rotate the object with their hand. Fig. 2a shows a user moving an object with GAZE+PINCH after selection. To move the object to the right, the user moves the hand towards the right. To move the object forward, the user stretches the arm and moves the hand forward.

Our implementation follows the latest published version of GAZE+PINCH as described in [17]. The method applies a distance-based hand amplification, scaling hand translation by the ratio between the eye-to-object distance and the eye-to-hand distance. Object rotation is mapped isomorphically to hand rotation. Releasing the pinch disengages hand input, allowing users to clutch. Full implementation details are provided in Sec. A.1.

3.2 MAGIC

MAGIC is built on GAZE+PINCH and inherits all the functions. Following the MAGIC principle, while holding the pinch to move and rotate the object, users can look away to snap the object to the gaze direction while preserving its distance. As shown in Fig. 2b, to move the object to the right, the user simply looks right without needing to move the hand. To move the object forward, the user stretches the arm and moves the hand forward.

The gaze and hand input are cascaded for directional (XY) change of the object position and integrated via mode switching. In hand mode, the object is fully controlled by the hand and behaves identically to GAZE+PINCH. In gaze mode, the object snaps to and follows the gaze ray, with hand input for direction ignored to avoid unintended offsets (e.g., when users move their hand while shifting gaze). In both modes, depth is always controlled by the hand, and the object rotation always follows the hand rotation isomorphically. Pinch release ends manipulation.

Mode switching is based on both the distance between the gaze and the object, and gaze saccade and fixation behaviour. In addition to prior distance-based approaches (referred to as “circular safe region” in [34, 39, 48], we also incorporate gaze saccade and fixation behaviour into the mode-switching criteria to improve robustness. Having both as mode-switching criteria prevents unintended switches caused by elicited saccades during smooth pursuit of rapid object movements with amplified hand input, and avoids disruptive multi-snapping behaviour during large saccades. The hand-to-gaze mode switch occurs upon detecting a saccade and when the gaze point is sufficiently separated from the object. The gaze-to-hand mode switch happens when a fixation is detected. Full implementation details are provided in Sec. A.2.

3.3 MAGICPITCH

The implementation of MAGICPITCH is based on MAGIC and inherits all the functions. While holding the pinch, the user can move and rotate the object, and look away to snap the object to gaze. The user can also tilt their head up and down for moving the object back and forth (up for further and down for closer [42]). As shown in Fig. 2c, to move the object to the right, the user simply looks right, without needing to move the hand (as with MAGIC). To move the object forward, the user tilts their head up, without needing to stretch the arm. Head-pitch depth translation is continuously applied along the head-to-object axis.

MAGICPITCH inherits the gaze/hand mode switching component from MAGIC, with an additional component of head pitch depth control that is always activated in hand mode. This means that in hand mode, while the object is moved by the hand, it can be translated in depth by head pitch. Conversely, in gaze mode, head pitch input does not affect the object to avoid unintended depth input during saccade [42]. The object rotation always follows the hand rotation isomorphically. Pinch release ends manipulation, which allows clutching for both hand and head pitch.

Depth control was adapted from EdgeGain [42], which is a relative pitch-to-depth mapping technique that responds to head pitch speed and the current eye-in-head angle. Faster head pitch produces a larger depth translation (up to 0.8m per degree of pitch rotation), and vice versa; a larger eye-in-head angle, reflecting an eccentric object position in the field of view, further scales the translation. The mapping was tuned so that a 10m depth translation could be achieved with roughly 12° of fast pitch movement (even less when the object is eccentric). The adopted depth control algorithm and its full implementation are detailed in Sec. A.3.

3.4 MAGICMODPITCH

User naturally rotate their head to maintain neck and visual comfort. Using head pitch for depth control can lead to uncomfortable neck postures and viewing angles [42], causing users to implicitly rotate their head to restore comfort during hand manipulation. This unintended head rotation would trigger the depth changes, disrupting the hand manipulation. The problem is exacerbated when high pitch-to-depth gains are used for efficiency, as they would produce larger depth shifts that are hard to correct with the hand.

To address this, we experiment with a variation of MAGICPITCH and build MAGICMODPITCH with a modulation to the depth control. MAGICMODPITCH behaves identically to MAGICPITCH, except when hand and head-pitch movements result in depth translations of opposite directions (whether judged in full 3D space or in their transverse-plane projections), the technique prioritises hand contribution. This means that the effect of head pitch decreases as hand speed increases, until the pitch-depth offset is reduced to zero at a hand speed of 0.1 m/s. This ensures that the hand control is prioritised during the interaction, so that the user can easily compensate for any large depth offset created by the implicit head movement with their hand. Implementation details are provided in Sec. A.4.

4 EVALUATION

To better understand the integration of gaze, head, and hand inputs for 3D object manipulation and assess the techniques’ efficiency (speed and effort) as well as overall user experience, we conducted an experiment involving a distant-object 3D docking task. In this study, we asked participants to pick up a virtual cube and move it across the scene to a target position indicated via a ghost image of the same cube (Fig. 3). The experiment employed a within-subject 4 Technique $\times$ 3 Depth $\times$ 3 Amplitude factorial design (detailed in Sec. 4.2).

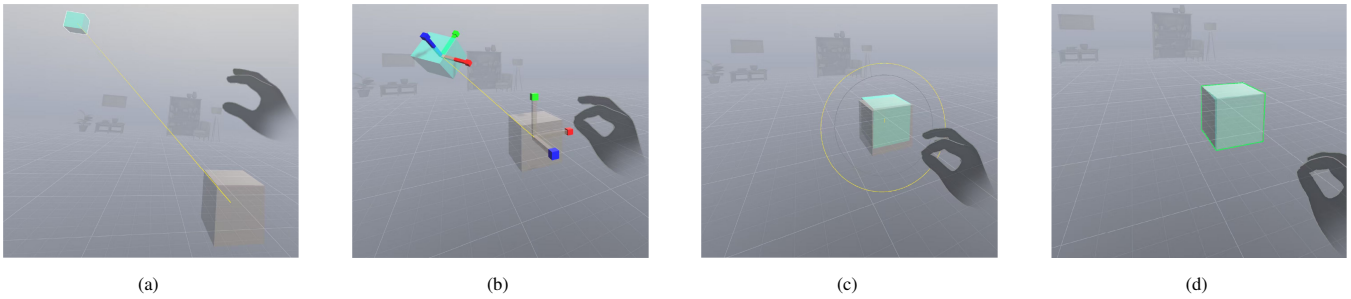


Figure 3: Task procedure and visual feedback from a first-person perspective. At the trial’s start, a cyan cube and a grey target spawn, connected by a yellow line (a). The participant starts the trial by looking at the cube, which then becomes outlined in white (a), and performs a pinch gesture (not shown in the figure) to move it toward the target. During manipulation, if orientation misalignment occurs, virtual gizmos appear (b). When the cube is near the target but its depth is misaligned, concentric rings (yellow and grey) appear to aid depth perception (c). Once position and orientation are within thresholds, a green outline appears (d). The participant then releases the pinch to finish the trial.

4.1 Task

To investigate 3D object manipulation, we devised a 3D docking task [10, 37, 48]. At the start of the trial, a cyan cube and its grey replica were positioned at the initial and target locations. Participants docked the cyan cube into the grey target using the assigned interaction technique, aligning both position and orientation. Positional tolerance was set to $\pm 1.5^\circ$ of visual angle [40] (equivalent to 0.05m at 2m, 0.10m at 4m, 0.16m at 6m, and 0.26m at 10m), with a rotational tolerance of $\pm 10^\circ$ [10]. To minimise the Heisenberg effect’s impact on the results [45], a trial was considered complete if the docking criteria were met at the frame of pinch release or at 200ms prior. Successful docking was confirmed with visual and audio feedback, after which participants were presented with the next trial. Participants were allowed to release the pinch for clutching during the trial.

Additional visual feedback was introduced during the task to minimise confounding factors such as target search and determining the correct translation and rotation direction. To guide the spatial relationship between the cube and the target, a thin yellow line continuously connected the cubes (Fig. 3a and Fig. 3b). Although no initial orientation offset was imposed, hand translation of the object could cause rotational changes [17]. To support this adjustment, gizmos (green, red, and blue axes extending from cube centres) appeared when orientation was misaligned (Fig. 3b). To aid depth perception — particularly for distant targets — two concentric rings centred on the target indicated relative depth (Fig. 3c). A static grey ring marked the target depth, while a dynamic yellow ring scaled with the cube’s depth relative to the target. Rings appeared within a depth distance of four times the position threshold, ensuring they supported depth perception rather than turning the task into simple ring matching. A green outline indicated successful alignment in both translation and rotation (Fig. 3d).

Trial start and end positions were always located at the corners of a frustum (Fig. 4) centred at the participant’s eye level. Each trial involved moving the cyan cube diagonally from one corner to the opposite, creating four diagonals, each with two directions (close-to-far and vice versa). This resulted in eight trials per frustum. Frustum size varied based on the current *Depth* and *Amplitude*.

4.2 Independent Variables

Technique – G+P, MAGIC, MAGICPITCH, and MAGICMOD-PITCH— represented the studied interaction techniques, as described in Sec. 3. G+P is short for GAZE+PINCH.

Depth – 2m, 4m, and 8m – represented the translation distances between the object’s start and target positions, measured along the line of sight (Fig. 5). At 2m, participants moved the target between the absolute depths of 2m and 4m in both directions. At 4m, the

range expanded to 2m-6m, and at 8m, to 2m-10m. These levels were chosen to reflect realistic room-scale interaction ranges [10].

Amplitude – 30° , 60° , 120° – represented radial translation distances, measured as a circular arc between the object’s start and target positions (Fig. 5). Levels included 30° for minimal radial displacement, 60° for medium displacement with start and end positions at the edges of the headset’s FoV [42], and 120° for large displacement and increased eccentricity, where the start and end positions were outside the FoV.

To control for order effects, we counterbalanced the presentation of *Techniques* using a balanced Latin square design. Each technique was presented in 9 blocks, each corresponding to a unique frustum defined by the combination of 3 *Depths* x 3 *Amplitudes*, with each block containing 8 trials. This resulted in a total of 72 trials per technique. The orders of the frustums and trials within each frustum were randomised.

4.3 Study Setup

We used a Meta Quest Pro ($106^\circ \times 96^\circ$, 90 Hz) with Quest Link running on a computer with an Intel Core i7-12700KF CPU, 32 GB RAM, and an NVIDIA GeForce RTX 4070 GPU. The prototype was developed in Unity 6000.0.44f1 with the OculusVR toolkit v4.5.1. The virtual environment is detailed in the appendix (Sec. B).

4.4 Participants

24 participants (9 self-identified as male, 15 as female) aged 22-60 ($M = 34.42$, $SD = 9.40$) were recruited from our local university via flyers and word-of-mouth. Most participants were right-handed (1 reported left-hand dominance). Visual acuity was normal in 10 participants; 11 wore glasses, and 3 wore contact lenses. Most participants reported playing video games monthly (14), with occasional (less than monthly) VR (21) and eye tracking (16) experience.

4.5 Procedure

Upon arrival, participants received an information sheet, completed a consent form and a demographic questionnaire, as well as received a brief task introduction. Before each technique, the experimenter explained the interaction method. Participants were seated in a fixed chair without armrests and fitted with a VR headset. Eye tracking calibration was performed before each technique. Each technique began with familiarisation and practice runs. During familiarisation, participants freely explored the technique by moving a cyan cube. Practice involved completing a block of 8 trials with *Amplitude* = 60° and *Depth* = 4m (the mid level for both factors).

Following practice, participants proceeded to the task and were instructed to complete it as quickly as possible, with short breaks allowed between trials while wearing the headset. After completing

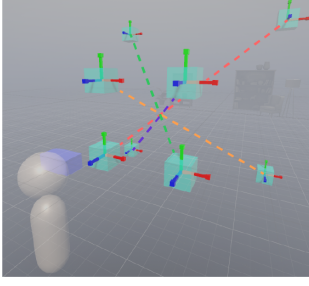


Figure 4: Task positions defined by a given amplitude-depth pair form a frustum. Docking trajectories ran diagonally from one corner to the opposite (four shown, different colours), with both directions resulting in 8 trials per frustum.

all trials for a technique, participants removed the headset, completed a post-task questionnaire, shared any comments, and took a short break before starting the next technique. Finally, after all four techniques were completed, participants ranked them and discussed their preferences.

The study was approved by the university ethics committee and took approximately 90 minutes. Participants received a 15 GBP voucher as compensation.

4.6 Measures

We measured completion time for speed, cumulative hand translation and hand rotation for hand effort, and head rotation for neck effort for each trial as objective measures (their definitions can be found in Sec. 5). Subjective measures included taskload via the Raw NASA-TLX [5], and hand, neck, and eye fatigue via the BORG CR10 [3]. We also included 7-point Likert questions for ease of use (“The technique was easy to use”) and comfort (“The technique was comfortable to use”).

5 RESULTS

We collected 72 data points per technique, 288 per participant, yielding a total of 6,912 data points. Outlier trials were systematically identified based on three criteria: tracking loss (pinch-tip displacement exceeding 0.5m within a single frame; 399 trials, 5.77%), erratic manipulation (hand rotations exceeding 4,000° within a trial, approximately Q3 + 5 × IQR; 81 trials, 1.17%), and instruction violations (participant taking a break during a trial—releasing the object for more than 5s after initial pickup; 31 trials, 0.45%). In total, 452 trials (6.53%) were excluded, leaving 6,460 valid trials for analysis. The proportion of excluded trials was normally distributed across participants (Shapiro–Wilk test, $p = .357$; further validated via histograms and Q-Q plots), suggesting that outlier removal did not introduce systematic participant-level bias. After outlier filtering, data were aggregated by computing participant-level means for each unique combination of independent variables, yielding 864 observations for statistical analysis.

We conducted 3-way repeated measures ANOVAs ($\alpha = .05$) with Technique × Depth × Amplitude as independent variables (cf. Table 2). When sphericity was violated (Mauchly’s test), Greenhouse-Geisser corrections were applied. Normality was validated using Shapiro-Wilk tests, histograms, and Q-Q plots. Aligned Rank Transform [44] was applied when the data was not normally distributed. Bonferroni-corrected post-hoc tests were used as needed. Subjective data were analysed with Friedman tests and Bonferroni-corrected Wilcoxon signed-rank post-hoc tests.

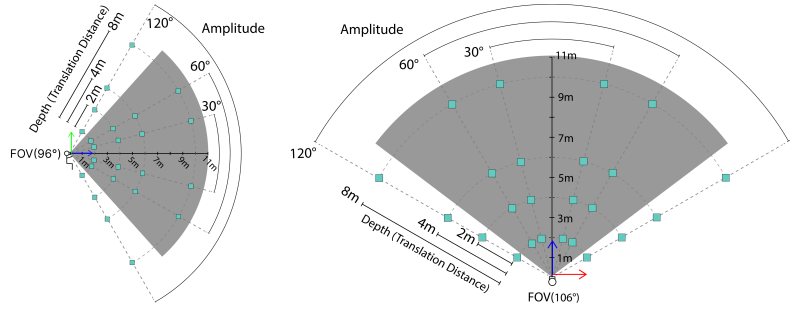


Figure 5: Detailed side (left) and top (right) views of all task positions for each depth-amplitude pair. The cyan squares represent the task positions. The grey region indicates the headset’s FOV. The red, green, and blue arrows indicate the world x-, y-, and z-axis, respectively.

Table 2: Summary of repeated measures ANOVA on the dependent variables. Significant results are highlighted in grey.

Variable	ANOVA			
	Effect	F value	p	η_p^2
Completion Time	T × D × A	F (12, 805) = .653	.797	.010
	T × D	F (6, 805) = 2.318	.032	.017
	T × A	F (6, 805) = .266	.953	.002
	D × A	F (4, 805) = .589	.671	.003
	T	F (3, 805) = 16.029	<.001	.056
	D	F (2, 805) = 107.246	<.001	.210
	A	F (2, 805) = 273.964	<.001	.405
Hand Movement	T × D × A	F (12, 805) = .232	.998	.003
	T × D	F (6, 805) = 13.825	<.001	.093
	T × A	F (6, 805) = 4.123	<.001	.030
	D × A	F (4, 805) = .668	.615	.003
	T	F (3, 805) = 109.198	<.001	.289
	D	F (2, 805) = 72.242	<.001	.152
	A	F (2, 805) = 454.864	<.001	.529
Hand Rotation	T × D × A	F (12, 805) = .183	.999	.003
	T × D	F (6, 805) = 6.008	<.001	.043
	T × A	F (6, 805) = 2.241	.038	.016
	D × A	F (4, 805) = .417	.796	.002
	T	F (3, 805) = 51.040	<.001	.160
	D	F (2, 805) = 38.429	<.001	.087
	A	F (2, 805) = 417.054	<.001	.509
Neck Rotation	T × D × A	F (12, 805) = 1.285	.222	.019
	T × D	F (6, 805) = 1.031	.403	.008
	T × A	F (6, 805) = .470	.830	.003
	D × A	F (4, 805) = 1.914	.106	.009
	T	F (3, 805) = 114.701	<.001	.299
	D	F (2, 805) = 62.657	<.001	.135
	A	F (2, 805) = 1697.153	<.001	.808

T = Technique; D = Depth; A = Amplitude.

5.1 Completion Time

Fig. 6 illustrates trial completion time in seconds (the time interval between the first-time object pick-up and the trial completion).

We did not find a significant 3-way interaction. However, we found a significant 2-way interaction for Technique × Depth. Comparing between the techniques, G+P and MAGIC were faster than MAGICPITCH and MAGICMODPITCH for 2m translations in depth (all $p < .001$), with no significant differences shown within each pair (all $p = 1$). For 4m depth translations, this advantage remained for G+P (all $p \leq .006$), but not for MAGIC (all $p \geq .612$). For 8m depth translations, no significant differences were observed across techniques (all $p = 1$). Within techniques, 8m depth translations were significantly slower than 2m and 4m for G+P, MAGIC, MAGICMODPITCH, with MAGIC also slower with 4m than 2m (all $p \leq .033$). No significant difference was found between depths for

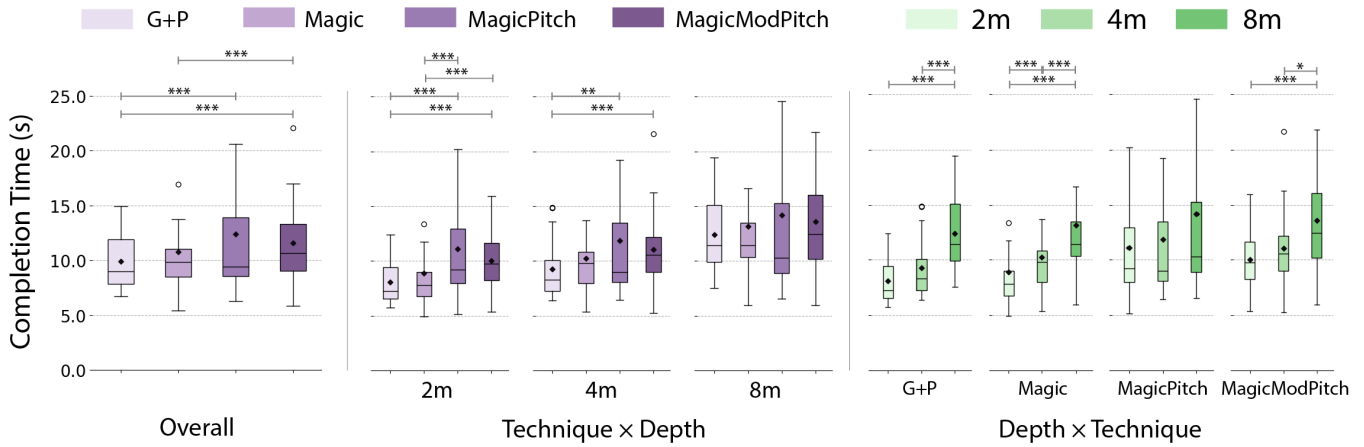


Figure 6: Completion Time of the Technique main effect and the interaction effect with Depth translation distance. Mean values are represented by black diamonds. Medians are represented by horizontal lines.

MAGICPITCH (all $p \geq .066$). No significant 2-way interaction for Technique $\times$ Amplitude was found.

We found a significant main effect of Technique. MAGICMODPITCH was slower than both G+P and MAGIC (all $p < .001$), while MAGICPITCH was slower only against G+P ($p < .001$). This is partially explained by the Technique $\times$ Depth interaction effect.

5.2 Hand Translation

Fig. 7 illustrates hand translation in meters (the cumulative Euclidean distance between the thumb tip positions in successive frames throughout the duration of a trial).

We did not find a significant 3-way interaction. However, we found a significant 2-way interaction for Technique $\times$ Depth. Comparing between the techniques, MAGICPITCH and MAGICMODPITCH consistently required less hand translation than G+P across all depth conditions (all $p \leq .020$). They also required less hand translation than MAGIC for 4 m and 8 m depth translation distances (all $p < .001$), but not for 2m (all $p = 1$). No significant difference was found between MAGICPITCH and MAGICMODPITCH (all $p = 1$), and between G+P and MAGIC (all $p = 1$). Within techniques, we found larger depth translations led to significantly more hand translation for G+P and MAGIC (all $p \leq .029$), but not for MAGICPITCH and MAGICMODPITCH (all $p = 1$).

We also found a significant 2-way interaction for Technique $\times$ Amplitude (see Fig. 17 in the appendix for details). Between the techniques, MAGICPITCH and MAGICMODPITCH consistently required less hand translation than both G+P and MAGIC across all amplitude conditions (all $p < .001$). Within all techniques, larger amplitudes led to greater hand translations (all $p \leq .006$).

We found a significant main effect on Technique. MAGICPITCH and MAGICMODPITCH required less hand translation than both G+P and MAGIC (all $p < .001$; explained by the interaction effects). MAGIC had slightly less hand translation against G+P ($p = .017$).

5.3 Hand Rotation

Fig. 8 illustrates hand rotation in degrees (the cumulative angular difference between the thumb tip orientations in successive frames throughout the duration of a trial).

We did not find a significant 3-way interaction. However, we found a significant 2-way interaction for Technique $\times$ Depth. Comparing between the techniques, with 4m and 8m depth translations, MAGICPITCH and MAGICMODPITCH required less hand rotation than both G+P and MAGIC (all $p < .001$). During the 2m depth translations, MAGICMODPITCH also required less hand rotation compared to G+P ($p = .010$). No significant difference was found

between MAGICPITCH and MAGICMODPITCH (all $p = 1$), and between G+P and MAGIC (all $p = 1$). Within techniques, both G+P and MAGIC required significantly more hand rotation for greater depth translations, as opposed to MAGICPITCH and MAGICMODPITCH, which maintained consistent hand rotation (all $p = 1$).

We also found a significant 2-way interaction for Technique $\times$ Amplitude (see Fig. 18 in the appendix for details). MAGICPITCH and MAGICMODPITCH consistently required less hand rotation compared to both G+P and MAGIC across all amplitude conditions (all $p \leq .012$). Within all techniques, larger amplitudes led to significantly greater hand rotation (all $p < .001$).

We found a significant main effect on Technique. MAGICPITCH and MAGICMODPITCH required less hand rotation than both G+P and MAGIC (all $p < .001$; explained by the interaction effects).

5.4 Neck Rotation

Fig. 9 illustrates neck rotation in degrees (the cumulative angular difference between the head-forward vectors in consecutive frames throughout the duration of a trial). We did not find a significant 3-way interaction nor any significant 2-way interactions. However, we found a significant main effect of Technique. MAGICPITCH and MAGICMODPITCH required higher neck rotation compared to both G+P and MAGIC (all $p < .001$). MAGICMODPITCH also required higher neck rotation than MAGICPITCH ($p < .001$).

5.5 Hand Position Distribution

We also conducted a post hoc analysis of the spatial distribution of hand positions (the thumb tip positions relative to the head; only included data from the 23 right-handed participants). Fig. 10 illustrates the spatial distribution of hand position in the transverse plane (World-XZ). MAGICPITCH and MAGICMODPITCH have less variance and more consistency compared to G+P and MAGIC in the joint X-Z distributions. For both axes, there is an approximately 25% difference in spatial peak concentration.

5.6 Subjective Metrics

For raw NASA-TLX scores, we found a significant main effect of technique for mental demand, temporal demand and frustration (all $p \leq .035$, $W \geq .012$). MAGICMODPITCH (Median 8.85, IQR [4.95-13.65]) was rated as requiring higher mental demand compared to G+P (Median 4.30, IQR [1.50-12.18]) and MAGIC (Median 5.05, IQR [2.68-10.53]; all $p \leq .050$). Participants found MAGICMODPITCH (Median 7.65, IQR [2.35-13.20]) to be more frustrating than MAGIC (Median 5.40, IQR [0.82-10.38]; $p = .031$).

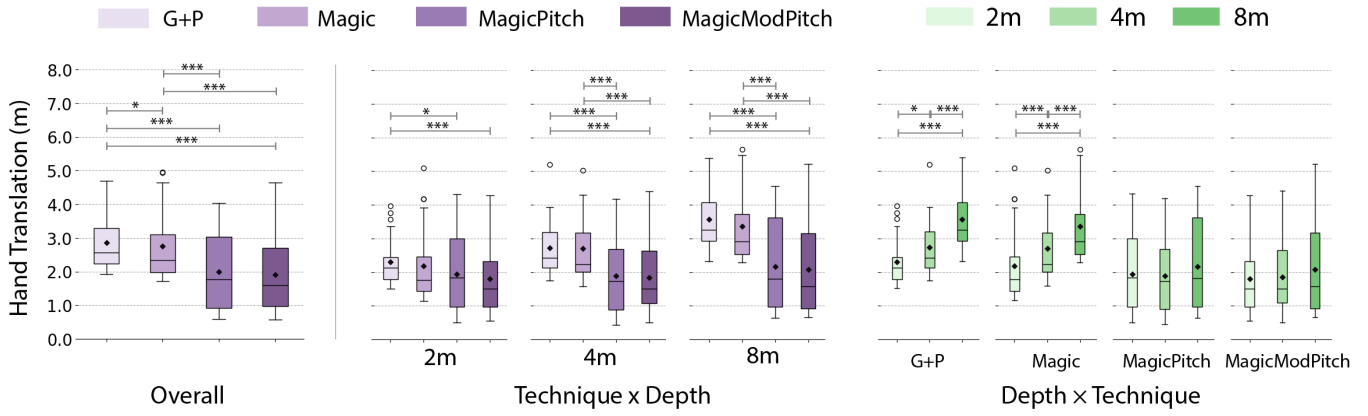


Figure 7: Hand Translation of each Technique for overall performance and the interaction effect with Depth translation distance. Mean values are represented by black diamonds. Medians are represented by horizontal lines.

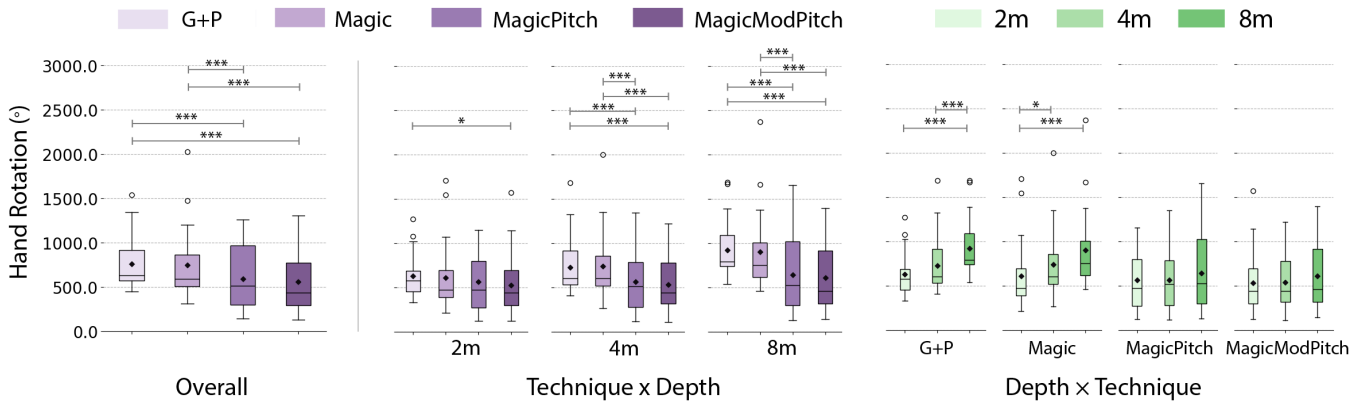


Figure 8: Hand Rotation of each Technique for overall performance and the interaction effect with Depth translation distance. Mean values are represented by black diamonds. Medians are represented by horizontal lines.

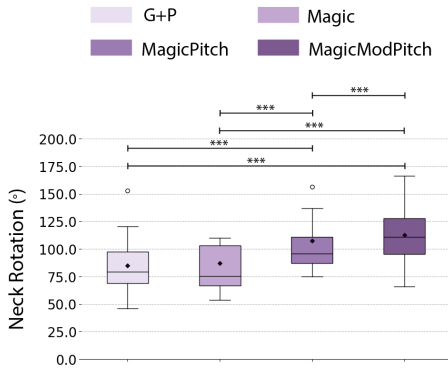


Figure 9: Neck Rotation of each Technique overall. Mean values are represented by black diamonds. Medians are represented by horizontal lines.

No significant difference was found in the post hoc analysis for temporal demand.

We also found a significant difference between techniques for hand and neck fatigue (measured in BORG CR10); all $p < .001$, $W \geq .234$). MAGICPITCH (Median 3.00, IQR [1.00-5.00]) significantly reduced hand fatigue compared to G+P (Median 5.00, IQR [3.00-7.00]; $p = .011$). MAGICMODPITCH (Median 4.00, IQR

[2.75-7.00]) was rated as higher neck fatigue compared to MAGIC (Median 2.00, IQR [1.00-4.00]; $p = .008$). No significant difference was found for eye fatigue (in BORG CR10), comfort, and ease of use ($p \geq .246$, $W \leq .058$). For detailed plots of subjective metrics, see Fig. 19 in the appendix.

5.7 Preferences and Subjective Feedback

Fig. 11 shows post-study preference ranking. G+P received the worst ranking, with 13 participants (over half) citing it as the worst technique due to high workload, fatigue, slowness, and frequent clutches. MAGIC received mixed preferences but was rarely ranked worst. Supporters saw it as a good middle ground between having only the hand and all three modalities, as gaze made it faster, easier, and less hand-intensive, but not overwhelming. Critics found it fast for radial translations, but still difficult and fatiguing for depth.

MAGICMODPITCH received the most “best” rankings (9 out of 24 participants). Together, MAGICMODPITCH and MAGICPITCH were rated as best by 13 participants in total, which is more than half of the sample. Among these 13 participants, 11 of them also rated G+P as the worst. They valued head input for depth as less tiring, faster, easier, and better for large movements. Participants who didn’t like having head pitch for depth noted drawbacks such as reduced control, coordination difficulty, unfamiliarity, and neck fatigue compared to G+P. Some noted they could adapt over time, while others relied mainly on the hand or forgot to use head pitch.

Compared to MAGICPITCH, MAGICMODPITCH was seen as less fatiguing, more stable, and easier to use, especially for far-to-

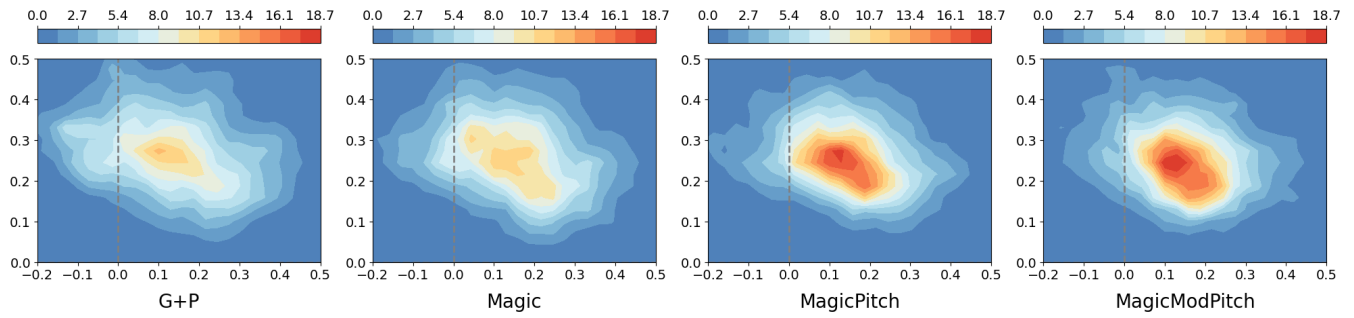


Figure 10: 2D Probability Density Distributions of Hand Positions in the Transverse Plane. The plots visualise the spatial distribution of hand positions relative to the participant’s head for each interaction technique (i.e. top-down view; $N = 23$, all right-handed). Contour colours represent the joint probability density function (PDF) value, with deeper red indicating higher spatial concentration. The horizontal axis represents the Hand Lateral Position (World-X) relative to the head (positive = rightward), while the vertical axis represents the Hand Depth (World-Z) relative to the head (positive = forward). The origin (0,0) represents the headset position at the time of interaction.

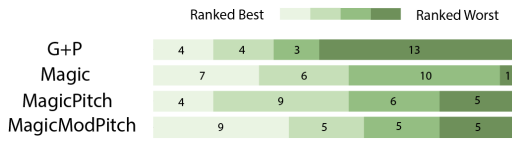


Figure 11: Preference ranking for each technique.

close docking. While two participants ranked MAGICPITCH worst due to poor control and difficulty of use, those who preferred MAGICPITCH found it smoother, easier, and more accurate. MAGICMODPITCH was reported to require excessive head tilt or occasionally fail to respond, resulting in erratic, tiring, and annoying behavior. Some could not distinguish between the two techniques.

6 DISCUSSION

In this work, we investigated integrating eye, head, and hand for 6DoF object manipulation in 3D, aiming to overcome the limitation of GAZE+PINCH (translation is limited by the hand’s movement range) by integrating hand, gaze, and head pitch input. To this end, we implemented MAGICPITCH, MAGICMODPITCH, and MAGIC, and compared them with GAZE+PINCH in a docking task.

Results show that MAGICPITCH and MAGICMODPITCH significantly reduced hand translation compared to the baselines, an important factor in reducing fatigue during prolonged interaction [26]. Specifically, the benefit of having the head pitch available for depth control becomes more pronounced as the depth translation distance increases during docking: While hand movement increases for GAZE+PINCH and MAGIC during tasks that require larger translations, it remains stable for MAGICPITCH and MAGICMODPITCH. This illustrates the successful cooperation and coordination between the hand and the neck, as participants tended to use the neck more to cover longer distances. An additional benefit is a more concentrated hand motor space during translation, allowing interaction in slightly narrower spaces, such as public transport [19]. Interestingly, although originally aimed at optimising translation, these effects also hold for rotation, though to a lesser extent. We attribute this to the fact that moving the hand less results in less unintended rotation [17], thereby eliminating the need for eventual compensatory movements to complete docking.

The lower hand effort comes, as expected, at the cost of increased neck effort. On average, we recorded an additional 25° of head rotation per trial, which is equivalent to one extra nod. Interestingly, however, this did not translate into a significant neck fatigue penalty in the BORG CR10 ratings for MAGICPITCH. This repre-

sents a favourable ergonomic trade-off: as a small rotational movement supported by the neck, head pitch imposes minimal perceived effort and is less strenuous than the large translational movement demanded of the arm. Controlling depth via head pitch reduces hand movement, lowering subjective hand fatigue without a significant increase in neck fatigue. This is particularly valuable in scenarios where hand travel, arm extension, and clutching accumulate over time (e.g. seated VR, constrained spaces, room-scale object placement, and extended authoring or gaming sessions). In these cases, offloading longer-range depth control to head pitch would be especially beneficial, as hand fatigue accumulates more slowly.

Participants also tended to need more time to complete the task with MAGICPITCH and MAGICMODPITCH compared to GAZE+PINCH and MAGIC. This might have been due to the novelty of the interaction paradigm and the greater mental effort required. A post-hoc analysis confirmed this and revealed a significant interaction between Technique $\times$ Order for completion time ($p < .001$; see Fig. 16 for detail). Participants who first experienced GAZE+PINCH and MAGIC did not take longer using MAGICPITCH and MAGICMODPITCH. It only took longer when participants experienced MAGICPITCH and MAGICMODPITCH first. This indicates that, for sufficiently trained users, the completion time with MAGICPITCH and MAGICMODPITCH is competitive. Together, our results suggest that MAGICPITCH and MAGICMODPITCH can successfully reduce hand effort by redistributing parts of depth control to the head pitch, and that completion time is competitive with that of other methods, given sufficient training or experience.

Originally, we assumed that synchronous depth control across two modalities in MAGICPITCH would lead to conflicts, thereby frustrating users, impairing performance, and increasing mental demand. We thus implemented MAGICMODPITCH to resolve this by prioritising hand input over head input during conflicts. While generally well received, as reflected in the final ranking, MAGICMODPITCH — contrary to our assumptions — led to the highest total neck rotation of all techniques, higher mental demand than GAZE+PINCH and MAGIC, and higher frustration and neck fatigue than MAGIC. Some participants had difficulty anticipating control and speed with MAGICMODPITCH, finding it frustrating when pitch-to-depth control did not respond as expected due to hand prioritisation. This reveals an interesting tension between the two variants: while hand prioritisation resolves conflicts, it can also implicitly suppress head pitch contribution to depth. This suggests room to improve how user intent is inferred (i.e., whether the hand is actively controlling and head input should be filtered) beyond our current hand-head conflict detection algorithm. Alternatively, this module may not be needed at all: to our surprise, we observed that participants could manage the conflict effectively simply through

the pinch gesture during the study.

Large depth translations and amplitudes led to behaviours that differed from those observed in the medium and small cases. Trials with a large amplitude induced a large head rotation to support gaze, which persisted after the saccade ended [28]. This triggered implicit depth control right after the gaze landed and before the hand began to manipulate, inducing synergistic and antagonistic effects [42]. While some participants appreciated that the object sometimes “magically” moved to the target, most were hesitant about these effects and tried to prevent them. This suggests that trimodal techniques should account for head rotations after gaze shifts, for example, by delaying or stabilising head-based depth input immediately after large gaze movements. Meanwhile, we found that when the object was redirected to a large eccentricity, the control axis perceived in the visual field could become misaligned with proprioceptive perception. For example, when an object appeared far in the periphery (e.g., 120° to the left), participants intended to move it rightward and would perform a rightward hand movement relative to their torso. As the hand offset was computed in world coordinates, this movement unintentionally brought the object closer in depth. Even though they soon realised the correct control is to move the hand towards both the right and the torso, this may disrupt performance and the user experience.

Participants showed asymmetric strategies for forward and backward depth translations. This is caused by our hand amplification implementation, which is calculated based on the current object distance, making it easier to bring the object closer and harder to send it farther [12]. Accordingly, participants preferred using head pitch for forward translation, while using the hand to pull objects back and correct head-pitch overshoot. This suggests that techniques may benefit from an asymmetric allocation of modalities, for example, using head pitch primarily to assist forward-depth translations.

Subjective feedback revealed strong variation in preferences: some participants preferred combining gaze, pinch, and head pitch for higher throughput, while some favoured hand-only control and found additional inputs disruptive. This suggests future implementations could modularise gaze, head pitch, and pitch, allowing users to enable or disable them. Following this, we suggest exploring a broader design space for 3D manipulation that combines the three modalities.

Our results indicate that head pitch can be integrated without degrading gaze effectiveness, and that interference between modalities can be managed through careful design. This allows for the flexible allocation of modalities across different DOF or manipulation phases. For example, the hand could support coarse depth alignment while the head provides refinement to improve comfort. Modalities may also be integrated differently, such as concurrent eye-hand input [10, 39], cascaded head-hand control [29], or weighted fusion [30]. Additional signals could extend the design space, for instance, using head yaw to refine pointing [29].

7 LIMITATIONS AND FUTURE WORK

Our task was demanding, and the rotation component of the docking task introduced noise, particularly for VR novices. Even with GAZE+PINCH, several participants struggled and frequently lifted their hands above the headset, especially when first encountering the trimodal technique. Consequently, some results may reflect novice behaviour. Furthermore, due to a lack of experience, they relied on visual aids (orientation gizmos and depth rings) to complete the task, especially for targets at a distance. While these aids were intentionally introduced to isolate technique effects, they may have scaffolded performance in ways that limit the generalisation. Future work should include more extensive training, recruit participants with broader VR experience, or adjust task difficulty to better isolate technique effects.

The task is limited to abstract object manipulation in an empty

space within a 1–11m interaction range, and the technique may require adaptation to denser target layouts, different distances, or more realistic application tasks. When targets are close together, gaze-based selection becomes challenging due to limited pointing accuracy [29]. Techniques integrate head or hand input [21, 29] could then refine the gaze selection before handing off to MAGICPITCH. Occlusion within clusters or during pitch-depth control could also cause users to lose track of the target. With this, visual aids such as making occluders semi-transparent [8] may help, and prior depth-layer selection techniques [38] could be adapted for head-pitch control to aid target-depth disambiguation. Regarding distance, performance below 1m or beyond 11m remains unexplored. At close range, transitioning to direct hand manipulation would likely improve stability. At far range, depth perception degrades, and additional visual aids (e.g. the ones we implemented in the study) become necessary. For extreme ranges, alternative strategies such as teleportation may be more appropriate.

Our results are limited by pinch-gesture tracking in our apparatus (Sec. 4.3). Regrabbing the object introduced small position and rotation offsets, disrupting refinement. In addition, reliable pinch detection required participants to bring and twist their hands towards the tracking camera on the headset, which influenced hand movement. This may have increased refinement time and hand fatigue, especially in techniques that relied on clutching.

Pinch served as a grab and as a clutching gesture for head-pitch input. Therefore, our findings are most directly applicable to techniques in which head-pitch is integrated with a hand gesture, such as a pinch, palm grab, or button press. Designs with head-pitch input independent from hand gestures may yield different performance and user strategies. Moreover, as with gaze-redirected manipulation [49, 48], our technique requires users not to look away while manipulating the object. To look around freely, they still needed to release the pinch to disengage the interaction.

8 CONCLUSION

In this work, we investigated the use of gaze, pinch, and head pitch together for 3D object manipulation. Our results show that this integration is effective and feasible: users can coordinate these three modalities to simultaneously manipulate different DoF of objects. This is significant because it shows that interaction techniques could be designed to support more than two, but three distinct, actively controlled input modalities beyond implicit model fusion. Redistribution of depth control to head pitch yields substantial benefits in reduced hand effort and improved ergonomics, but meanwhile introduces a trade-off in increased neck motion and, when lacking experience, slightly slower task completion. This suggests an interesting direction for optimising multimodal XR interaction to dynamically balance the workload across input modalities. For practical use, as the MAGICPITCH technique is readily available, with modularised components (gaze, head pitch, and pitch modulation) can be selected on demand and used in different interaction contexts. Taken together, our findings highlight the value of complementing hand input with coordinated gaze and head movement in the design of multimodal 3D interaction techniques.

SUPPLEMENTAL MATERIALS

The open-source code, demo Unity project, video figure, and raw data are available at: <https://doi.org/10.5281/zenodo.21173851>

ACKNOWLEDGMENTS

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

REFERENCES

- [1] M. R. Bashar, M. Amini, W. Stuerzlinger, M. Sarac, K. Pfeuffer, M. D. Barrera Machuca, and A. U. Batmaz. Depth3DSketch: Freehand Sketching Out of Arm's Reach in Virtual Reality. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25, pp. 1–8. Association for Computing Machinery, New York, NY, USA, Apr. 2025. doi: 10.1145/3706599.3719717 2
- [2] R. Bates and H. Istancë. Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices. *Universal Access in the Information Society*, 2(3):280–290, Oct. 2003. doi: 10.1007/s10209-003-0053-y 2
- [3] G. A. Borg. Psychophysical bases of perceived exertion. *Medicine and science in sports and exercise*, 14(5):377–381, Jan. 1982. doi: 10.1249/00005768-198205000-00012 5
- [4] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics - SI3D '97*, pp. 35–ff. ACM Press, Providence, Rhode Island, United States, 1997. doi: 10.1145/253284.253301 2
- [5] J. Byers, A. Bittner, and S. Hill. Traditional and raw Task Load Index(TLX) correlations : Are paired comparisons necessary? *Advances in Industrial Ergonomics and Safety*, pp. 481–485, 1989. 5
- [6] 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, pp. 2527–2530. Association for Computing Machinery, New York, NY, USA, May 2012. doi: 10.1145/2207676.2208639 12
- [7] H. J. Chae, J.-i. Hwang, and J. Seo. Wall-based Space Manipulation Technique for Efficient Placement of Distant Objects in Augmented Reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 45–52. ACM, Berlin Germany, Oct. 2018. doi: 10.1145/3242587.3242631 2
- [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, pp. 1–11. Association for Computing Machinery, New York, NY, USA, Oct. 2023. doi: 10.1145/3611659.3615693 9
- [9] 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, pp. 1–14. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3581201 2
- [10] J. Jeong, S.-H. Kim, H.-J. Yang, G. Lee, and S. Kim. GazeHand2: A Gaze-Driven Virtual Hand Interface with Improved Gaze Depth Control for Distant Object Interaction. *Electronics*, 14(13):2530, Jan. 2025. doi: 10.3390/electronics14132530 2, 4, 9
- [11] J. Jeong, S.-H. Kim, H.-J. Yang, G. A. Lee, and S. Kim. GazeHand: A Gaze-Driven Virtual Hand Interface. *IEEE Access*, 11:133703–133716, 2023. doi: 10.1109/ACCESS.2023.3337372 2
- [12] J. J. L. Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, Apr. 2017. 1, 9
- [13] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billingham. 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 2
- [14] Y. Lee, D. Park, and Y. M. Kim. The effect of wearing a head-mounted display on the boundaries of the cervical range of motion based on perceived comfort in a static posture. *Virtual Reality*, 27(2):815–828, June 2023. doi: 10.1007/s10055-022-00684-w 2
- [15] X. Liu, L. Wang, S. Luan, X. Shi, and X. Liu. Distant Object Manipulation with Adaptive Gains in Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 739–747, Oct. 2022. doi: 10.1109/ISMAR55827.2022.00092 2
- [16] 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, pp. 1–12. Association for Computing Machinery, New York, NY, USA, May 2019. doi: 10.1145/3290605.3300842 2
- [17] T. Mikkelsen, Q. Zhou, M. N. Lystbæk, Y. Liu, H. Gellersen, and K. Pfeuffer. DOF-Separation for 3D Manipulation in XR: Understanding Finger-Wrist Separation to Simultaneously Translate and Rotate Objects. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1246–1256, Oct. 2025. doi: 10.1109/ISMAR67309.2025.00130 2, 3, 4, 8
- [18] P. Mohan, W. Boon Goh, C.-W. Fu, and S.-K. Yeung. Head-Fingers-Arms: Physically-Coupled and Decoupled Multimodal Interaction Designs in Mobile VR. In *Proceedings of the 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pp. 1–9. ACM, Brisbane QLD Australia, Nov. 2019. doi: 10.1145/3359997.3365697 2
- [19] A. Ng, D. Medeiros, M. McGill, J. Williamson, and S. Brewster. The Passenger Experience of Mixed Reality Virtual Display Layouts in Airplane Environments. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 265–274, Oct. 2021. doi: 10.1109/ISMAR52148.2021.00042 8
- [20] 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, Budapest, Hungary, Oct. 2020. doi: 10.1109/ISMCR51255.2020.9263775 2
- [21] K. Patel, V. Phadnis, E. J. Gonzalez, H. Gellersen, K. Pfeuffer, and M. Gonzalez-Franco. H2E: Hand, Head, Eye a Multimodal Cascade of Natural Inputs. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 84–89, Mar. 2025. doi: 10.1109/VRW66409.2025.00026 2, 9
- [22] K. Pfeuffer, H. Gellersen, and M. Gonzalez-Franco. Design Principles and Challenges for Gaze + Pinch Interaction in XR. *IEEE Computer Graphics and Applications*, 44(3):74–81, May 2024. doi: 10.1109/MCG.2024.3382961 1, 2
- [23] 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, pp. 99–108. Association for Computing Machinery, New York, NY, USA, Oct. 2017. doi: 10.1145/3131277.3132180 1, 2, 3
- [24] 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, pp. 91–98. Association for Computing Machinery, New York, NY, USA, Oct. 2017. doi: 10.1145/3131277.3132182 2
- [25] A. Ramcharitar and R. J. Teather. EZCursorVR: 2D Selection with Virtual Reality Head-Mounted Displays. In *Proceedings of the 44th Graphics Interface Conference*, GI '18, pp. 123–130. Canadian Human-Computer Communications Society, Waterloo, CAN, June 2018. doi: 10.20380/GI2018.17 2
- [26] R. Shi, Y. Wei, X. Qin, P. Hui, and H.-N. Liang. Exploring Gaze-assisted and Hand-based Region Selection in Augmented Reality. *Proc. ACM Hum.-Comput. Interact.*, 7(ETRA):160:1–160:19, May 2023. doi: 10.1145/3591129 8
- [27] 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, pp. 1–15. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3580685 2
- [28] L. Sidenmark and H. Gellersen. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Transactions on Computer-Human Interaction*, 27(1):1–40, Jan. 2020. doi: 10.1145/3361218 2, 9
- [29] 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*, pp. 1–9. ACM, Stuttgart Ger-

- many, June 2020. doi: 10.1145/3379155.3391312 9
- [30] L. Sidenmark, M. Parent, M. Glueck, D. Wigdor, T. Grossman, and M. Giordano. Weighted Pointer: Error-aware Gaze-based Interaction through Fallback Modalities. *IEEE transactions on visualization and computer graphics*, 28(11):3585–3595, Nov. 2022. doi: 10.1109/TVCG.2022.3203096 9
- [31] A. Simonson, D. Gretton, and C. Harteveld. Bird: A Point Cursor for Virtual Immersive Environments. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '25, pp. 1–6. Association for Computing Machinery, New York, NY, USA, Apr. 2025. doi: 10.1145/3706599.3720045 2
- [32] S. Sultana, J. F. Gonzalez, and R. J. Teather. Comparing Controller-Free Pointing Techniques Across Depth for 2D Selection in Augmented Reality, June 2026. doi: 10.48550/arXiv.2606.08441 2
- [33] 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, pp. 265–272. Association for Computing Machinery, New York, NY, USA, Apr. 2000. doi: 10.1145/332040.332443 2
- [34] J. Turner, J. Alexander, A. Bulling, and H. Gellersen. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pp. 4179–4188. Association for Computing Machinery, New York, NY, USA, Apr. 2015. doi: 10.1145/2702123.2702355 2, 3
- [35] J. Turner, J. Alexander, A. Bulling, D. Schmidt, and H. Gellersen. Eye Pull, Eye Push: Moving Objects between Large Screens and Personal Devices with Gaze and Touch. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2013*, pp. 170–186. Springer, Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-40480-1_11 2
- [36] J. Turner, A. Bulling, and H. Gellersen. Combining gaze with manual interaction to extend physical reach. In *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction*, PETMEI '11, pp. 33–36. Association for Computing Machinery, New York, NY, USA, Sept. 2011. doi: 10.1145/2029956.2029966 2
- [37] E. K. Tütüncü, M. Gonzalez-Franco, and E. J. Gonzalez. HandOver: Enabling Precise Selection & Manipulation of 3D Objects with Mouse and Hand Tracking. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–11. ACM, Busan Republic of Korea, Sept. 2025. doi: 10.1145/3746059.3747757 4
- [38] U. Wagner, M. Albrecht, A. A. Jacobsen, H. Wang, H. Gellersen, and K. Pfeuffer. Gaze, Wall, and Racket: Combining Gaze and Hand-Controlled Plane for 3D Selection in Virtual Reality. *Proc. ACM Hum.-Comput. Interact.*, 8(ISS):534:189–534:213, Oct. 2024. doi: 10.1145/3698134 9
- [39] U. Wagner, A. Asferg Jacobsen, T. Feuchtner, H. Gellersen, and K. Pfeuffer. Eye-Hand Movement of Objects in Near Space Extended Reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, UIST '24, pp. 1–13. Association for Computing Machinery, New York, NY, USA, Oct. 2024. doi: 10.1145/3654777.3676446 1, 2, 3, 9
- [40] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbaek, K. Pfeuffer, and H. Gellersen. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–15. Association for Computing Machinery, New York, NY, USA, Apr. 2023. doi: 10.1145/3544548.3581423 2, 4
- [41] H. Wang, L. Sidenmark, F. Weidner, J. Newn, and H. Gellersen. Head-Shift: Head Pointing with Dynamic Control-Display Gain. *ACM Trans. Comput.-Hum. Interact.*, 32(1):2:1–2:28, Apr. 2025. doi: 10.1145/3689434 2
- [42] H. Wang, F. Weidner, Y. Abdrabou, K. Pfeuffer, and H. Gellersen. HeadDepth: Gaze Raycasting with Head Pitch for Depth Control. In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 370–380, Oct. 2025. doi: 10.1109/ISMAR67309.2025.00048 1, 2, 3, 4, 9, 12
- [43] M. R. Williams and R. F. Kirsch. Evaluation of head orientation and neck muscle EMG signals as three-dimensional command sources. *Journal of NeuroEngineering and Rehabilitation*, 12(1):25, Mar. 2015. doi: 10.1186/s12984-015-0016-6 2
- [44] 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, pp. 143–146. Association for Computing Machinery, New York, NY, USA, May 2011. doi: 10.1145/1978942.1978963 5
- [45] 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, pp. 1–10. Association for Computing Machinery, New York, NY, USA, Apr. 2020. doi: 10.1145/3313831.3376876 4
- [46] W. Xu, X. Meng, K. Yu, S. Sarcar, and H.-N. Liang. Evaluation of Text Selection Techniques in Virtual Reality Head-Mounted Displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 131–140. IEEE, Singapore, Singapore, Oct. 2022. doi: 10.1109/ISMAR55827.2022.00027 2
- [47] 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. IEEE, Beijing, China, Oct. 2019. doi: 10.1109/ISMAR.2019.00-20 2
- [48] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves. Gaze-Supported 3D Object Manipulation in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 1–13. Association for Computing Machinery, New York, NY, USA, May 2021. doi: 10.1145/3411764.3445343 1, 2, 3, 4, 9
- [49] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99*, pp. 246–253. ACM Press, Pittsburgh, Pennsylvania, United States, 1999. doi: 10.1145/302979.303053 1, 2, 9
- [50] 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, pp. 1–18. Association for Computing Machinery, New York, NY, USA, May 2024. doi: 10.1145/3613904.3642589 2
- [51] K. Zhang, C. Moreira, P. Belchior, G. Lee, M. Billingham, and J. Jorge. HeadZoom: Hands-Free Zooming and Panning for 2D Image Navigation Using Head Motion, Aug. 2025. doi: 10.48550/arXiv.2508.01765 2
- [52] D. B. Zhao and W. Stuerzlinger. A Novel Bare-Handed Manipulation Technique for Distant Objects in Virtual Reality. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 353–360, Mar. 2025. doi: 10.1109/VRW66409.2025.00083 2

A DETAILED TECHNIQUE IMPLEMENTATIONS

A.1 GAZE+PINCH

Objects were selected with gaze pointing. Both gaze ray origin and direction was 1€ filtered [6] with $f_{c_{min}} = 0.05$, $\beta = 10$. To minimise the effects of asynchronous sampling between the gaze and head trackers in our apparatus, where head rotation could shift the estimated gaze direction during fixation, we corrected the gaze direction by applying a rotation offset computed from the change in head orientation between the current frame and the frame 10 samples earlier. Meanwhile, to minimise the effect of gaze accuracy for object selection, the object could be selected either by gaze raycast hit or falls in a gaze cone with 10° radius. A pinch gesture was detected with a threshold of 1.5 cm Euclidean distance between the index and thumb tips.

While holding the pinch, the hand translate and rotates the object with a relative mapping. At each frame, the algorithm takes the position difference of the thumb tip of the dominant hand from the previous frame, passes it to the visual angular hand amplification function, and then applies it to the object:

$$p'_{obj} = p_{obj} + (\vec{s}_{obj})_{hand} \quad (1)$$

where p'_{obj} and p_{obj} denote the object's position in the current and previous frame, respectively. $(\vec{s}_{obj})_{hand}$ is the object translation offset created by hand:

$$(\vec{s}_{obj})_{hand} = \min\left(\frac{\|p_{obj} - p_{gaze}\|}{\|p_{hand} - p_{gaze}\|}, 1\right) \cdot \vec{s}_{hand} \quad (2)$$

where G_{hand} is the difference between the thumb tip positions between consecutive frames $\vec{s}_{hand}$, representing the hand translation. This is amplified by the visual angular gain function, which is derived by the ratio of the distance between the object and gaze origin (p_{gaze} ; the mid position of the eyes), and the distance between the pinching hand (p_{hand} ; again, tracked by thumb tip) and the gaze origin.

Object rotation follows the hand rotation isomorphically, also with a relative mapping while holding the pinch:

$$\mathbb{H}'_{obj} = \Delta\mathbb{H}_{hand} * \mathbb{H}_{obj} \quad (3)$$

where $\mathbb{H}'_{obj}$ and $\mathbb{H}_{obj}$ represent the current and previous object orientations, respectively. $\Delta\mathbb{H}_{hand}$ is the hand rotation offset between consecutive frames.

A.2 MAGIC

In gaze mode, the object is snapped and follows the gaze direction while keeping the same distance from the user:

$$p'_{obj} = p_{gaze} + \|p_{obj} - p_{gaze}\| \cdot \hat{d}_{gaze} \quad (4)$$

where p_{gaze} is the gaze origin and $\hat{d}_{gaze}$ is the unit vector of gaze direction. $\|p_{obj} - p_{gaze}\|$ represents the current distance between the object and gaze origin.

Fig. 12 visualises the gaze-hand mode switching and input in each mode. Fixation time threshold set to 0.25s and dispersion threshold set to 3° for the real-time I-DT. Safe region radius $\theta_{thr} = 15^\circ$. Values were tuned via pilot tests.

A.3 MAGICPITCH

Fig. 13 illustrates the integration of head pitch input to MAGIC. The full integration algorithm is summarised in Algorithm 1. MAGICPITCH adds depth offset during hand mode at each frame, following a relative pitch-to-depth mapping [42]:

$$p'_{obj} = p_{obj} + (\vec{s}_{obj})_{pitch} \quad (5)$$

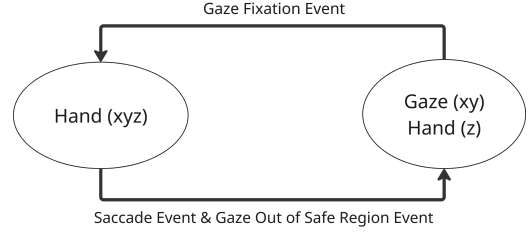


Figure 12: Integration and mode switching between gaze and hand in MAGIC. “xy” represents 2D direction from users’ first-person perspective, “z” represents depth. If a saccade is detected and the gaze point is outside the safe region in hand mode, it switches to gaze mode, where gaze takes over direction, and hand input only controls depth. If gaze fixation is detected in gaze mode, it switches to hand mode, where the hand fully controls the object’s translation. Note that rotation is always controlled by hand and independent of the mode.

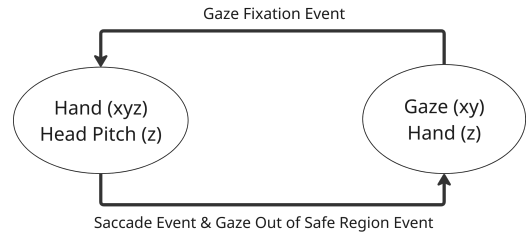


Figure 13: Integration and mode switching between gaze, hand, and head pitch in MAGICPITCH. The only difference from Fig. 12 of MAGIC is having head pitch for depth control (“z”) during hand mode.

$(\vec{s}_{obj})_{pitch}$ is the object translation offset created by head pitch, calculated by:

$$(\vec{s}_{obj})_{pitch} = G_{pitch} \cdot \Delta pitch \cdot \hat{u}(P_{obj} - P_{gaze}) \quad (6)$$

$\Delta pitch$ is a scalar that denotes the change in head pitch angle (1€ filtered with $f_{c_{min}} = 1$, $\beta = 0$) since the last frame (positive for upward tilt, negative for downward tilt). $\hat{u}(P_{obj} - P_{gaze})$ is the unit vector from the gaze origin to the object, defining the forward direction of the depth axis. G_{pitch} is the gain function for pitch-to-depth mapping, with algorithm implemented identical to EdgeGain [42]:

$$G_{pitch} = F_{scale} \cdot G_{rel} \quad (7)$$

where G_{rel} and F_{scale} are the corresponding functions for speed-sensitive gain and the scaling factor for optimising the viewing angle. G_{rel} sensitivity was tuned to adapt to the range of depth translation for the task. The minimum gain value G_{min} in the original function was tuned to zero to have a head pitch speed deadzone to enhance the stability:

$$G_{rel} = \begin{cases} 0, & \text{if } v_{pitch} \leq v_{min} \\ \frac{G_{max}}{v_{max} - v_{min}}(v_{pitch} - v_{min}), & \text{if } v_{min} < v_{pitch} < v_{max} \\ G_{max}, & \text{if } v_{pitch} \geq v_{max} \end{cases} \quad (8)$$

where v_{min} was set lower to 0.1 rad/s to accommodate setting G_{min} to zero. $v_{max} = 0.6$ rad/s remained the same as in previous work. $G_{max} = 0.8$ m/°, doubled to adopt the doubled depth range in the task. F_{scale} parameters remained identical to previous work [42]. Object depth distance from the gaze was clamped according to the task setting (1m-11m).

Algorithm 1 Object Translation and Rotation with MAGICPITCH.

```

1:  $p_{obj} \leftarrow p_{obj} + (\vec{s}_{obj})_{hand}$ 
2:  $\mathbb{H}_{obj} \leftarrow \Delta \mathbb{H}_{hand} * \mathbb{H}_{obj}$ 
3: if Mode = Gaze then
4:    $p_{obj} \leftarrow p_{gaze} + \|p_{obj} - p_{gaze}\| \cdot \hat{d}_{gaze}$ 
5:   if GazeFixating then
6:     Mode  $\leftarrow$  Hand
7:   end if
8: else
9:    $p_{obj} \leftarrow p_{obj} + (\vec{s}_{obj})_{pitch}$ 
10:  if  $\neg$ GazeFixating and GazeOffset  $> \theta_{thr}$  then
11:    Mode  $\leftarrow$  Gaze
12:  end if
13: end if

```

$\triangleright$ Apply amplified hand translation to object.
 $\triangleright$ Apply hand rotation to object.
 $\triangleright$ Align object to gaze direction.
 $\triangleright$ Apply head-pitch offset to object.

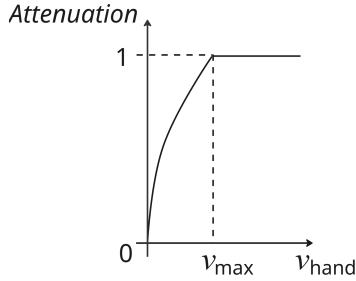


Figure 14: Visualized *Attenuation* function in Equation 10.

A.4 MAGICMODPITCH

Head pitch control was modified by applying an attenuation function to the object translation offset created by head pitch:

$$p'_{obj} = p_{obj} + (1 - \text{Attenuation}) \cdot (\vec{s}_{obj})_{pitch} \quad (9)$$

where *Attenuation* is the attenuation function for head pitch input, visualised in Figure 14. The attenuation is activated when the hand translation vector opposes the translation offset vector induced by head pitch. This condition is detected both in 3D space and in its projection onto the transverse plane:

$$\text{Atten.} = \begin{cases} \min\left(\sqrt{\frac{v_{hand}}{v_{hmax}}}, 1\right), & \text{if } \vec{s}_{hand} \cdot (\vec{s}_{obj})_{pitch} < 0 \text{ || } \\ & \text{proj}_{xz}(\vec{s}_{hand}) \cdot \text{proj}_{xz}((\vec{s}_{obj})_{pitch}) < 0 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where v_{hand} is hand translation speed, and $v_{hmax} = 0.1$ m/s is the maximum hand speed threshold, at which attenuation reaches its maximum value of 1. $\text{proj}_{xz}(\vec{s}_{hand})$ and $\text{proj}_{xz}((\vec{s}_{obj})_{pitch})$ are the projections of the 1€-filtered hand translation vector $\vec{s}_{hand}$ and the head-pitch offset $(\vec{s}_{obj})_{pitch}$ onto the transverse (XZ) plane, respectively. We apply a square-root scaling to v_{hand} to achieve higher responsiveness at low speeds.

B VIRTUAL ENVIRONMENT

We adopted the virtual environment from the Meta Quest Interaction SDK Sample scene¹, which is in a grey-themed infinite environment, to let participants not be distracted by the test environment. A grid was displayed at ground level to enhance depth perception. Fig. 15 shows such a scene. We placed furniture 11

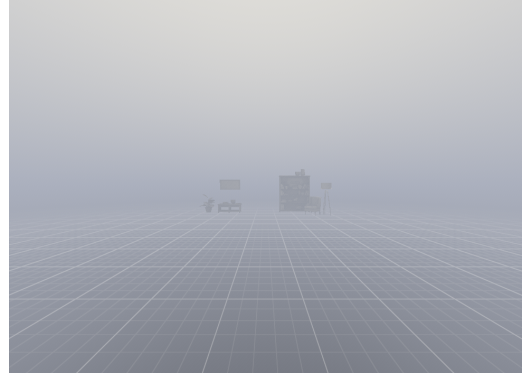


Figure 15: The virtual environment used in the study, shown from the participants' first-person perspective when facing straight ahead.

meters away from the participants, indicating the maximum interaction range, and a grid was shown on the ground level (calibrated to align the real-world floor level) to enhance the depth cue.

Object and target size were the same and remained the same throughout the study (0.4m in diameter; visual degrees at 2m: 11.42°; 4m: 5.72°; 6m: 3.82°; 10m: 2.30°; visualised with cyan squares in Fig. 5).

C SUPPLEMENTARY PLOTS FOR RESULT

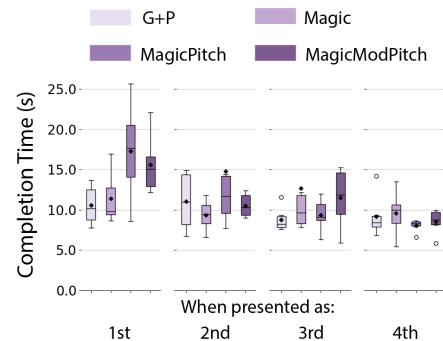


Figure 16: Completion Time of techniques when being presented as the first, second, third, and fourth (last) to the participants. Mean values are represented by black diamonds. Medians are represented by horizontal lines. Specifically for this figure, no statistical analysis was performed to assess differences.

¹<https://developers.meta.com/horizon/documentation/unity/unity-isdk-comprehensive-rig-example-scene>

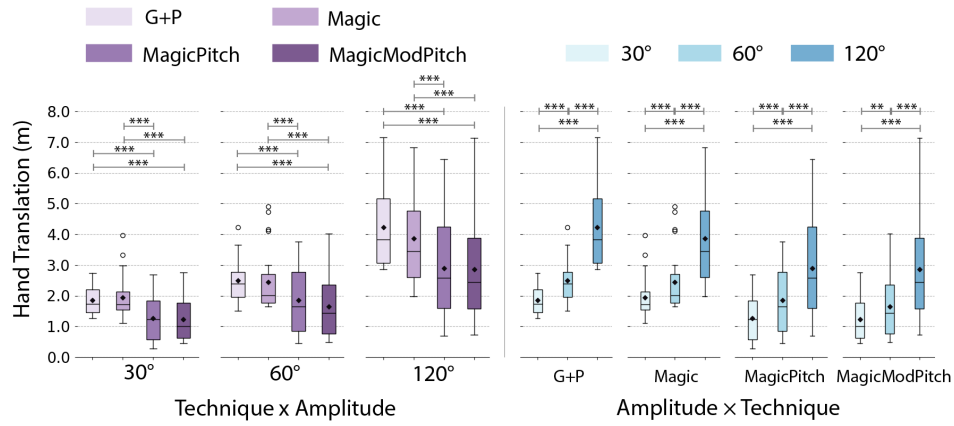


Figure 17: The interaction effect of Technique $\times$ Amplitude for Hand Translation. Mean values are represented by black diamonds. Medians are represented by horizontal lines.

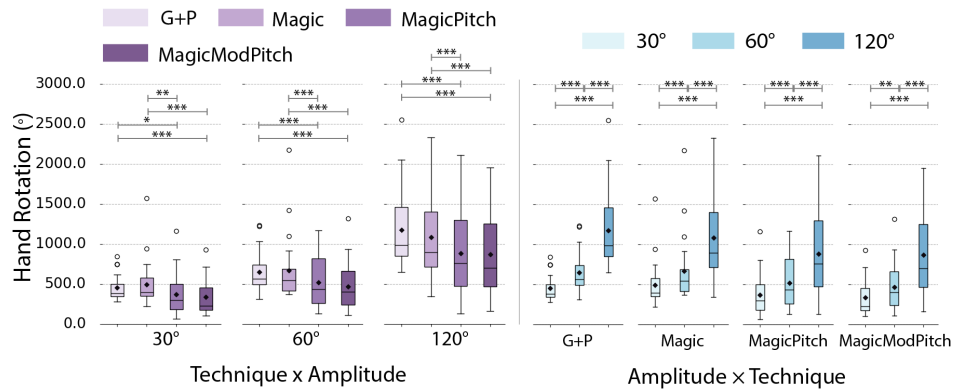


Figure 18: The interaction effect of Technique $\times$ Amplitude for Hand Rotation. Mean values are represented by black diamonds. Medians are represented by horizontal lines.

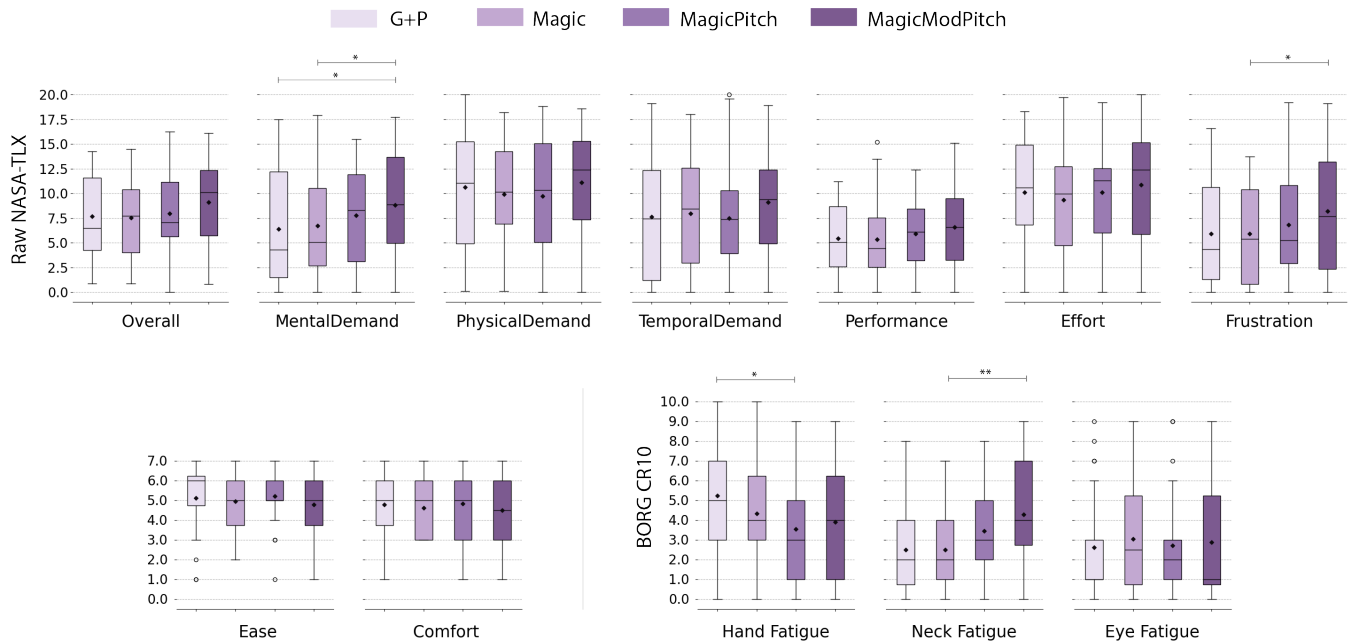


Figure 19: Subjective metrics for each Technique. Mean values are represented by black diamonds. Medians are represented by horizontal lines.