DOF-Separation for 3D Manipulation in XR: Understanding Finger-Wrist Separation to Simultaneously Translate and Rotate Objects

Thorbjørn Mikkelsen* Aarhus University
Qiushi Zhou† Aarhus University
Mathias N. Lystbæk‡ Aarhus University
Yang Liu§ Aarhus University
Hans Gellersen¶ Lancaster University, Aarhus University
Ken Pfeuffer‖ Aarhus University

Abstract

Hand-tracking based 3D object manipulation in Extended Reality (XR) typically employs a pinch gesture for acquisition and manipulation through a direct mapping from 6-degrees-of-freedom (DOF) hand movement to that of the object. In this work, we investigate the effect of separating this mapping to concurrent 3DOF controls (DOF-Separation) of translation and rotation of the virtual object using the position and orientation of the hand independently. We aim to understand how DOF-Separation could ease manipulation for different techniques with varying requirements for hand position and orientation during acquisition, including Virtual Hand, Hand Ray, and Gaze&Pinch. Through a user study that features a docking task in VR, we found that DOF-Separation significantly improves the manipulation performance of Hand Ray, while improving that of Virtual Hand only in difficult tasks of combined translation and rotation. We suggest future XR systems to adopt DOF-Separation for input in manipulation-heavy applications, such as 3D design.

Index Terms: gaze input, 3D manipulation, extended reality, VR

1 Introduction

3D interaction is widely recognised as a key challenge for XR [3, 5]. While interaction with 2D interface elements, such as windows, buttons, and scrollbars, follows the "WIMP" metaphor [45], acquiring and manipulating a virtual object in 3D space is a unique affordance of XR. Modern XR devices typically feature 6 degrees-of-freedom (DOF) hand-tracking for 3D input techniques, useful to operate a variety of tasks from basic object manipulation to sophisticated CAD design and advanced workflows in 3D modelling, gaming, architecture, automotive, and manufacturing [23, 30]. For instance, design software like Arkio [1] allows users to manipulate distant objects in context for a broad overview while also enabling rapid, precise adjustments to elements in their immediate near-space. These techniques implement different movement mappings between the hand and the target object for acquisition following different metaphors. Virtual Hand renders users' hands in their physical locations for direct manipulation of virtual objects within reach. Hand Ray extends the user's reach to acquire an object over a distance using a ray originating from their hand, usually combined with a pinch gesture to confirm pointing. Gaze&Pinch frees the hand from pointing by replacing it with eye gaze, combined with a pinch gesture performed from anywhere. Each acquisition metaphor has unique requirements for the position and orientation of the pinching hand that affect its capabilities of subsequent manipulation. In this work, we investigate how 3D input techniques could improve manipulation capabilities by untangling it from the preceding acquisition task using DOF-Separation.

*e-mail: thormik@cs.au.dk
†e-mail: qiushi.zhou@cs.au.dk
‡e-mail: mathiasl@cs.au.dk
§e-mail: yangliu.hci@gmail.com
¶e-mail: h.gellersen@lancaster.ac.uk
‖e-mail: ken@cs.au.dk

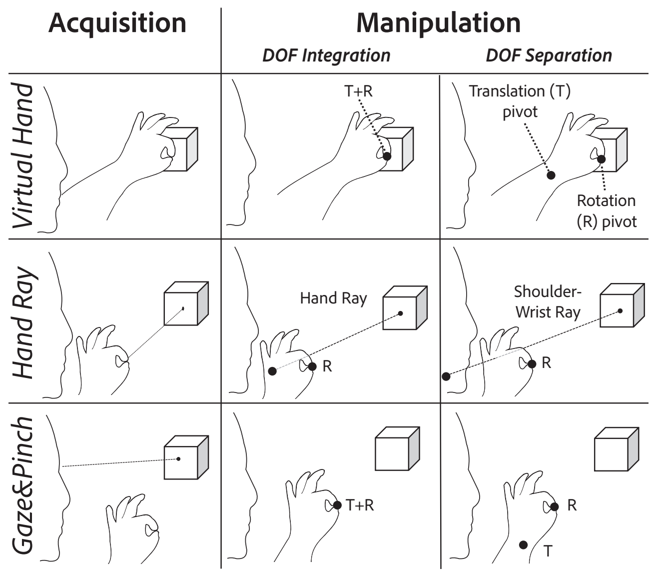


Figure 1: To efficiently manipulate in 3D space, current XR systems support three common input methods for 6 degrees-of-freedom (DOF) object manipulations: direct grasping with a unified input/output mapping, as well as gaze and hand-ray-based input where the hand is decoupled from the manipulation object. In our study, we examine how this affects 3D object docking (translation + rotation) tasks, for both DOF integration and separation mappings to uncover the speed-accuracy performance trade-offs.

Pinching marks the end of the acquisition task and the beginning of the manipulation task. The former determines the complexity of the latter through its different requirements for the hand at pinch: Virtual Hand demands the position of the pinch point to be on the object, Hand Ray demands the hand to be pointing toward the object, whereas Gaze&Pinch does not require any spatial configuration of the hand. These different restrictions significantly affect the difficulty of the subsequent manipulation. For instance, whereas Gaze&Pinch allows a user to easily rotate an object in-place for 180° by flexing the wrist joint of the pinching hand, the same task is nearly impossible with Hand Ray due to the tight coupling between the hand orientation and the displacement of the manipulated object. Novel solutions have been proposed that address this issue by easing the restrictions on the hand for acquisition and manipulation. MRTK [8] and Leap Motion [44] have implemented versions of the Hand Ray employing a shoulder-stabilized vector that decouples the orientation of the hand from acquisition, such that hand rotation post-acquisition causes the object to rotate in-place. This separated mapping of 6DOF hand movement into 3DOF controls of translation and rotation has been formalised as "DOF-Separation" [31]. By dividing the manipulation task to the translational and rotational movements of the hand, DOF-Separation significantly strengthens the manipulation capabilities of Hand Ray.

Though the benefit of DOF-Separation seems obvious for Hand Ray, its effects on Virtual Hand and Gaze&Pinch are unclear. Whereas separating translation and rotation theoretically frees the hand from orienting for acquisition in all techniques, it breaks the direct manipulation metaphor of Virtual Hand, and may have little effect on Gaze&Pinch. In this paper, we present a study that contributes novel understandings of single-hand pinch-manipulation of virtual objects by investigating the interplay between Input-Metaphor (Virtual Hand, Hand Ray, Gaze&Pinch) and DOF-Separation (Integrated and Separated) of the 6DOF hand movement into 3DOF controls of translation and rotation. We implemented six techniques (Figure 1) and compared them in a VR docking task with varying complexity levels: Translation, Rotation, and Combined. Overall, our results suggest that DOF-Separation benefits Hand Ray in all tasks by imposing less restriction on the orientation of the hand. For Virtual Hand, similar performance benefits are found in more complex tasks where the extra freedom of hand orientation is helpful. Conversely, Gaze&Pinch is less affected by DOF-Separation due to its relaxed requirement for the hand at acquisition. Based on our findings, we recommend DOF-Separation at the wrist joint for Hand Ray and for Virtual Hand in use cases where frequent complex manipulation tasks are expected, such as in 3D designing and modelling applications.

2 Related Work

We summarise the state-of-the-art 3D hand-based interaction techniques in XR, including Virtual Hand, Hand Ray, and Gaze&Pinch with varying requirements of hand position and orientation for acquisition and subsequent 6DOF manipulation. We also revisit related work on using DOF-Separation to map hand movement to 3DOF controls for easier manipulation.

2.1 3D hand-based interaction techniques in XR

Equipped with advanced hand tracking, state-of-the-art XR systems typically support Virtual Hand input techniques that render the user's hands in the virtual dimension to interact with virtual objects by touching and pinching [23]. Virtual Hand is often regarded as the "most natural" due to its emulation of hand interaction in the physical reality, and to the direct 6DOF mapping between the movement of the user's hand and of virtual objects [30].

To extend users' access to the virtual space beyond reach, previous work has explored different means to acquire and manipulate distant objects while adapting the Virtual Hand metaphor [37, 18]. Poupyrev et al. proposed the "go-go" technique that prolongs the user's arm to reach distant objects, allowing for seamless direct manipulation for near and far objects [37]. Bowman and Hodges proposed the "HOMER" (Hand-centered Object Manipulation Extending Ray-casting) technique that enabled a movement mapping which is intuitive to learn and use [6, 52]. These techniques similarly involve a "teleportation" of the user's gestures toward the direction of the target object that the hand is pointing at, to enable Virtual Hand interaction over distances. Following this metaphor, many modern XR systems adopt Hand Ray as the default option for distant 3D interaction. These techniques use a virtual ray originating from or relying on the user's hand as a pointer, while acquisition is confirmed with a pinch gesture. Some of these systems implement a "stick" metaphor for manipulation, including the Meta XR SDK [33]. In these cases, the ray becomes rigid upon pinching, and subsequent wrist flexion will result in the displacement of the object. Conversely, Mixed Reality Toolkit (MRTK) [8] and Leap Motion guidelines [44] adopt a Hand Ray that originates from the user's shoulder (or shoulder and wrist) position and is aimed through the hand, enabling the use of wrist movement to control the rotation of the object without displacement. This approach addresses the limitation of the "stick" metaphor by freeing hand rotation from the pointing task.

Previous work explored gaze pointing to replace Hand Ray while further freeing the hand and arm from acquisition [43, 56]. For instance, Apple Vision Pro adopts Gaze&Pinch as its default input technique for system navigation [35][1]. Gaze&Pinch enables acquisition by looking at the target while performing a pinch gesture with the hand being anywhere within the tracking range, and subsequent manipulation through a direct mapping between 6DOF hand movement and that of the object [35]. The combination of indirect pointing and the direct movement mapping has been explored to successfully enable bi-manual 3D manipulation in XR [26]. Compared to Hand Ray, Gaze&Pinch enables distant interaction without requiring the hand to point at the object. This property enables users to move the hand to extreme positions to anticipate a complex manipulation task without hand clutching. In this work, we evaluate how the different requirements for the position and orientation of the pinching hand for acquisition using Virtual Hand, Hand Ray, and Gaze&Pinch affect the incorporation of an alternative control mapping from the 6DOF hand movement to separate 3DOF translation and rotation of the object.

2.2 DOF-Separation for 3D object manipulation in XR

Moving and orienting virtual objects in the 3D space is a unique affordance of XR and a challenge for interaction design due to the complexity of the task [10, 17, 19, 32]. Previous work has explored different approaches for easier manipulation, including *World In Miniature* metaphors that allow users to use smaller-scale proxies to manipulate distant objects [34, 36, 41], bi-manual input for two-point control [28, 26], varying the control-display gains for the control mapping between movement of hand and of object [14, 9], and the use of virtual handles, constraints, widgets, and gizmos to break down the 6DOF movement into lower DOF controls that are easier to perceive and perform [15, 12, 40, 31, 32, 42, 58] (for reviews, see [4, 30]).

The concept of mapping higher-DOF input to lower-DOF control has been explored and discussed in interaction design. For computer input devices in general, Jacob defined integrality as allowing users to manipulate all task dimensions at once, and separability as supporting the control of different dimensions individually [20]. For 3D manipulation, integrality and separability can be applied to the task space. A common approach is to apply 6DOF hand movement to drag virtual handles or gizmos in order to translate or rotate objects in single dimensions [40, 31, 32]. Evaluation studies of DOF-Separation have shown mixed results depending on the context. For instance, in an orientation task, users may be faster with a 1DOF technique than with a 3DOF technique without sacrificing precision [46]. However, studies of 3DOF [32] and 6DOF [21, 11] interactions consistently demonstrate that simultaneous manipulation enhances user performance, and is universally recommended [21, 19, 30].

Previous work has also explored applying DOF-Separation to the control space by mapping 6DOF hand movement to translation and rotation of the object concurrently, similar to how the MRTK Hand Ray uses hand position and orientation [8, 11]. In a review of 3D virtual object manipulation, Mendes et al. reported an observed lack of techniques that employ DOF-Separation for hand movement, while most techniques mimic physical world interactions using the Virtual Hand metaphor, despite suffering from human inaccuracy [30]. Previous work has found that DOF-Separation can

[1]https://support.apple.com/guide/apple-vision-pro

offer precise control over sub-tasks without interference from concurrent dimensions [51, 54, 59]. Specifically, the *AMP-IT* technique, using adaptive control-display gains, was found to benefit from DOF-Separation while retaining the benefit of direct manipulation [38]. Inspired by these work, we investigate the effect of DOF-Separation on manipulation performance of different techniques with varying requirement of the hand.

3 STUDY

In this section, we present a user study that is designed to answer the **Research Question: What are the different impacts of the separate mapping of 6DOF hand movement to concurrent 3DOF controls of translation and rotation on the performance of virtual object manipulation using Virtual Hand, Hand Ray, and Gaze&Pinch techniques?** We describe the design of a within-subject study featuring a docking task in virtual reality (VR), and the design and implementation of six techniques as conditions.

3.1 Study design

We employ a within-subject study design with two independent variables: Input-Metaphor (Virtual Hand, Hand Ray, and Gaze&Pinch) and DOF-Separation (Integrated and Separated) that derive $3 \times 2 = 6$ techniques as conditions. Previous work on manipulation commonly adopted docking tasks that encapsulate selection, translation, and rotation, which are the basic building blocks to form many complex interactions [39, 4, 30]. Docking tasks can be investigated separately for translation and rotation [52, 13, 46] or together to understand the suitability of integrated techniques [49, 55, 24]. Previous evaluations of multi-DOF manipulation typically involves docking tasks that allow *clutching*—using multiple consecutive gestures to complete a trial [57, 25, 38, 11]. Whereas the trial completion criterion in these studies is based on an accuracy threshold that helps quantify performance using completion time, it provides limited insight into the combined ease and accuracy afforded by the technique in a similar way to Fitts' Law tasks for evaluating selection [27]. To address this, we employ a docking task that can be completed using a single gesture without using an accuracy threshold. While measuring both speed and accuracy as the *goal of the evaluation*, our results emphasize users' ability to accurately manipulate using a single pinch [4].

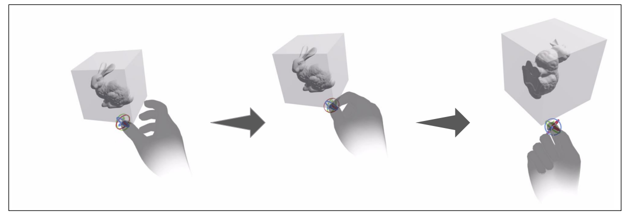
3.2 Design of Interaction Techniques

In this section we describe the interaction techniques used in the study. By joining Input-Metaphor and DOF-Separation, we derive six manipulation techniques: Integrated Virtual Hand (VH_I), Separated Virtual Hand (VH_S), Integrated Hand Ray (HR_I), Separated Hand Ray (HR_S), Integrated Gaze&Pinch (GP_I), and Separated Gaze&Pinch (GP_S). Following common practice and for consistency with Virtual Hand, we implement a visual-angle-based translation mapping for Hand Ray and Gaze&Pinch, such that the manipulated object travels the same angular distance around the user as the pinching hand.

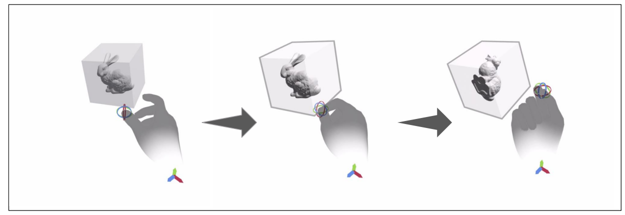
3.2.1 Virtual Hand

VH_I uses a 1:1 mapping between the 6DOF movement of the pinching hand to that of the object. Following common practice, we choose the tip of the thumb to represent the point of acquisition and map its changes in position and orientation to the translation and rotation of the manipulated object post-acquisition (Figure 2a).

For VH_S, we separate the translation mapping from the pinch point to the wrist joint to allow a more independent rotation without affecting translation. Hand rotation post-acquisition in VH_S is mapped to egocentric rotation of the object, reducing unintended translation of the object caused by e.g. wrist flexion. Additionally, to balance the effective virtual reach between the two techniques,



(a) DOF-Integrated Virtual Hand (VH_I).



(b) DOF-Separated Virtual Hand (VH_S).

Figure 2: Acquisition and manipulation using Virtual Hand. User reach towards the object and pinch to acquire it. When they manipulate: **(a)** the object translates following translation of thumb at pinch point (marked by translation gizmo), and rotate following the orientation of the thumb also at pinch point (marked by rotation gizmo) **(b)** the object translates following translation of wrist joint (marked by translation gizmo), and rotates following the orientation of the thumb at pinch point (marked by rotation gizmo).

we compensate for the different translation origin of VH_S. A forward offset, matching the distance from the wrist to the thumb, is applied when computing the CD-mapping, to ensure that the object remains at the pinch-point in the absence of rotation, consistent with VH_I (Figure 2b).

3.2.2 Hand Ray

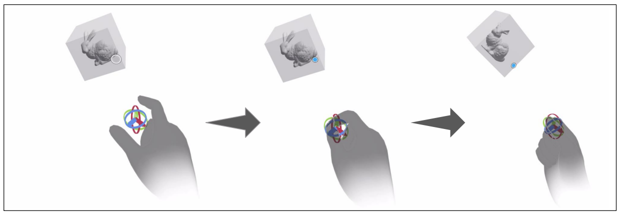
We implement HR_I following the example of the Hand Ray input from the Meta XR SDK [33], where both translation and rotation mappings are integrated at the base of the hand. This can be considered as a "stick" metaphor via a wrist-based ray, where only the pronation and supination of the hand is applied exactly to the object as "roll" through the acquisition point. The two remaining rotation DOFs of hand movement determine the pointing direction and are mapped to object translation during manipulation after acquisition. Visual feedback for pointing and acquisition is rendered through a circular cursor on the object that shrinks in radius upon pinching, and a small directional ray at the hand, following the implementation of Quest devices [33] (Figure 3a).

HR_S follows the implementation style and feedback of the Hand Ray defined in MRTK [8] and Leap Motion [44] by casting the ray from the midpoint of the estimated shoulder position and the wrist through the hand, thus freeing the hand from the pointing task for acquisition. Upon pinch-acquisition, changes in orientation of the hand relative to the wrist joint are mapped to rotation control of the object, without affecting its position (Figure 3b). For HR_S, we render the ray to help users understand the pointing direction due to its abstract origin and deliberate decoupling from the hand's orientation.

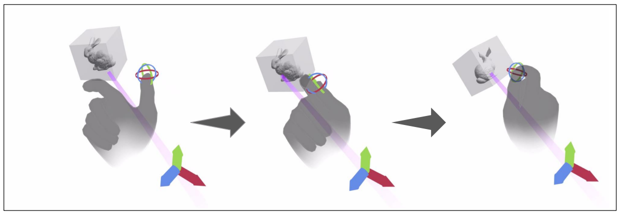
3.2.3 Gaze&Pinch

We implemented GP_I following common practice as described by Pfeuffer et al. [35] and by Apple[2]. By using gaze for pointing, the hand is free to pinch to confirm acquisition and to initiate manipulation from any point in reachable space (Figure 4a). This feature

[2]https://support.apple.com/guide/apple-vision-pro

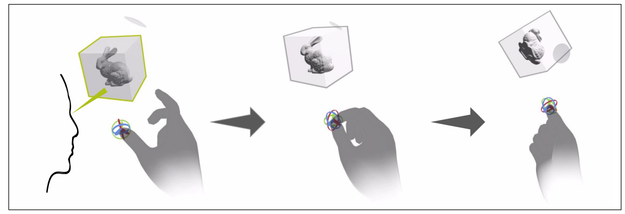


(a) DOF-Integrated Hand Ray (HR_I).

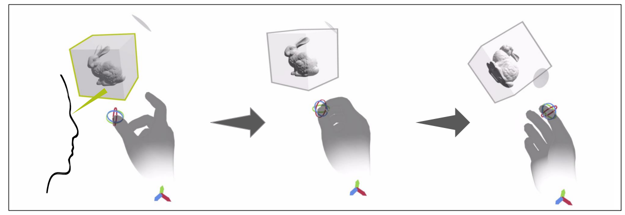


(b) DOF-Separated Hand Ray (HR_S).

Figure 3: Acquisition and manipulation using Hand Ray. **(a)** User points towards the object using virtual ray originating from the hand and changes direction following change in hand orientation. The grey cursor indicates the current pointer on the object surface, where a pinch confirms acquisition, upon which the cursor turns blue. During manipulation, the object translates and rotates as if the user is holding it by a stick represented by the ray (marked by translation and rotation gizmos) **(b)** User points towards the object using a ray originating from the shoulder and points through the hand. During manipulation, the object translates as if the user is holding it by a stick originating at the ray origin (marked by translation gizmo), and rotates following the rotation of the thumb at pinch point (marked by rotation gizmo).



(a) DOF-Integrated Gaze&Pinch (GP_I).



(b) DOF-Separated Gaze&Pinch (GP_S).

Figure 4: Acquisition and manipulation using Gaze&Pinch. User's gaze on the object triggers the rendering of a yellow outline and a virtual disc around the object. A pinch gesture confirms the acquisition of the gaze-pointed object. During manipulation, **(a)**: the object translates following the translation of the thumb at pinch point (marked by translation gizmo), and rotates following the rotation of the thumb at pinch point (marked by rotation gizmo); and **(b)**: the object translates following the translation of wrist joint (marked by translation gizmo), and rotates following the rotation of the thumb at pinch point (marked by rotation gizmo).

of acquisition allows the upper limb to perform a bigger range of preparatory movement needed for the manipulation task. However, because the hand is away from the object, it is unclear how the user can comprehend how the orientation of their hand relative to the object may affect the preparation for rotation. Previous work suggested that the spatial referencing of the hand during Gaze&Pinch manipulation is important for its performance [26]. For this reason, we designed a visual feedback system to compensate for the lack of visual access of the (representation of) hand in comparison to Virtual Hand and Hand Ray. Upon gaze-hover in Gaze&Pinch, the object is highlighted, and the user sees an indicator of the hand relative to the object in the form of an abstract curved disc. This disc rotates around the object following the change in orientation of the hand, representing the palm during natural prehension. The disc responds to pinching by shrinking, similarly to the HR_I cursor.

The control mapping in GP_S is analogous to VH_S and HR_S, with changes in hand orientation mapped to egocentric rotation of the object after acquisition, without causing translation (Figure 4b).

3.3 Task

Following suggestions and examples from previous work, we employ a docking task in VR to understand users' ability to accurately manipulate virtual objects using the six techniques [30, 4]. To understand the effect of Input-Metaphor and DOF-Separation on different components and difficulty levels of the manipulation task, we adopt three types of tasks: Translation, Rotation, and Combined translation and rotation (Figure 5). In each task, the user needs to acquire an object with their dominant hand and then dock the object at a matching target [4].

We designed the object for manipulation as a white semi-transparent cube inscribing an opaque Stanford bunny [56]. A trial starts upon object appearance and ends on pinch release. The cube has a side length of 12.5 cm and appears in reachable distance 50 cm in front of the participant's chest level, approximated based on the position of the VR headset. At the beginning of each trial, the object and the target appears overlapping, then the target is animated to its final destination through linear interpolation in translation and rotation. This presents to a user an optimal integral solution to the trial, minimising potential task difficulty caused by visual clarity and mental effort for rotation planning.

For the Translation task, only the position of the target is changed relative to the object while they face the same direction in global coordinates. The target destination is at a ± 30 cm offset along a single axis. Each axis (X,Y,Z) is tested with both positive and negative Euclidean distance offset in position.

For the Rotation task, only the rotation of the target is different from that of the object with a $\pm 45°$ angular offset, while they overlap in position at their pivots (centre of cube). Each axis (X,Y,Z) is tested with both positive and negative angular offset in rotation.

For the Combined task, both the translation and rotation of the target is different from those of the object. The target destination is at a ± 30 cm offset along the X axis with $\pm 45, 90, 135°$ angular offsets along one or two axes. Each of the 18 combinations of angular offset and axes (X,Y,Z,XY,XZ,YZ) is tested, while the direction of both translation and rotation is chosen uniformly random.

The sequence of trials for each task were chosen uniformly random, resulting in 3 INPUT-METAPHOR $\times$ 2 DOF-SEPARATION $\times$ 2 REPETITIONS $\times$ (6 TRANSLATIONS $+$ 6 ROTATIONS $+$ 18 COMBINED) $=$ 360 trials per participant.

3.4 Participants

We recruited 18 participants (12 M, 6 F) from the local area with age ranged from 22 to 30 ($M = 25.33, SD = 2.35$). All participants were right-handed and had normal or corrected vision. On a scale between 1 (low) and 5 (high), participants rated themselves as hav-

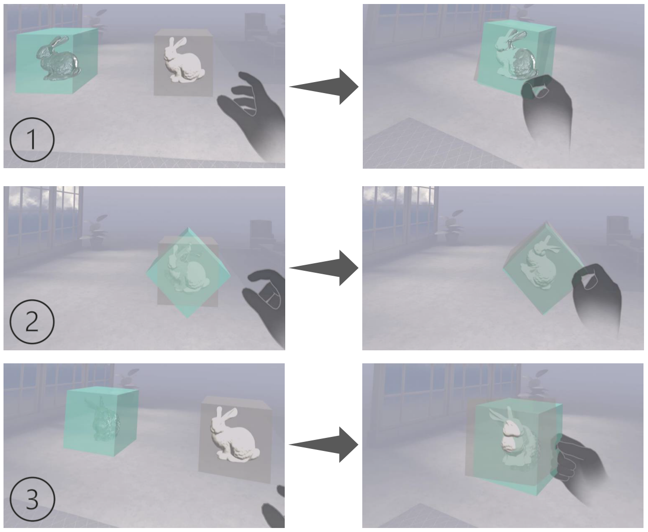


Figure 5: Examples of trials in the docking task: The white (object for manipulation) and blue (target position) cubes appear initially overlapping at the origin point. The target immediately moves to its final position and orientation via a linear interpolation animation. In Translation tasks (**1**), the target is offset ± 30 cm along either the X,Y, or Z axis from the object. In Rotation task (**2**), the target is offset $\pm 45°$ along the X,Y, or Z axis from the object. For Combined tasks (**3**), the target is offset ± 30 cm along the X axis, and $\pm 45, 90, 135°$ along either X,Y,Z,XY,XZ, or YZ relative to the object. Participants then pinch to acquire the object, manipulate the object to the target position, and release pinch to end the trial.

ing medium experience with VR/AR ($M = 2.89, SD = 1.23$), 3D hand gestures ($M = 2.56, SD = 1.34$), and medium-low experience with eye-gaze ($M = 2.28, SD = 1.13$).

3.5 Apparatus and implementation

The study is implemented with Unity (2022.3.27f1) for the Meta Quest Pro (90 Hz, 30 Hz eye tracker) using the Meta XR SDK v67. VH_I and HR_I implementations are based on defaults from the SDK, while the remaining 4 techniques rely on the SDK for tracking data. Eye-tracking accuracy varies from 1.5° to 3° [50, 2] and all techniques use the same selection collider, 1:1 with visuals. Hand tracking data was filtered through a 1€ Filter [7] for all techniques. Pinch-detection was tweaked, using the SDK's pinch strength parameter, towards early acquisition rather than early release - meaning that a participant encounters fewer frustrating early releases, but may encounter more early acquisitions and late releases, based on how the system evaluates the tracked hands. This detail is important as the tasks necessitates complex hand movement that challenge the tracking system.

3.6 Procedure

Participants were briefed and completed consent and demographics forms. Afterwards, and before each condition, two 1-minute videos of the techniques and the tasks were shown to understand the study elements. Participants then wore the headset and underwent fit adjustment and eye-tracking calibration. Throughout the study, users remained seated in a fixed chair that only afforded the movement of their upper body. For each condition the participant underwent fitting calibration and were allowed training in a task-less environment. They would manipulate the object using the current technique until comfortable in its use. The participant would then solve 12 Translation trials, 12 Rotation trials and finally 36 Combined trials, in this order. They were instructed to be as fast and as accurate as possible. Between each task the participants could take a break. After each condition (technique), participants completed a post-condition questionnaire. After finishing all conditions, the participants completed a final post-study questionnaire. The study took 80 min on average.

3.7 Evaluation metrics

We include the following measures.

- Positional Offset: The offset between the manipulated object and the target centres in Euclidean distance, representing translation accuracy.
- Rotational Offset: The angular offset between the manipulated object and the target aggregated along all 3 axes, representing rotation accuracy.
- Trial Completion Time: Time from start (object appears) to end (pinch release).
- Acquisition Time: Time from start (object appears) to acquisition (pinch down).
- NASA-TLX: Task load was measured after each condition using the NASA-TLX questionnaire [16].
- Post-Study User Feedback: After completing all conditions, the participants scored each of the six techniques in terms of preference and provided feedback about their choices.
- Observations: During the study, the experimenter observed participant behaviour.

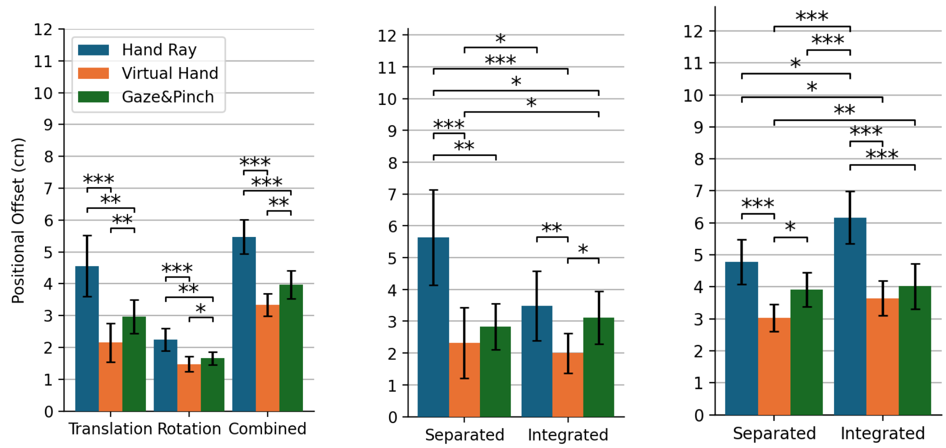
4 RESULTS

For performance data, we filtered outliers (99 trials, 1.5%) by Trial Completion Time when exceeding $Mean \pm 3 \cdot SD$. Furthermore, due to data loss, 16 out of the expected 6480 trials (0.2%) are missing. For the selection performance data, we applied the Aligned Rank Transform (ART) due to persisting deviations from normality [53]. Next, we performed a repeated measures ANOVA with the task performance, and post hoc pairwise comparisons (Homl-Bonferroni corrected). We present the task performance results in Figures 6 to 8, in which the error bars indicate 95% confidence interval and statistical significance is shown as * for $p < .05$, ** for $p < .01$, and *** for $p < .001$. We mark the significant main effect of Input-Metaphor and interaction effects on Input-Metaphor x DOF-Separation. For NASA-TLX and user ratings, we performed a Kruskal-Wallis Test with post-hoc Dunn tests (Holm-Bonferroni corrected) [29]. We present the TLX results in Figure 9 and subjective ratings in Figure 10, with significant effects marked.

4.1 Positional Offset (Figure 6)

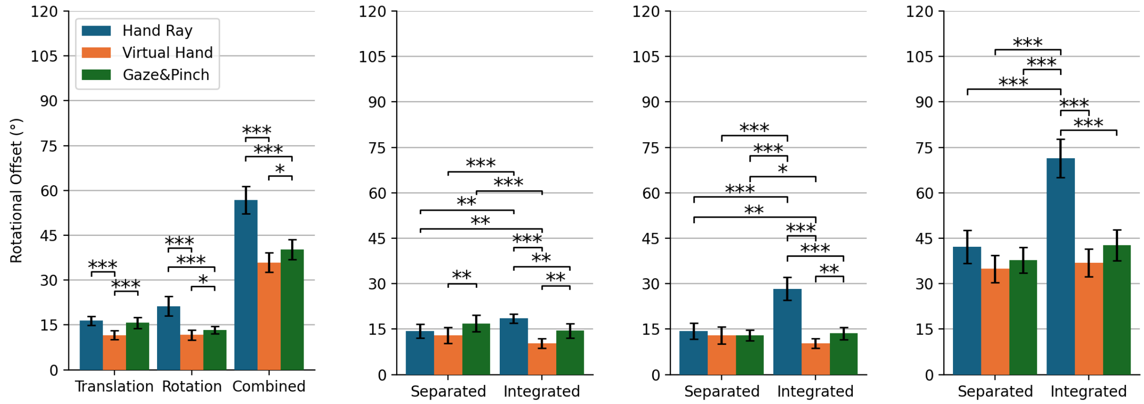
For Translation task, we found significant effects in Input-Metaphor ($F_{2,85} = 20.85, p < 0.001$), DOF-Separation ($F_{1,85} = 5.41, p = 0.022$), and Input-Metaphor x DOF-Separation ($F_{2,85} = 5.29, p = 0.007$). Post hoc comparisons show that Hand Ray induced more Positional Offset than both Virtual Hand ($p < 0.001$) and Gaze&Pinch ($p = 0.003$), and Gaze&Pinch induced more Positional Offset than Virtual Hand ($p = 0.002$). Furthermore, interaction analyses revealed that HR_S induced more Positional Offset for Translation than the other Input-Metaphors across both levels of DOF-Separation (all $p \leq 0.024$) (i.e., excluding HR_I). Additionally, VH_S induced less Positional Offset than HR_I ($p = 0.009$) and GP_I ($p = 0.038$), while VH_I induced less Positional Offset than HR_I ($p = 0.007$) and GP_I ($p = 0.024$).

For Rotation task, we found significant effects in Input-Metaphor ($F_{2,85} = 16.9, p < 0.001$) and DOF-Separation ($F_{1,85} = 13.92, p <$



(a) Input-Metaphor. (b) Translation Task. (c) Combined Task.

Figure 6: Positional Offset between target position and the final position of the object upon release of pinch for each technique. Overall, the offset with Hand Ray is larger than Gaze&Pinch, while Virtual Hand yielded the lowest offset. This trend is consistent across all tasks.



(a) Input-Metaphor. (b) Translation Task. (c) Rotation Task. (d) Combined Task.

Figure 7: Rotational Offset between target orientation and the final orientation of the object upon release. Separated techniques yielded similar Rotational Offsets. Virtual Hand yielded the lowest offsets, while HR_I is the least accurate, performing significantly worse than HR_S.

0.001). Post hoc comparisons show that Hand Ray induced more Positional Offset than Gaze&Pinch ($p = 0.001$), which measured worse than Virtual Hand ($p = 0.029$).

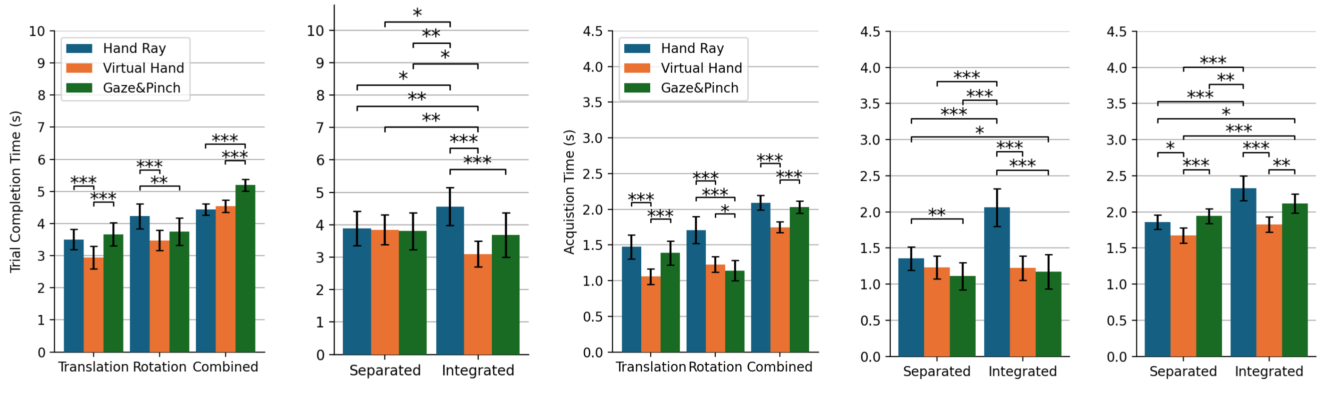
For Combined task, we found significant effects in Input-Metaphor ($F_{2,625} = 30.19, p < 0.001$), DOF-Separation ($F_{1,625} = 12.13, p < 0.001$), and Input-Metaphor x DOF-Separation ($F_{2,625} = 4.71, p = 0.009$). Post hoc comparisons revealed that Hand Ray induced more Positional Offset than both Virtual Hand ($p < 0.001$) and Gaze&Pinch ($p < 0.001$), and Gaze&Pinch induced more Positional Offset than Virtual Hand ($p = 0.002$). Furthermore, interaction analyses revealed that HR_I induced significantly more Positional Offset than all other combinations of Input-Metaphor and DOF-Separation (all $p \leq 0.041$), while HR_S induced significantly more Positional Offset than both VH_S ($p < 0.001$) and VH_I ($p = 0.017$), resulting in most of the overall difference in Input-Metaphor. VH_S induced significantly less Positional Offset than GP_S ($p = 0.022$) and GP_I ($p = 0.008$).

4.2 Rotational Offset (Figure 7)

For the Translation task, we found significant effects in Input-Metaphor ($F_{2,85} = 25.09, p < 0.001$) and Input-Metaphor x DOF-Separation ($F_{2,85} = 10.87, p < 0.001$). Post hoc comparisons revealed that Virtual Hand induced less Rotational Offset than both Hand Ray ($p < 0.001$) and Gaze&Pinch ($p < 0.001$). Furthermore, post hoc analyses on interaction effects revealed that HR_S induced less Rotational Offset than HR_I ($p = 0.003$), but more Rotational Offset than VH_I ($p = 0.005$). VH_S induced less Rotational Offset than HR_I ($p < 0.001$), while GP_S induced more Rotational Offset than VH_I ($p < 0.001$). Additionally, VH_I induced less Rotational Offset that both HR_I ($p < 0.001$) and GP_I ($p = 0.006$). GP_I induced less Rotational Offset than HR_I ($p = 0.002$), while VH_S less Rotational Offset than GP_S ($p = 0.003$).

For the Rotation task, we found significant effects in Input-Metaphor ($F_{2,85} = 49.5, p < 0.001$), DOF-Separation ($F_{1,85} = 34.39, p < 0.001$), and Input-Metaphor x DOF-Separation ($F_{2,85} = 46.5, p < 0.001$). Post hoc comparisons show that Hand Ray induced more Rotational Offset than both Virtual Hand ($p < 0.001$) and Gaze&Pinch ($p < 0.001$), while Virtual Hand induced less Rotational Offset than Gaze&Pinch ($p = 0.011$). Furthermore, post hoc comparisons on interaction effects revealed that HR_I induced more Rotational Offset than all other combinations of Input-Metaphor and DOF-Separation (all $p < 0.001$), resulting in most of the overall difference between Input-Metaphor. Additionally, VH_I induced less Rotational Offset than GP_I ($p = 0.004$), GP_S ($p = 0.023$), and HR_S ($p = 0.002$).

For the Combined task, we found significant effects in Input-



(a) Input-Metaphor. (b) Rotation Task. (c) Input-Metaphor. (d) Rotation Task. (e) Combined Task.

Figure 8: Trial Completion Time (a-b): time taken to complete trials. Acquisition Time (c-e): time taken to acquire the object. Both measures rises with rotational complexity.

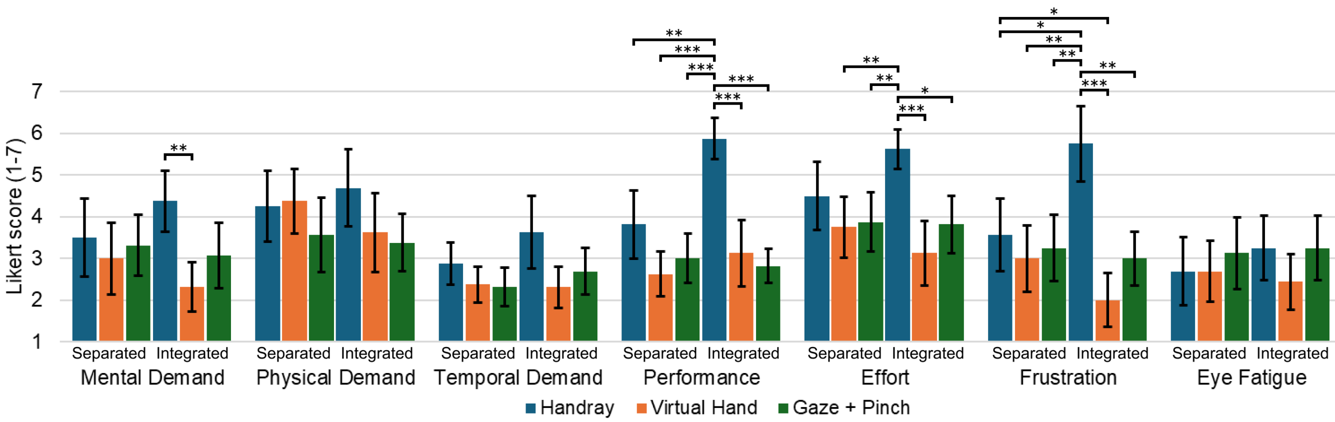


Figure 9: NASA-TLX ratings. HR_I was perceived as more effort-inducing, frustrating, and less performant than the other techniques.

Metaphor ($F_{2,625} = 38.49, p < 0.001$), DOF-Separation ($F_{1,625} = 41.87, p < 0.001$), and Input-Metaphor x DOF-Separation ($F_{2,625} = 21.88, p < 0.001$). Post hoc comparisons revealed that Hand Ray induced more Rotational Offset than both Virtual Hand ($p = 0.001$) and Gaze&Pinch ($p < 0.001$), while Virtual Hand induced less Rotational Offset than Gaze&Pinch ($p = 0.021$). Furthermore, interaction effect analyses revealed that HR_I induced more Rotational Offset than all other combinations of Input-Metaphor and DOF-Separation (all $p < 0.001$), resulting in all of the overall differences in Input-Metaphor.

4.3 Trial Completion Time (Figure 8a-b)

For the Translation task, we found significant effects in Input-Metaphor ($F_{2,85} = 19.42, p < 0.001$) and DOF-Separation ($F_{1,85} = 10.29, p = 0.002$). Post hoc comparisons revealed that Virtual Hand was faster than both Hand Ray ($p < 0.001$) and Gaze&Pinch ($p < 0.001$). For the Rotation task, we found significant results for Input-Metaphor ($F_{2,85} = 12.73, p < 0.001$) and Input-Metaphor x DOF-Separation ($F_{2,85} = 11.39, p < 0.001$). It showed that Hand Ray was slower than both Virtual Hand ($p < 0.001$) and Gaze&Pinch ($p = 0.001$). Furthermore, post hoc comparisons revealed that HR_I was significantly slower than all other combinations (all $p \leq 0.038$), while VH_I was faster than both HR_S ($p = 0.004$) and GP_S ($p = 0.014$), showing that the major overall differences in Input-Metaphor stems from differences in DOF-Separation between the Integrated techniques (HR_I, VH_I, and GP_I). For the Combined task, we found significant effects in Input-Metaphor ($F_{2,625} = 31.3, p < 0.001$) and DOF-Separation ($F_{1,625} = 21.01, p < 0.001$). Post hoc comparisons show that Gaze&Pinch was slower than both Hand Ray ($p < 0.001$) and Virtual Hand ($p < 0.001$).

4.4 Acquisition Time (Figure 8c-e)

For the Translation task, analysis of Input-Metaphor ($F_{2,85} = 21.83, p < 0.001$) revealed that acquisition was faster with Virtual Hand than both Hand Ray ($p < 0.001$) and Gaze&Pinch ($p < 0.001$). For the Rotation task, we found significant effects in Input-Metaphor ($F_{2,85} = 48.61, p < 0.001$), DOF-Separation ($F_{1,85} = 27.77, p < 0.001$), and Input-Metaphor x DOF-Separation ($F_{2,85} = 21.66, p < 0.001$). Acquisition was slower with Hand Ray than both Virtual Hand ($p < 0.001$) and Gaze&Pinch ($p < 0.001$), and Virtual Hand was slower than Gaze&Pinch ($p = 0.038$) As well, HR_I was slower than all other techniques (all $p < 0.001$). We also found that HR_S was similarly significantly slower than both GP_I ($p = 0.012$) and GP_S ($p = 0.004$), resulting in most of the overall difference in Input-Metaphor. For the Combined task, Input-Metaphor ($F_{2,625} = 28,36, p < 0.001$), DOF-Separation ($F_{1,625} = 33.14, p < 0.001$), and Input-Metaphor x DOF-Separation ($F_{2,625} = 4.7, p = 0.009$) were significant. We find that acquisition was faster with Virtual Hand than both Hand Ray ($p < 0.001$) and Gaze&Pinch ($p < 0.001$). In particular, HR_I was slower than all Separated combinations (i.e. HR_S, VH_S, and GP_S) (all $p \leq 0.003$), while GP_I was slower than both HR_S ($p = 0.036$) and VH_S ($p < 0.001$). Virtual Hand was faster than the other integrated Input-Metaphors (all $p \leq 0.012$).

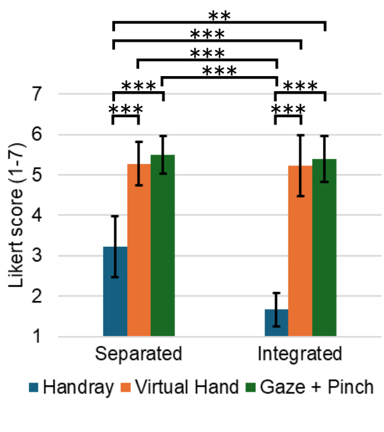


Figure 10: User preference rating for different techniques. Hand Ray was rated as the least favourite, especially so for the integrated variant HR_I. No significant differences was found for preferences between Gaze&Pinch and Virtual Hand.

4.5 NASA-TLX (Figure 9)

Significant differences were indicated in Mental Demand ($H(5) = 14.69, p = 0.012$), Performance ($H(5) = 40.36, p < 0.001$), Effort ($H(5) = 25.55, p < 0.001$), and Frustration ($H(5) = 37.52, p < 0.001$). Participants perceived HR_I as more mentally demanding than VH_I ($p = 0.004$). HR_I was rated as least performant (all $p \leq 0.007$) and more frustrating (all $p \leq 0.017$) to use than all other conditions, and rated more effortful than all other combinations except for HR_S (all $p \leq 0.01$).

4.6 User Feedback (Figure 10)

The preferences results for techniques ($H(5) = 57.26, p < 0.001$) showed that Hand Ray is least preferred, but without significant differences in scores between Gaze&Pinch and Virtual Hand variations. User feedback reveals common themes across the techniques. For Virtual Hand, seven participants noted that it felt natural (or similar descriptions), but three (VH_I) and five (VH_S) participants noted rotation difficulties, while four participants reported arm fatigue across both variants. This aligns with prior results on added physical effort in direct input in contrast to indirect methods [47, 48, 26]. For Hand Ray, six participants mentioned that HR_I felt difficult for rotation, with HR_S being comfortable for this purpose. Yet, with HR_S, four participants stated that translation is poor, and five reported that it does not feel natural to use it. With Gaze&Pinch, six participants noted the ease to understand and control with GP_I, and four stated comfortable use with GP_S. Three participants noted wrist fatigue with GP_I, while GP_S was reported to feel harder for positioning by six.

4.7 Observations

In general, Virtual Hand techniques imposed more movement of the upper body, motivating participants to lean forwards and extend their reach for solving translation and combined tasks comfortably, which was especially noticeable for participants with shorter reach. Remarkably, the HR_I technique motivated participants to initiate manipulation from eccentric positions.With HR_I, unusual configuration of the upper limb and shifting of the torso was observed for 12 participants, for acquiring the object from either side or top surfaces. All other techniques allowed participants to remain reclined, while Gaze&Pinch Techniques allowed the participants' arm to remain on the arm rest.

HR_S motivated participants to engage their shoulders and arm for solving the pointing-for-acquisition subtask. Both Hand Ray techniques seemed to enforce fatiguing arm positions and distract the user from the task. Participants tended to acquire the object with a neutral wrist (in default hand orientation), limiting subsequent rotation. Some participants seemed to struggle relatively more with acquisition using Hand Ray compared to the other techniques.

Most participants (16) knew intuitively, or discovered through training, the importance of pre-shaping their arm and hand. Every participant mentioned the acquisition strategy related to the task they were solving—11 explicitly referring to the preparatory movement of the hand. Those who invested more time in planning and coordinating their movements prior to acquisition generally achieved greater accuracy and minimized eccentric motions. Furthermore, manipulations executed after establishing an anticipatory hand and wrist configuration typically addressed translation and rotation in parallel. Acquisition with a neutral wrist more often led to participants addressing the translation sub-task before rotation. All participants engaged in more careful acquisition during the Combined task, especially when the rotation became more complex (e.g. larger angular offset or rotation along 2 axis).

Participants would adapt their strategy to each of the tasks, with a few exceptions. For planning the acquisition, participants tend to spend time looking at the target. Seven participants exhibited difficulty in selecting targets using Gaze&Pinch techniques, often looking ahead to the target before acquiring the object (Late-Trigger issue [22]). This issue was particularly evident for complex trials. Pinch-detection was observed as a major cause of inaccuracy and frustration. Participants with shorter arms experienced this often.

5 Discussion

Through the user study, we aim to understand the impact of DOF-Separation when applied to different input techniques for 3D manipulation of virtual objects in XR. In this section, we focus on synthesising results on interaction effects observed from the study to understand how DOF-Separation yielded better or worse speed and accuracy performance with each Input-Metaphor.

5.1 Hand Ray and DOF-Separation

From nearly all performance measures, we can easily observe the significant worse performance yielded by HR_I. Because of the use of the "stick" metaphor that tightly couples the change in hand orientation post-acquisition to the displacement of the object, the user's capability of manipulating the object is significantly limited. For instance, while it is nearly impossible to rotate an object 90°around the Y axis in-place using HR_I without walking around the object, this task can be easily achievable using HR_S while benefiting from the capability of using wrist flexion to only induce local rotation of the object. This separated mapping of the 6DOF hand movement to the translation and rotation of the object post-acquisition frees the hand from the pointing task, while enabling more flexibility in manipulation. Despite being featured in the Meta XR SDK as the default distant interaction paradigm for the Quest devices [33], HR_I is significantly limited in usability comparing with the HR_S approach adopted by earlier devices, including the HoloLens series [8]. The difference between HR_I and HR_S clearly illustrate the effect of DOF-Separation on 3D input techniques for combined acquisition and manipulation.

5.2 Virtual Hand and DOF-Separation

In most performance measures, Virtual Hand outperformed the other methods, indicating the benefit of the intuitiveness of using a natural metaphor that mimics real-world physics. While Hand Ray yielded significantly worse performance in most measures due to the Integrated version, we focus on the comparison between Virtual Hand and Gaze&Pinch to reveal deeper insights into the effect of DOF-Separation, thanks to the benefit of Gaze&Pinch that decouples the hand completely from the acquisition task prior to manipulation. While Virtual Hand require the hand's position to overlap with that of the object, comparing these two metaphors can tell us more about DOF-Separation in light of extreme differences in the requirement for the hand.

The results of Positional Offset shows that while GP_I yielded worse measures than both Virtual Hand versions in the Translation task, VH_S yielded better measures than both versions of Gaze&Pinch in the Combined translation and rotation task. These results suggest that Virtual Hand techniques may be better for the translation component of manipulation tasks due to the consistent visual and proprioceptive feedback of the hand coupled with the object comparing with Gaze&Pinch. Moreover, we posit that VH_S yielded significantly better performance in Positional Offset in Combined tasks, which are theoretically more difficult, because the DOF-Separation allowed the hand to more freely prepare for the manipulation task during acquisition using wrist flexion. This benefit is observed only in Combined tasks that feature more extreme rotation angles because the extra freedom at the wrist afforded by DOF-Separation is only needed in these scenarios, whereas it may not be easily reflected in easier tasks.

Notably, VH_S is also significantly faster than both Gaze&Pinch techniques for acquisition in Combined tasks. This indicates that participants may have needed less time for preparatory hand posture before acquisition, thanks to the freedom in hand orientation afforded by DOF-Separation in comparison to VH_I, which may demand them to configure the hand in more extreme positions to prepare for the same manipulation task. We also observe that for Rotational Offset, VH_I yielded better measures than both Gaze&Pinch techniques in Translation and Rotation tasks, respectively. However, this effect is absent in Combined tasks, and for which VH_I yielded observably worse (though non-significant) performance than VH_S. This may suggest that the increased difficulty in the manipulation task favours DOF-Separation more while diminishing the benefit of Virtual Hand techniques, especially VH_I, due to the close coupling of hand orientation and object displacement.

Overall, the results suggest that though VH_I has an Input-Metaphor that is closer to real-world physics with a direct mapping between the 6DOF hand movement to that of the object in pinch-manipulation, DOF-Separation still benefits performance for more complex manipulation tasks that involve more extreme rotation angles. Based on these results, we suggest that it is plausible to feature DOF-Separated versions of Virtual Hand techniques for XR applications that involve frequent performances of difficult manipulation tasks, such as 3D design and modelling.

5.3 Gaze&Pinch and DOF-Separation

While the results suggest that Virtual Hand may have benefitted from DOF-Separation in more difficult manipulation tasks, this effect was not as pronounced with Gaze&Pinch techniques. While we did not find consistent patterns of difference between GP_I and GP_S, the performance of Gaze&Pinch is mostly worse than but still comparable with that of Virtual Hand techniques. Though Virtual Hand techniques benefit from the realistic physics metaphor and the visibility of the hand coupled with the object for guiding manipulation, Gaze&Pinch enable usable performance of acquisition and manipulation in distant space that is inaccessible with Virtual Hand. Further, Gaze&Pinch is the technique that has the least requirement for the hand for acquisition as the pinch can be performed from anywhere while the pointing task is delegated to gaze input. We posit that this may be a main reason why the effect of DOF-Separation is limited on Gaze&Pinch as users would be able to freely prepare their hands for acquisition to anticipate the manipulation tasks with or without DOF-Separation in the mapping of hand movement.

5.4 Limitations & Future Work

This study used hand-tracking provided by Quest Pro to represent the state of the art. However, pinch detection errors led to cases of premature trial endings. More robust hand tracking that may become available in the future could help increase the validity of the obtained results. We measured task completion time to reflect users' capability of performing acquisition and manipulation using the described techniques. However, we observed that in some cases when the tasks were too difficult, participants tended to quickly drag and release the objects to move on to the next trial while giving up on accuracy, typically with HR_I. Future research could explore alternative study designs that allow clutching of the hand to understand if the potential benefits of different techniques can be directly reflected by task completion time. Lastly, we experimented with DOF-Separation on the hand and forearm following existing practice, whereas the anatomy of human hands may offer more options to separate movement control to other joints that could be explored.

6 Conclusion

In this work, we presented a study that contributes novel understandings of single-hand pinch-manipulation of virtual objects by investigating the interplay between Input-Metaphor and DOF-Separation of the 6DOF hand movement into 3DOF controls of translation and rotation. We implemented DOF-integrated and DOF-separated versions of Virtual Hand, Hand Ray, and Gaze&Pinch techniques, and compared them in a VR docking task with varying complexity levels. Our results suggests that DOF-Separation significantly benefits Hand Ray in all tasks by freeing the hand orientation from controlling translation of the object while only mapping it to the 3DOF rotation control. For Virtual Hand, similar performance benefits are found in more complex tasks where the extra freedom of hand orientation is helpful, but not for all tasks. Gaze&Pinch is less affected by DOF-Separation due to its relaxed requirement for the hand at acquisition, either regarding its position or orientation. Our results offer insights into the interplay between the DOF-Separation of 6DOF hand movement and the popular 3D input techniques in XR that have different requirement of the hand's position and orientation at acquisition that affect the subsequent manipulation task. Based on our findings, we recommend DOF-Separation for Hand Ray and for Virtual Hand in use cases where frequent complex manipulation tasks are expected, such as in 3D designing applications.

Acknowledgments

This work was supported by funds from a Google Research Gift award ('Multimodal and Gaze + Gesture Interactions in XR'), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant no. 101021229 GEMINI), and the Danish National Research Foundation under the Pioneer Centre for AI in Denmark (DNRF grant P1).

References

[1] Arkio. https://www.arkio.is. [Online; accessed 8-6-2025]. 1

[2] S. Aziz, D. J. Lohr, L. Friedman, and O. Komogortsev. Evaluation of eye tracking signal quality for virtual reality applications: A case study in the meta quest pro, 2024. 5

[3] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21(6):34–47, 2001. 1

[4] J. Bergström, T.-S. Dalsgaard, J. Alexander, and K. Hornbæk. How to evaluate object selection and manipulation in vr? guidelines from 20 years of studies. In *proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–20, 2021. 2, 3, 4

[5] M. Billinghurst, A. Clark, G. Lee, et al. A survey of augmented reality. *Foundations and Trends® in Human–Computer Interaction*, 8(2-3):73–272, 2015. 1

[6] 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*, I3D '97, p. 35–ff. Association for Computing Machinery, New York, NY, USA, 1997. doi: 10.1145/253284.253301 2

[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 5

[8] M. Cooperation. MRTK - Mixed Reality Toolkit's Hand Interaction Examples with HoloLens 2. https://www.youtube.com/watch?v=wogJv5v9x-s&t=56s, 2019. [Online; accessed 9-9-2024]. 1, 2, 3, 8

[9] L. Dominjon, A. Lécuyer, J.-M. Burkhardt, P. Richard, and S. Richir. Influence of control/display ratio on the perception of mass of manipulated objects in virtual environments. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 19–25. IEEE, 2005. 2

[10] T. Drey, J. Gugenheimer, J. Karlbauer, M. Milo, and E. Rukzio. Vrsketchin: Exploring the design space of pen and tablet interaction for 3d sketching in virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376628 2

[11] T. Drey, M. Montag, A. Vogt, N. Rixen, T. Seufert, S. Zander, M. Rietzler, and E. Rukzio. Investigating the effects of individual spatial abilities on virtual reality object manipulation. 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: 10.1145/3544548.3581004 2, 3

[12] J. Feng, I. Cho, and Z. Wartell. Comparison of device-based, one and two-handed 7dof manipulation techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, SUI '15, p. 2–9. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2788940.2788942 2

[13] S. Frees and G. D. Kessler. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 99–106. IEEE, Institute for Electrical and Electronics Engineers, New York, NY, USA, 2005. 3

[14] S. Frees, G. D. Kessler, and E. Kay. Prism interaction for enhancing control in immersive virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 14(1):2–es, may 2007. doi: 10.1145/1229855.1229857 2

[15] P. C. Gloumeau, W. Stuerzlinger, and J. Han. Pinnpivot: Object manipulation using pins in immersive virtual environments. *IEEE transactions on visualization and computer graphics*, 27(4):2488–2494, 2020. 2

[16] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52:139–183, 1 1988. doi: 10.1016/S0166-4115(08)62386-9 5

[17] D. Hayatpur, S. Heo, H. Xia, W. Stuerzlinger, and D. Wigdor. Plane, ray, and point: Enabling precise spatial manipulations with shape constraints. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, p. 1185–1195. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3332165.3347916 2

[18] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 1063–1072. ACM, New York, NY, USA, 2014. doi: 10.1145/2556288.2557130 2

[19] K. Hinckley, J. Tullio, R. Pausch, D. Proffitt, and N. Kassell. Usability analysis of 3d rotation techniques. In *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology*, UIST '97, p. 1–10. Association for Computing Machinery, New York, NY, USA, 1997. doi: 10.1145/263407.263408 2

[20] R. J. Jacob, L. E. Sibert, D. C. McFarlane, and M. P. Mullen Jr. Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(1):3–26, 1994. 2

[21] A. Kulik, A. Kunert, and B. Froehlich. On motor performance in virtual 3d object manipulation. *IEEE transactions on visualization and computer graphics*, 26(5):2041–2050, 2020. 2

[22] 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 8

[23] 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. 1, 2

[24] J. Lee, M. Sinclair, M. Gonzalez-Franco, E. Ofek, and C. Holz. Torc: A virtual reality controller for in-hand high-dexterity finger interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300301 3

[25] A. Leganchuk, S. Zhai, and W. Buxton. Manual and cognitive benefits of two-handed input: an experimental study. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 5(4):326–359, 1998. 3

[26] M. N. Lystbæk, T. Mikkelsen, R. Krisztandl, E. J. Gonzalez, M. Gonzales-Franco, H. Gellersen, and K. Pfeuffer. Hands-on, hands-off: Gaze-assisted bimanual 3d interaction. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST'24)*. Association for Computing Machinery, 2024. 2, 4, 8

[27] I. S. MacKenzie. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction*, 7(1):91–139, 1992. 3

[28] D. P. Mapes and J. M. Moshell. A two-handed interface for object manipulation in virtual environments. *Presence: Teleoper. Virtual Environ.*, 4(4):403–416, jan 1995. doi: 10.1162/pres.1995.4.4.403 2

[29] P. E. McKight and J. Najab. Kruskal-wallis test. *The corsini encyclopedia of psychology*, pp. 1–1, 2010. 5

[30] 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. In *Computer graphics forum*, vol. 38, pp. 21–45. Wiley Online Library, 2019. 1, 2, 3, 4

[31] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge. The benefits of dof separation in mid-air 3d object manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, p. 261–268. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2993369.2993396 2

[32] D. Mendes, M. Sousa, R. Lorena, A. Ferreira, and J. Jorge. Using custom transformation axes for mid-air manipulation of 3d virtual objects. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3139131.3139157 2

[33] Meta. Ray interactions. https://developers.meta.com/horizon/documentation/unity/unity-isdk-ray-interaction/, 2025. [Online; accessed 5-4-2025]. 2, 3, 8

[34] M. R. Mine, F. P. Brooks, and C. H. Sequin. Moving objects in space: Exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, p. 19–26. ACM Press/Addison-Wesley Publishing Co., USA, 1997. doi: 10.1145/258734.258747 2

[35] 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, 3

[36] J. S. Pierce, B. C. Stearns, and R. Pausch. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *Proceedings of the 1999 Symposium on Interactive 3D Graphics*, I3D '99, p. 141–145. Association for Computing Machinery, New York, NY, USA, 1999. doi: 10.1145/300523.300540 2

[37] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-

go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, UIST '96, p. 79–80. Association for Computing Machinery, New York, NY, USA, 1996. doi: 10.1145/237091.237102 2

[38] F. Rodrigues, A. Giovannelli, L. Pavanatto, H. Miao, J. C. de Oliveira, and D. A. Bowman. Amp-it and wisdom: Improving 3d manipulation for high-precision tasks in virtual reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 303–311. IEEE, 2023. 3

[39] R. Schmidt, K. Singh, and R. Balakrishnan. Sketching and composing widgets for 3d manipulation. In *Computer graphics forum*, vol. 27, pp. 301–310. Wiley Online Library, 2008. 3

[40] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 1297–1306. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208585 2

[41] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a wim: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, p. 265–272. ACM Press/Addison-Wesley Publishing Co., USA, 1995. doi: 10.1145/223904.223938 2

[42] J. Sun and W. Stuerzlinger. Extended sliding in virtual reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3359996.3364251 2

[43] 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, p. 4179–4188. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2702123.2702355 2

[44] P. Turner. Hand-powered far field interaction - how we did it. `https://docs.ultraleap.com/ultralab/far-field-ray-blog.html`, 2022. [Online; accessed 9-9-2024]. 1, 2, 3

[45] A. van Dam. Post-wimp user interfaces. *Commun. ACM*, 40(2):63–67, Feb. 1997. doi: 10.1145/253671.253708 1

[46] M. Veit, A. Capobianco, and D. Bechmann. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST '09, p. 51–58. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1643928.1643942 2, 3

[47] U. Wagner, A. Jacobsen, T. Feuchtner, H. Gellersen, and K. Pfeuffer. Eye-hand movement of objects in near space. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24)*. Association for Computing Machinery, 2024. 8

[48] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbæk, K. Pfeuffer, and H. Gellersen. A fitts' law study of gaze-hand alignment for selection in 3d user interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581423 8

[49] J. Wang and R. W. Lindeman. Object impersonation: Towards effective interaction in tablet- and hmd-based hybrid virtual environments. In *2015 IEEE Virtual Reality (VR)*, pp. 111–118. Institute for Electrical and Electronics Engineers, New York, NY, USA, 2015. doi: 10.1109/VR.2015.7223332 3

[50] S. Wei, D. Bloemers, and A. Rovira. A preliminary study of the eye tracker in the meta quest pro. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*, IMX '23, p. 216–221. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3573381.3596467 5

[51] D. Wigdor, H. Benko, J. Pella, J. Lombardo, and S. Williams. Rock & rails: Extending multi-touch interactions with shape gestures to enable precise spatial manipulations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 1581–1590. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1979173 3

[52] C. Wilkes and D. A. Bowman. Advantages of velocity-based scaling for distant 3d manipulation. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, VRST '08, p. 23–29. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1450579.1450585 2, 3

[53] 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. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963 5

[54] S. Wu, A. Chellali, S. Otmane, and G. Moreau. Touchsketch: A touch-based interface for 3d object manipulation and editing. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*, VRST '15, pp. 59–68. ACM, New York, NY, USA, 2015. doi: 10.1145/2821592.2821606 3

[55] J. J. Yang, H. Horii, A. Thayer, and R. Ballagas. Vr grabbers: Ungrounded haptic retargeting for precision grabbing tools. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, p. 889–899. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3242587.3242643 3

[56] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves. *Gaze-Supported 3D Object Manipulation in Virtual Reality*, pp. 1–13. Association for Computing Machinery, New York, NY, USA, 2021. 2, 4

[57] S. Zhai, P. Milgram, and W. Buxton. The influence of muscle groups on performance of multiple degree-of-freedom input. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 308–315, 1996. 3

[58] D. B. Zhao and W. Stuerzlinger. A novel bare-handed and widget-based manipulation technique for distant objects in virtual reality. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 525–528, 2024. doi: 10.1109/ISMAR-Adjunct64951.2024.00152 2

[59] Q. Zhou, D. Yu, M. N. Reinoso, J. Newn, J. Goncalves, and E. Velloso. Eyes-free target acquisition during walking in immersive mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3423–3433, 2020. doi: 10.1109/TVCG.2020.3023570 3