Gaze-Supported 3D Object Manipulation in Virtual Reality

Difeng Yu*
The University of Melbourne
difeng.yu@student.unimelb.edu.au

Xueshi Lu
Xi'an Jiaotong-Liverpool University
xueshi.lu17@student.xjtlu.edu.cn

Rongkai Shi
Xi'an Jiaotong-Liverpool University
rongkai.shi19@student.xjtlu.edu.cn

Hai-Ning Liang*
Xi'an Jiaotong-Liverpool University
haining.liang@xjtlu.edu.cn

Tilman Dingler
The University of Melbourne
tilman.dingler@unimelb.edu.au

Eduardo Velloso
The University of Melbourne
eduardo.velloso@unimelb.edu.au

Jorge Goncalves
The University of Melbourne
jorge.goncalves@unimelb.edu.au

ABSTRACT

This paper investigates integration, coordination, and transition strategies of gaze and hand input for 3D object manipulation in VR. Specifically, this work aims to understand whether incorporating gaze input can benefit VR object manipulation tasks, and how it should be combined with hand input for improved usability and efficiency. We designed four gaze-supported techniques that leverage different combination strategies for object manipulation and evaluated them in two user studies. Overall, we show that gaze did not offer significant performance benefits for transforming objects in the primary working space, where all objects were located in front of the user and within the arm-reach distance, but can be useful for a larger environment with distant targets. We further offer insights regarding combination strategies of gaze and hand input, and derive implications that can help guide the design of future VR systems that incorporate gaze input for 3D object manipulation.

CCS CONCEPTS

• **Human-centered computing → User interface design**; *User studies*; *Virtual reality*; *Interaction techniques*.

KEYWORDS

3D object manipulation, gaze input, multimodal interface

ACM Reference Format:
Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. Gaze-Supported 3D Object Manipulation in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3411764.3445343

*The corresponding authors from each institution are Difeng Yu and Hai-Ning Liang.

ACM ISBN 978-1-4503-8096-6/21/05...$15.00
https://doi.org/10.1145/3411764.3445343

1 INTRODUCTION

As one of the primary tasks in virtual reality (VR) systems, object manipulation is used in many different application domains such as 3D modeling [14, 20], game development [18], online collaboration [25, 38], and immersive data exploration [2, 8]. However, its primary input modality, which uses virtual hands to "direct manipulate" an object, has long been criticized for being inefficient and imprecise [7, 36], and likely to induce arm-fatigue in longer interaction scenarios [16, 29, 35].

Alternatively, gaze has been identified as a light-weight and fast input method, and has shown its potential for assisting with object manipulation tasks (e.g., [33, 40, 59]). However, previous work in VR mostly focused on the use of gaze for target selection [39, 45], which is only a sub-phase of the whole manipulation process, while how gaze input can be incorporated into the "manipulate" phase (translation, rotation, and scaling [31]) is still underexplored. Thus, this research aims to understand *whether* the incorporation of gaze input can benefit the hand manipulation process in VR, and *how* gaze input should be combined with hand input for convenient and efficient 3D object manipulation.

This research investigates different integration, coordination, and transition strategies when incorporating gaze into current systems with mid-air hand input for 3D object manipulation in VR. Specially, we examine a design space that considers how gaze and hand input are *integrated* into different phases of the manipulation task, how they *coordinate* with each other when starting the manipulation, and how to *transition* from one to the other during manipulation. Based on this design space, we developed four gaze-supported manipulation techniques and evaluated them through two user studies. In the first study, we focused on the *primary working space*, where all objects located in front of the user and were within arm-reach distance, and assessed the techniques in terms of user performance and experience. In the second study, we further evaluated our techniques in a larger virtual environment with distant objects and embedded the designed techniques into realistic workflows.

Our findings show that gaze might not offer significant performance benefits for transforming objects in the primary working space, but can be useful in a larger environment with distant targets, while also mitigating the arm fatigue issue. We further derived a set of design implications that reveal the usefulness of different strategies, including hand-only *vs.* eye-hand manipulation, direct *vs.*

remote hand mappings, and implicit *vs.* explicit eye-hand transitions. Our findings and implications provide a helpful guide for the design of future gaze-supported object manipulation techniques in VR.

To summarize, the main contributions of the paper include:

- The design space of how to incorporate gaze into the traditional hand-based object manipulation workflow.
- A novel implicit transition-based approach (called *ImplicitGaze*).
- The evaluation of the techniques, which has led to useful findings and design implications (whether it is beneficial to incorporate gaze and what can be done to improve interaction).

2 RELATED WORK

Here we introduce the most commonly used approaches and recent advances regarding VR object manipulation (also see more thorough recent reviews [31, 36]). We further discuss gaze-supported techniques used in VR and other domains.

2.1 Object Manipulation in VR

Mid-air interaction based on *Virtual Hand* is one of the primary input paradigm for modern VR systems [36]. With spatially tracked hand positions, typically with 6 degrees-of-freedom (DoF), users are able to directly translate and rotate objects in virtual environments in a similar way as they manipulate them in the physical world [44]. Although it has been criticized to be inefficient and imprecise [7, 36], due to its simplicity and intuitiveness of the control, Virtual Hand has been widely applied in various VR applications [14, 20, 27].

Further approaches have been used to enhance Virtual Hand. For example,*Go-Go* [43] and its recent extension [69], which scale up the speed of the virtual hand, enable users to reach distant targets, even at a potentially infinite distance [5]. *Raycasting* also provides an easy solution for acquiring distant objects, but users may not be able to rotate the object precisely with one single hand as they are attached to the end of the ray [5]. Other methods [41, 57, 73] scale-down the whole virtual world to enable the interaction with out-of-reach objects.

To offer fine-grained manipulation control, several interaction techniques decrease the control-display ratio of the hand movement based on hand velocity [17, 70]. Degree-of-freedom (DoF) separation [37, 65] is another promising way to increase the accuracy of mid-air object manipulation—that is, rather than manipulating all the six DoF simultaneously, only one or two of them are controlled each time. For instance, in a recent work, researchers tried to reduce the DoF during object manipulation by constraining it to the shape of a point, ray, or plane, thereby increasing precision [22].

Nevertheless, many mid-air interaction techniques fall short in supporting prolonged manipulation due to cumulative arm muscle fatigue (the so-called "gorilla arm" effect) [29]. This is especially detrimental to interaction scenarios such as 3D modeling in VR, which require fine-grained, focused, and prolonged usage of mid-air interfaces. To address these challenges, providing indirect mappings [35] or integrating other less effort-demanding input modalities such as gaze into object manipulation techniques in VR can be potentially helpful.

2.2 Gaze-Supported Manipulation

Gaze-supported object manipulation has been widely explored in contexts outside VR. In general, while gaze offers fast and natural pointing, it suffers from the lack of precision and the difficulty of confirming a selection. To overcome these challenges, many techniques combine gaze with an additional modality, such as the principle of "gaze select, hands manipulate" [9, 39, 56, 66]. For example, Pfeuffer et al. proposed *Gaze-touch* [39], which enabled users to control gaze-selected targets indirectly using multi-touch gestures on interactive surfaces. Another example is the method proposed by Turner et al. [62], which casts the object being looked at by the user to the touch/cursor position to allow further manipulation. In contrast, other approaches [48, 55, 60–63, 67] for content-transfer between different displays, have embedded gaze movement into the translation process. These prototypes typically require the use of a hand trigger to "attach" the object to the gaze direction and then release the hand trigger to "drop" it. In a follow up research, Turner et al. [59] pushed this concept further by developing techniques that maintain concurrent rotation and scaling operations when performing translation tasks using gaze and touch.

Limited work has investigated gaze input for object manipulation in VR or 3D virtual space. Simeone et al. [52] combined bi-manual touch gestures with gaze input to allow the scaling of objects on the XYZ-axis inside a touchscreen. Liu et al. have presented *OrthoGaze* [34], in which gaze is issued to move an object along three orthogonal planes in VR. Other researchers have used eye gaze to select objects, and leveraged indirect freehand gestures to manipulate them [40, 42, 45, 54]. All of them still followed the idea of "gaze select, hands manipulate". In contrast to these approaches, the gaze input in our work was not only used for the selection of objects but also was involved in the whole target manipulation process, which requires continuous actions rather than the discrete selection operation [59]. Our aim is to understand how different methods of hand-eye integration, coordination, and transition can result in improved user performance and their suitability to be applied to a variety of scenarios.

2.3 Transition Between Gaze and Hand Input

Different collaboration strategies have been explored to combine gaze and other input modalities, such as hands or head [49, 50], and the transition between different modalities can be classified according to whether they are explicit and implicit. Explicit transitions rely on specifically issued commands to switch between gaze and other forms of input. The "switch" orders include actuating the input device or pressing a trigger. In contrast, implicit transitions do not rely on distinct commands to switch between multiple input mechanisms; all modalities always have an effect on the cursor/object that users interact with, and users do not need to concern about the transition during the interaction.

An example of the explicit transition is *Pinpointing* [30], which starts with a fast but imprecise modality like gaze, and then refining it with a slower but more precise input modality such as hand gestures with an explicit button click or a finger gesture for mode transition. An example of the implicit transition is "liberal" *MAGIC* pointing [76], where users can always move the cursor with manual or gaze input once they have decided to do so, without activating any trigger. While explicit transitions offer more robust control in many cases [30], implicit transitions fade the boundary between the input mechanisms and smooth the "flow" of interaction.

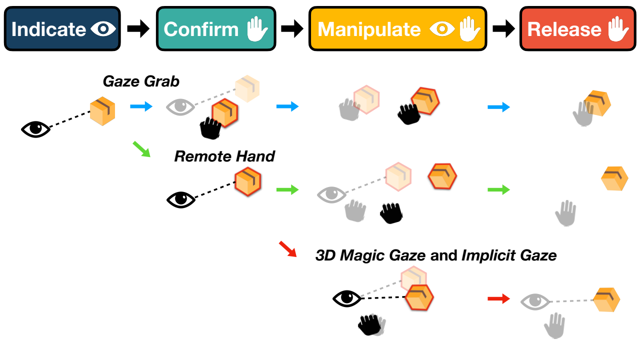

Figure 1: An illustration of the target manipulation process, where gaze is used for indicate, hands are used for trigger confirm and release, and both inputs are applied (as hand only, or gaze and hand collaboratively) for object manipulation.

It is important to note that we distinguish "implicit" from other names presented in the literature like "seamless" [51, 55], the smoothness of the transition, and "concurrent" [59], the ability to manipulate multiple degrees-of-freedom simultaneously. Seamlessness and concurrency do not ensure an implicit transition; we only consider if an explicit triggering mechanism is used. As discussed, previous works on gaze-supported VR object manipulation mainly used gaze as a selection technique, rather than incorporated it into the manipulate phase which includes translation, rotation, and scaling. Therefore, how to transition between gaze and hand input for manipulation tasks is still underexplored.

3 DESIGN SPACE

3D object manipulation techniques can be broken down into an initial selection of object and the later manipulation steps including translation, rotation, and scaling [6, 40]. We first introduce this process and propose corresponding input modalities for each sub-phase. We then formulate a design space that considers how gaze and hand input are integrated into different phases of the manipulation task, coordinate with each other when starting the manipulation, and transition from one to the other during manipulation. Based on the design space, we further point out several gaps in the existing literature, and use this knowledge to design our proposed techniques.

3.1 Target Manipulation Process

We first introduce a target manipulation process that is based on previous works [6, 40]. The whole task can be decomposed into four phases: indicate, confirm, manipulate, and release (see Figure 1). We identified suitable input modalities for each phase, which is then useful to structure and narrow down our exploration space.

3.1.1 Indicate. Indicating is the action of determining the target of interest with an input device. The literature suggests that gaze-based pointing requires less effort and can be faster than manual input [39, 40, 55]. Further, gaze tracking has become more accurate with recent advances in the field [15]. Therefore, we consider gaze as our input mechanism in the indicate phase.

3.1.2 Confirm. Confirming the selection allows users to "pick up" and start manipulating the indicated target. Because gaze-based confirming techniques, such as dwell, can be inefficient and may induce unwanted selection [26], we decided to use a hand-based method, specifically, pressing the trigger on the hand-held controller for a robust control of the confirm phase.

3.1.3 Manipulate. Manipulation of objects, including translation, rotation, and scaling, can be achieved by hand input alone, or by gaze and hand input together. Gaze can be treated as a 2 degrees-of-freedom (DoF) modality as an estimated gaze point normally moves on a 2D spherical plane, while accurately predicting the depth of the gaze point can be challenging [24]. In contrast, hand-based mechanisms typically feature 6 DoF motion input (both translation and rotation along the 3 axes). Based on its properties, gaze offers more opportunities for rapid translating objects in the lateral direction [59]. As for hands, they are likely to be better in positioning objects in the depth dimension (the third DoF), and rotating or scaling them (as they either require the rotation of the input device or need multiple control points). To distinguish this phase from the whole target manipulation process, we call it *the manipulate phase* in this paper.

3.1.4 Release. Releasing the trigger signals the completion of one operation. Similar to the confirm phase, we use the trigger on the controller for the robust control of the release phase.

3.2 Design Dimensions

We considered the following three-dimensional design space by emphasizing the integration, coordination, and transition of gaze and hand input for the manipulate phase. While we acknowledge that exploring other design dimensions such as target properties and input techniques can be useful, this research focuses on exploring how to incorporate gaze-input into the traditional hand-based workflow.

- **D1. Integration**: which input mechanism(s) of gaze and hand has (have) been integrated into the manipulate phase.
- **D2. Coordination**: when starting the manipulate phase, if the indicated target will snap to the hand position or remain in its original place. This further corresponds to whether the object is directly mapped onto the hand position (direct mapping) or manipulated by hands remotely (remote mapping).
- **D3. Transition**: if both input mechanisms are involved in the manipulate phase, whether the transition between gaze and hand input is explicit or implicit (with or without specifically issued triggering commands like button pressing).

3.2.1 Synthesis of Prior Work. We further summarized how existing gaze-supported manipulation techniques fit into each dimension of the design space (see Table 1). We have focused on the ones that involve hand input in the manipulate phase, rather than relying on the gaze input alone. That is, the approaches that use gaze input only as a supporting mechanism for manipulation.

3.2.2 Research Gaps and Design Opportunities. The design space and the synthesis of prior work reveal some research gaps that are essential for framing the design of gaze-supported object manipulation techniques but are still underexplored in the literature and thus create new design opportunities.

	Techniques	Integration		Coordination		Transition		
		Gaze	Hand	Direct	Remote	Implicit	Explicit	None
	Eye drop [61, 62]		✓	✓				✓
	TouchGP [55]	✓	✓		✓		✓	
	Gaze-Touch [39]		✓		✓			✓
2D	TouchT [59]		✓		✓			✓
	GazeT [59]	✓	✓		✓		✓	
	MagicT [59]	✓	✓		✓		✓	
	Gaze [66]		✓		✓			✓
	Gaze + Non-touch [42]		✓		✓			✓
	Three-point [52]	✓	✓	✓				✓
	Gaze + pinch [40]		✓		✓			✓
	GG interaction [45]		✓		✓			✓
3D	Gaze + Gesture [9, 10, 54]		✓		✓			✓
	Gaze Grab		✓	✓				✓
	Remote Hand		✓		✓			✓
	3D Magic Gaze	✓	✓		✓		✓	
	Implicit Gaze	✓	✓		✓	✓		

Table 1: Summary of how existing gaze-supported manipulation solutions and ours (the bottom four) fit into the design space. Our techniques enabled us to explore explicit and implicit transitions, which have not been well-covered by previous research in 3D, and how different design dimensions may influence user performance and experiences in VR manipulation.

G1. *Transition mechanisms between gaze and hand input have not been investigated in the manipulate phase in VR*; most of the previous work focused on the rationale of "gaze select, touch manipulate". However, gaze can not only support discrete pointing tasks but can also be beneficial for target manipulation, which requires continuous actions [55, 59]. Further exploration is needed to understand how gaze input supports manipulation in VR environments, which offers 3D spatial input and stereo vision [31].

G2. *Implicit transition is still under-explored for target manipulation tasks in general.* According to Table 1, there is lack of implicit transition techniques in the manipulate phase. All transitions are based on either releasing a pressed trigger [55] or exceeding a hand movement threshold [59] to switch from gaze input to hand input.

G3. *Techniques that leverage different elements of the design space have not been compared in terms of their efficiency and usability.* For example, it is unclear how gaze-supported methods that allow remote (indirect) manipulation compare to direct manipulation-based solutions in terms of performance and user experiences, although they have been applied in different applications [68]. Furthermore, it is still unclear if gaze-supported techniques can provide more benefits than hand-only techniques in the manipulate phase in VR.

4 TECHNIQUE DESIGN

Based on the identified research gaps and design opportunities, we developed the following four techniques to (1) explore transition mechanisms (**G1 - 3DMagicGaze**), especially implicit transition (**G2 - ImplicitGaze**), for target manipulation in VR and (2) evaluate and compare approaches that leverage different elements of the design space in terms of user performance, experiences, and their suitability to be applied to a variety of scenarios (**G3**). Table 1 shows how each technique fits within the design space.

4.1 Gaze Grab

With *GazeGrab*, the gaze-indicated target snaps to the hand position once the selection is confirmed. Next, the hand takes full control of the selected target during the manipulation phase until the trigger is released. This technique allows the direct manipulation of objects and represents a VR-enhanced version of previous research on content transfer [61]. Similar techniques have also been demoed in VR applications [68], though it has not been empirically evaluated or compared with other techniques. In our design, the gaze-grabbed object is located slightly above the virtual hand position, to avoid visual occlusion.

4.2 Remote Hand

To manipulate an object through *RemoteHand*, a user first points at it with eye gaze and then confirms the selection with a hand trigger. The target then follows the rotation and translation of the hand, without snapping to the hand location. This technique enables the indirect manipulation of targets with hand movement. It can be seen as a 3D extension of existing approaches in 2D [39, 40, 66], which follow the underlying rationale of "gaze selects, hand manipulates".

4.3 3D Magic Gaze

3DMagicGaze establishes a circular safe region (10° radius, invisible to users) around the target once the initial eye-based selection is confirmed. If the gaze point is within the safe region, only the hand can control the transformation of the object. Otherwise, when the gaze point is outside of the safe region and if the hand movement distance exceeds a threshold (0.08m), the object snaps to the gaze point direction (without changing its depth to the user). A new safe region appears around the target after the snapping takes place. The design of this technique follows *MagicT* [59] in 2D, which requires an explicit command (hand movement) to switch from gaze input to manual input.

4.4 Implicit Gaze

ImplicitGaze also forms a circular safe region around the target once the eye-based selection is confirmed. If the gaze point is inside the safe region, hand input will control the object's transformation. Otherwise, if the gaze point is outside the safe region, the object snaps to the gaze point direction (without changing its depth to the user). A new safe region appears around the target after snapping. Unlike *3DMagicGaze*, this technique does not rely on any trigger mechanism to switch between gaze and hand input, thus features "implicit" transition between input modalities. To prevent the gaze cursor from being "over-active" [76], we introduced a dynamically-resized safe region, which resizes automatically based on the user's gaze behavior. It then increases its size from the original radius (6°) with a constant speed (10°/s) if the gaze point stays within the region until the maximum size (20°). This is to simulate users' search behavior, in which the longer the gaze point is fixed within specific regions, the more likely the user is approaching the target location [19, 39, 53]. Therefore, increasing the safe region's size can avoid unwanted snapping and allow robust, fine-grained hand translation.

The parameter values were obtained from our pilot tests. We kept the following design aspects consistent among the techniques: (1) the gaze pointer is invisible to users so as not to distract them; (2)

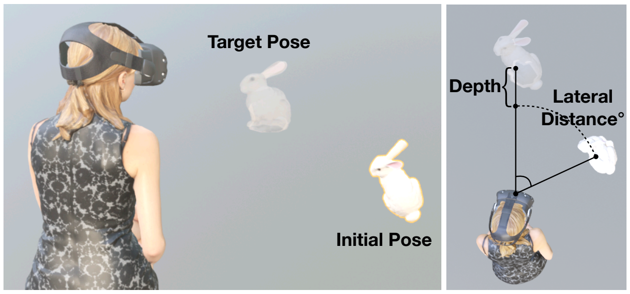

Figure 2: (left) Task illustration: participants were required to transform an object from its initial configuration to a target pose; (right) An illustration of Lateral Distance and Depth which were independent variables of the first study.

all techniques had the same control-display mapping (1:1) for hand manipulation; and (3) all techniques used the trigger button of the hand-held controller to confirm and release the selection.

Next, we present two user studies where we evaluated and compared the four gaze-supported manipulation techniques that employed different integration, coordination, and transition strategies. In the first study, we focused on the *primary working space*, where all objects located in front of the user and were within arm-reach distance, and assessed the techniques in terms of user performance and experience. In the second study, we further evaluated our techniques in a larger virtual environment with distant objects and embedded the designed techniques into realistic workflows.

5 STUDY 1: CONTROLLED EVALUATION

In this study, our goal was to evaluate and compare the four gaze-supported manipulation techniques (*GazeGrab*, *RemoteHand*, *3DMagicGaze*, and *ImplicitGaze*) that leveraged different design features from the presented design space in a controlled working space. By doing so, we aimed to better understand whether gaze input should be incorporated into hand manipulation process, and how gaze input could be combined with hand input for convenient and efficient 3D object manipulation in VR. The study mainly focused on the *primary working space*, where all targets of interest are located in front of the user (less than 90° horizontal offset when the user is looking forward) and are within arm-reach distance. Most of the work in VR is likely to happen within this area, so there is no need for users to frequently turn back or move around the virtual environment [1, 13, 72].

5.1 Participants and Apparatus

We recruited 12 university students (3 women, 9 men) between the age of 18 to 29 years (mean = 22.5) for this first study. All participants reported to be right-handed.

We developed the system using the Pico Neo 2 Eye, a standalone VR headset with 6 DOF tracking and Tobii eye-tracking features. The headset has 1920 × 2160 pixels screen resolution per eye and 101° field-of-view (FoV). The embedded eye tracker has 90Hz data output frequency, 0.5° estimated accuracy, and 25° left/right/down and 20° up trackable FoV. The software was implemented in C# in Unity3D.

5.2 Task

The task required participants to transform a 3D model from its initial configuration to a new target pose (see Figure 2 left). The target location was randomly selected within 30° of angle distance when the participant was looking straightforward along the z-axis (the depth axis) of his/her local space. The initial position was then calculated according to the target position based on our independent variables—lateral distance (the angular distance between the start and target location) and depth (the differences in the depth dimension). The target position was to be expected by participants. In other words, they knew where the object should be translated to when starting the manipulation task, even when the initial target was located outside the user's field-of-view (but within the primary working space). This allowed us to minimize the search time, which may confound with the manipulation time. Another factor, which is the object orientation, was adjusted according to the experiment requirement.

5.3 Evaluation Metrics

5.3.1 Performance Measures. To evaluate technique performance, we controlled transformation errors to be under a threshold (smaller than 0.015m and 3.5°) while comparing task completion time.

- *Manipulation Time*: the time elapsed between when object selection is confirmed and when both of the following conditions are satisfied: (1) the target is correctly placed with errors under the pre-determined threshold; and (2) the trigger is released.
- *Coarse Translation Time*: the time elapsed between the selection confirmation and the first time when the distance between the acquired object and target position is smaller than 0.05m. The rationale for including this variable was that, during our pilot studies, we found users took a long time to re-adjust the object orientation and fine-tune its position after reaching an approximate target location.
- *Re-position Time*: the elapsed time for fine-grained manipulation (= Manipulation Time - Coarse Translation Time).

5.3.2 Hand Manipulation Measures. We were also interested in investigating how techniques may influence hand movement and rotation for manipulation tasks, which may correlate to the arm fatigue measures, based on the simple rationale that more hand motion is likely to induce more arm fatigue [23].

- *Hand Movement Distance*: the accumulated distance (by accumulating the displacement of hand per frame) that the hand has travelled during the manipulation process.
- *Hand Rotation Angles*: the accumulated angle that the hand has rotated during the manipulation process.

5.3.3 Subjective Measures. We also compared the techniques based on subjective measures, including arm fatigue, ease of use, required workload, and individual rankings.

- *Borg CR10* [4, 29]: a categorical rating (0-10 points) which can be used to assess perceived arm exertion/fatigue. It has been shown to correlate well with objective measures from, for example, EMG data [58]. We adopted the same format and verbal description as previous works [29] in this experiment.
- *Single Easement Questionnaire* [46]: to measure the ease-of-use of the techniques with a 7-point scale.

- *Raw NASA-TLX* [21]: to measure the task load induced by the techniques with 7-point scales.
- *Subjective Ranking*: a rank of all the techniques according to participants' overall preference.

5.4 Design and Procedure

The study employed a $4 \times 3 \times 2$ within-subjects design with three independent variables: Technique (*RemoteHand*, *GazeGrab*, *ImplicitGaze*, and *3DMagicGaze*), Lateral Distance (35° and 55°), and Depth (0.05m, 0.10m, and 0.15m). Lateral distance represents the angular distance between the start and target location, whereas the depth factor looks at the differences in the depth dimension along the user's line of sight (see Figure 2 right). The current level and task setting made all objects to be located within the primary working space (from 0° to 85° horizontal offset and within arm-reach distance). The presentation order of Technique was counterbalanced using the Latin Square approach, whereas Lateral Distance and Depth were presented in random order. Additionally, the rotation factor (20°, 50°, 80°, 110°, and 140°), which is the required rotation (in angles) from the initial to the target transform, was pre-determined for each repetition and the same set of values was used across all conditions (though appeared with a randomized order). Exploring the effect of rotation was not our primary focus, as all techniques used a similar method to achieve that purpose. In the experiment, each condition was repeated 5 times which resulted in 1440 (= 12 participants × 4 techniques × 3 lateral distances × 2 depths × 5 repetitions) trials of data.

The whole experiment lasted approximately 50 minutes in total. Participants first completed a questionnaire to collect their demographic information. They were then introduced to the experiment task and the VR device, and instructed to complete the trials as fast and as accurately as possible. Next, we asked participants to put on the headset and start the experience in VR. The VR experience consisted of four sessions corresponding to four manipulation techniques. Each session began with ten warm-up trials for participants to get familiar with the input method, followed by the formal test trials. After each session, we collected user feedback with the Borg CR10, Single Easement, NASA-TLX, and Subjective Ranking questionnaires. Participants were required to have a rest between each session.

5.5 Results

To analyze the collected data, we first discarded the outliers that deviated more than three standard deviations from the mean value ($mean \pm 3std.$) in each condition (20 trials, 1.3%). Furthermore, a Shapiro-Wilk test indicated that the data is non-normally distributed. Therefore, all data underwent pre-processing through Aligned Rank Transform (ART) [71]. Next, we performed repeated-measures ANOVAs (RM-ANOVA) and Bonferroni-adjusted pairwise comparisons for each measurement. We also computed effect size (the non-parametric estimator for CL, symboled A_w [32, 64]) to accompany the pairwise tests based on unranked (non-normal) data. The results from performance measures, hand manipulation measures, and Borg CR10 are summarized in Figure 3.

5.5.1 Performance Measures. A RM-ANOVA indicated that Technique ($F_{3,253} = 4.141, p = .007$) and Lateral Distance ($F_{1,253} = 5.414, p = .021$) had significant main effects on Manipulation Time, but not Depth ($F_{2,253} = 0.186, p = .831$). No interaction between these variables was found. A post-hoc test indicated that *GazeGrab* (13.7*s*) was significantly slower ($p = .004$, $A_w = 0.64$) than *RemoteHand* (12.3*s*).

Another RM-ANOVA showed that both Technique ($F_{3,253} = 3.084, p = .030$) and Lateral Distance ($F_{1,253} = 25.024, p < .001$) had significant main effects on Coarse Manipulation Time, but not Depth ($F_{2,253} = 0.610, p = .544$). An interaction effect between Technique and Depth was also identified ($F_{6,253} = 3.396, p = .003$). When Depth increased, while *RemoteHand*, *ImplicitGaze*, and *3DMagicGaze* led to larger Coarse Manipulation Time, *GazeGrab* required less time. A post-hoc test indicated that *ImplicitGaze* (4.1*s*) was significantly faster ($p = .036$, $A_w = 0.61$) than *GazeGrab* (4.7*s*).

Finally, Technique ($F_{3,253} = 3.861, p = .010$) had a significant main effect on Re-position Time, but not Lateral Distance ($F_{1,253} = 1.377, p = .242$) or Depth ($F_{2,253} = 0.452, p = .637$). No interaction effects were found. According to a post-hoc test, *GazeGrab* (9.1*s*) was significantly slower ($p = .007$, $A_w = 0.64$) than *RemoteHand* (7.6*s*).

5.5.2 Hand Manipulation Measures. A RM-ANOVA showed that both Technique ($F_{3,253} = 13.559, p < .001$) and Lateral Distance ($F_{1,253} = 55.681, p < .001$) had significant main effects on Hand Movement Distance, but not Depth ($F_{2,253} = 0.126, p = .881$). No interaction effects were found. A post-hoc test indicated that *ImplicitGaze* required much smaller hand movement than *3DMagicGaze* ($p = .047$, $A_w = 0.38$), *GazeGrab* ($p < .001$, $A_w = 0.30$), and *RemoteHand* ($p < .001$, $A_w = 0.31$). Furthermore, *3DMagicGaze* required significantly less hand movement than *GazeGrab* ($p = .008$, $A_w = 0.38$).

Furthermore, Technique ($F_{3,253} = 15.663, p < .001$) and Lateral Distance ($F_{1,253} = 26.569, p < .001$) had significant main effects on Hand Rotation Angles, but not Depth ($F_{2,253} = 0.924, p = .398$). Additionally, no interaction effects between the variables were found. *ImplicitGaze* led to significantly less hand rotation than *3DMagicGaze* ($p = .003$, $A_w = 0.37$) and *RemoteHand* ($p = .008$, $A_w = 0.39$). Additionally, *GazeGrab* also resulted in significantly less hand rotation than *3DMagicGaze* ($p < .001$, $A_w = 0.29$) and *RemoteHand* ($p < .001$, $A_w = 0.30$).

5.5.3 Subjective Measures. A RM-ANOVA test indicated that *GazeGrab* induced more arm fatigue and higher physical workload than all other techniques (for all pairwise comparison, $p < .001$). It also led to higher (mental and physical) effort and created more frustration than *ImplicitGaze* and *RemoteHand* (all $p < .024$), and higher mental demand than *3DMagicGaze* ($p = .034$). Subjective ranking data indicated that participants significantly preferred *ImplicitGaze*, *RemoteHand*, and *3DMagicGaze* over *GazeGrab* (all $p < .001$). No other statistically significant effect was identified.

5.6 Discussion

5.6.1 Hand-Only vs. Eye-Hand Manipulation. When comparing the hand-only (during the manipulate phase) technique (*RemoteHand*) and the eye-hand techniques (*3DMagicGaze* and *ImplicitGaze*), we did not observe significant performance differences. This extends findings from previous research on 2D screens [59], where gaze was not be able to enhance the performance of manipulation tasks in the primary working space. On the other hand, adding transitions between gaze and hand also did not deteriorate performance compared

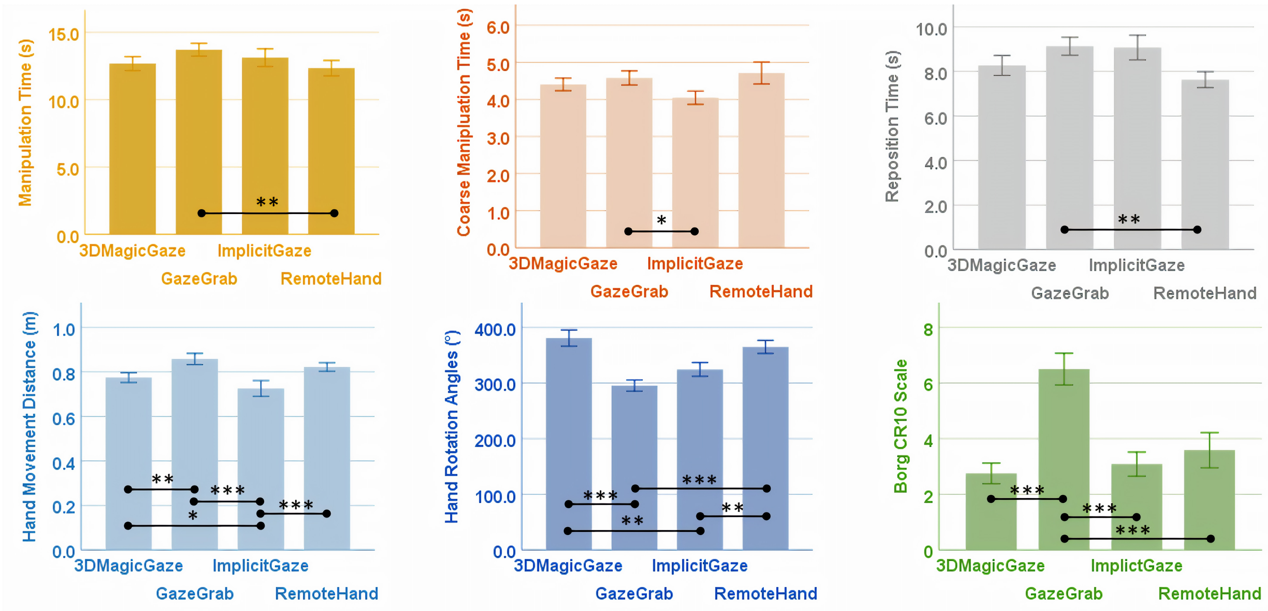

Figure 3: Plots of techniques' performance under different measurements. Error bars indicate the standard error. Statistical significant effects are marked (* = p < .05, ** = p < .01, and *** = p < .001).

to hand-only techniques; participants quickly learned/adapted to these new input methods. As expected, *RemoteHand* required more hand movement and rotations to achieve the same manipulation task. However, the results from the Borg CR10 and NASA-TLX questionnaires did not show significant benefits of eye-hand transitions over hand-only techniques regarding arm fatigue and perceived workload.

5.6.2 Direct vs. Remote Hand Mappings. When comparing *GazeGrab* to other techniques that allowed remote hand mappings (specifically *RemoteHand*), we observed substantial differences in performance measures and subjective feedback. *GazeGrab* required a much longer time frame to re-position an object than *RemoteHand* and caused significantly higher perceived arm fatigue. This was mostly because, as indicated in previous research, "direct manipulation" techniques are imprecise in nature [36]. Participants found it difficult to place the object in the correct position by holding it with their arms. Further, participants suggested that *GazeGrab* was cumbersome as it required them to "suspend" their arms in the air to perform the manipulation (in contrast with the techniques based on indirect mapping which allowed them to manipulate the target with their arms down). Interestingly, *GazeGrab* induced less hand rotation than *RemoteHand*. In fact, according to the mean value shown in Figure 3, *GazeGrab* had the smallest hand rotation angles. This is likely due to the presence of direct mappings, which leads to users finding it easier to determine how to optimally rotate an object to the target configuration.

5.6.3 Implicit vs. Explicit Eye-Hand Transitions. When comparing *ImplicitGaze* and *3DMagicGaze*, we found that they led to similar empirical performance, while *ImplicitGaze* required less hand movement and rotation to complete the manipulation task. This difference was most likely due to the transition mechanism we chose for *3DMagicGaze*, which entailed the use of hand movement to snap the target to the hand position. Our choice was based on Turner et al. [59] work, where they thought such input structure would demonstrate "some form of integrity" (as we usually use a final hand manipulation to fine-grain the translation made by gaze input). However, according to our study results, we found this explicit hand movement can have side effects. It required participants' hands to move for longer periods of time and rotate more to achieve the same task compared to the implicit approach. Even if we change to other mode switching mechanisms, like trigger tapping [55], it is likely that such extra efforts would still be needed for methods based on explicit transitions. In contrast, implicit transition techniques can be an ideal solution as they require minimum effort for mode switching. Our results also showed no issues regarding unwanted snapping (in other words, not inducing the Midas touch problem [28]) by using a dynamically-resized safe region.

5.6.4 Effect of Lateral Distance and Depth. As expected, our results showed that lateral distance influenced the technique performance in coarse manipulation time, but not re-position time (mostly orientation adjustment). Depth did not have a definite impact on selection performance, likely due to the differences between the levels not being substantial (as all of them were within arm-reach distance).

5.6.5 Summary of Study 1's Key Findings. Based on the discussion, we summarize the following key findings from the first study.

- Our results show no evidence that manipulating objects (mainly translation) based on both eye and hand input (*3DMagicGaze* and *ImplicitGaze*) can offer significant performance benefits in VR manipulation tasks over the hand-only approach (*RemoteHand*) in the primary working space.

- Direct hand mapping (*GazeGrab*) is less precise and can lead to more arm fatigue than remote hand mappings (like *RemoteHand*). However, it might help users to determine how to optimally rotate an object to the target configuration.
- Implicit transition (*ImplicitGaze*) and explicit transition (*3DMagicGaze*) led to similar task performance, while implicit transition required less effort (e.g., hand movement) than explicit transition. In particular, a dynamically-resized safe region was shown to be useful as there was little evidence of the Midas touch issue [28].

After assessing technical performance and initial user feedback in Study 1, we further extended the evaluation to a larger space which requires the use of locomotions in Study 2.

6 STUDY 2: APPLICATION

In this second study, we aimed to assess how gaze-supported manipulation techniques perform under a larger environment and when applied to realistic workflows. We also wanted to compare our techniques with Virtual Hand (hand input only for selection and manipulation), which is currently the most common method for manipulating objects. We also measured user experience and collected user feedback, which can help adapt the gaze-supported techniques to real use cases.

6.1 Participants and Apparatus

We recruited eight university students (3 women, 5 men) with previous experience in 3D modeling (1-8 years, mean = 2.75, using software like SolidWorks, 3DS MAX, CAD, Rhino, and Unity). We hope that more fruitful discussions could be triggered with experienced/expert users in the relevant domain. Their ages were between 21-29 years (mean = 24.4). All of them were right-handed. We used the same device as in the previous study.

6.2 Interaction Scenario

Participants were instructed to reconstruct an empty room following a miniature, as shown in Figure 4, using the manipulation techniques. However, they were not required to follow how the miniature looked like precisely; rather, it was used as a guide for them to make their own creations. Participants could move around the room using the teleportation mechanism, choose desired objects from a prefab list (see Figure 4), and manipulate (translate, rotate, and scale) the selected item. This differed from the first study, which controlled the participants in a static position (within primary working space) and had specific time-controlled task requirements. In this interaction scenario, we emphasized the "design-by-yourself" concept, where the techniques were integrated into users' own workflow and creative experiences [75]. Similar applications include Mozilla Hubs [25] or Minecraft VR [38], where users/players can decorate/build virtual space with different objects/building blocks.

6.3 Procedure

The whole experiment lasted approximately 60 minutes in total. Participants first completed a demographic questionnaire. Then, participants were briefed about the task and program functionalities, and were asked to put on the headset on and started interacting with the virtual space. The whole interaction experience was divided into five sessions (four gaze-supported techniques and virtual hand were presented in a randomized order). During each session, they learned about a manipulation technique and performed the task as described in the previous section. At the end of each session, they completed a short version of the User Experience Questionnaire (UEQ-S) [47] and answered a set of structured questions to provide their overall feedback towards the technique. The structured questions asked about the strengths and weaknesses of each method. After finishing the five sessions, they were also invited to provide their opinions regarding the different design features employed in the techniques (hand-only *vs.* hand-eye, direct *vs.* remote mappings, and implicit *vs.* explicit transitions). Responses were recorded for further analysis.

Figure 4: Participants were instructed to construct an empty room following a miniature (left) with the gaze-supported manipulation techniques. They were able to teleport around the room, select objects from a prefab list (right), and manipulate (translate, rotate, and scale) the selected item.

Technique	Pragmatic	Hedonic	Overall
GazeGrab	0.66	0.84	0.75
RemoteHand	0.21	-0.25	-0.02
3DMagicGaze	0.31	0.94	0.63
ImplicitGaze	0.68	1.10	0.89
Virtual Hand	-0.13	-0.63	-0.38

Table 2: The results from the short version of User Experience Questionnaires (UEQ-S) which outline the pragmatic quality, hedonic quality, and overall quality of each Technique (higher scores are better).

6.4 Results

The results from UEQ-S are summarized in Table 2, which indicates that the gaze-supported techniques performed better comparing to Virtual Hand in terms of pragmatic, hedonic, and overall quality. Next, we provide a summary of participant interview responses grouped by technique.

6.4.1 Gaze Grab. As a way of hand-eye coordination, *GazeGrab* has a unique feature of snapping the object to the hand position when starting the manipulate phase. A number of participants (N=5) commented that it was "*efficient*" and "*convenient*" way of achieving this; "*I normally moved to the destination first, and then brought the object to me with the technique. It was very quick.*" (P2). However, a couple of participants mentioned that "*the efficiency of GazeGrab was highly dependent on the accuracy of teleportation method, which was sometimes not very accurate.*" (P3). The inaccurate teleportation might require users to re-adjust their standing position when using

GazeGrab. Two participants also said that the technique "*required some learning*". Notably, P5 noticed that "*when object flew to me, especially big objects like a sofa, I was afraid that it might hit me.*", and P5 also found it challenging to fine-grain the position of an object as "*the object would fly to my hand again when pressing the trigger, and my previous effort was wasted*".

6.4.2 Remote Hand. Although this technique has the ability to manipulate objects remotely, almost all participants (N=7) noted that *RemoteHand* was inconvenient when moving objects that were at a far distance; "*It seemed that the object only moved a little bit when I moved my arms.*" (P2). "*This was fatiguing.*" (P1). Moreover, P5 mentioned that "*when I tried to move the object for a large distance, my arm's movement might also cause the rotation of the object. So I had to rotate it back.*" Despite these limitations, most participants (N=7) felt *RemoteHand* was accurate for manipulation. In addition, P5 commented on the agency provided by the technique "*manipulating objects remotely made me feel that I was taking control of the whole space*".

6.4.3 3D Magic Gaze. Half of the participants (N=4) explicitly mentioned that, with the help of their eyes, *3DMagicGaze* was quick for long-distance object translation. However, a few participants (N=5) mentioned some flaws in the hand confirmation mechanism: "*I needed time to get used to this (hand movement for confirmation).*" (P6) "*For small or medium movement, it was sometimes hard for me to decide whether using hand or gaze.*" (P5). Additionally, some participants (N=5) thought the switching between eye and hand input was confusing at times: "*I often forgot using hand to bring (snap) the object.*" (P7) "*I found sometimes waving my arms did not make the quick transformation (snap). For example, I wanted to put a bed adjacent to the wall, but it was hard to achieve—the movement was either too small or too large*" (P4). The later was because the gaze cursor was still inside the safe region, so the hand snapping did not happen. In contrast, P8 said that "*I did not feel any big difference comparing to ImplicitGaze.*" and indicated that hand movement was natural for confirmation. P6 further said *3DMagicGaze* felt more "*stable*" than *ImplicitGaze*, since the selected object would not frequently snap to the gaze direction.

6.4.4 Implicit Gaze. Participants (N=7) felt that *ImplicitGaze* was "*novel*" and "*efficient*"; "*I can just stand still and manipulate the objects quickly.*" (P4). However, several participants (N=4) also commented about the difficulty of using eyes to achieve precise manipulation. "*When I was searching for the places, the object, especially the big ones, would block my view. Also, there were some unwanted movements caused by eyes.*" (P3). On the positive side, P7 commented that "*I thought eye movement might cause some random movements before using it, but it actually didn't when trying.*" Noticeably, some participants (N=3) thought it was not as easy to move the object in the depth dimension with *ImplicitGaze*, as the movement in that dimension is particularly slower than lateral directions.

6.4.5 Virtual Hand. Almost all participants (N=7) thought *VirtualHand* was natural and realistic; "*I always know how to do it (the manipulation), as that's what we do in everyday life.*" (P3). The technique also felt more "*controllable*" due to these characteristics. However, all participants (N=8) acknowledged that *VirtualHand* was "*fatiguing*" and "*not efficient enough for long-distance translation*".

6.5 Discussion

In this section, we discuss and summarize the results and provide solutions for the identified limitations and design implications that can help future implementation of gaze-supported manipulation techniques in VR.

6.5.1 Hand-Only vs. Eye-Hand Manipulation. While the benefit of rapid eye movement for object translation is not salient in the primary working space (as shown in the first study), for manipulating faraway objects in a larger environment, participants clearly preferred the efficiency and convenience of gaze-hand combination for coarse translation. Indeed, theoretically, an exact control-display mapping (1:1) of hand movement has little effect (visually) on objects located in a far distance from a user's perspective. In such a situation, it is thus more ideal for translating the target according to visual angles (as what gaze input does), rather than exact distance mapping (as what hand input does in this research). Another solution, which can enhance hand-only approaches (e.g., *RemoteHand*) in the manipulate phase is to provide hand amplification (e.g., [69]), where the hand movement is amplified using specific functions, so the object appears to move a larger distance.

However, eye-hand manipulation became less useful for close and large objects, as it might occlude the user's line-of-sight (since the target follows gaze), which made location searching difficult. A quick fix could entail making the target under manipulation semi-transparent [11], so that the user's view is not fully-blocked. Some participants also found hand-only manipulation to be more manageable, as they reported being more used to this type of input.

6.5.2 Direct vs. Remote Hand Mappings. With the feature of bringing faraway objects to users' hands (turns a remote object to direct hand mapping), *GazeGrab* shifted how participants interacted with objects when compared to the other three gaze-supported techniques. With remote-mapping based methods like *ImplicitGaze*, participants tended to remain in the same standing position and transferred the items remotely. In contrast, with *GazeGrab*, they were likely to first move to a new target position and then bring the object to their location. As reported by the participants, this transformation was efficient in transporting distant targets but can be cumbersome for close ones. Repetitive snapping close objects to hands can make the adjustment difficult, and it is likely better to disable this function when the target is within arm-reach distance. Furthermore, users needed to re-adjust their standing position if there was any inaccuracy caused by the locomotion technique. If the VR locomotion/teleportation [3] is sufficiently smooth, efficient, and accurate, the snap-to-hand function could be useful by translating distant objects along the depth dimension.

Another issue brought by direct hand mappings is that if the object under manipulation is quite large, participants found it difficult to transform the object into a satisfiable configuration as a significant part of their view is occupied by the item. Some participants also reported that it made them feel unsafe as they thought the object might collide with their body. Potential solutions to these issues could entail providing a mini-map [12, 57] as an overlay to give an non-occluded vision to support fine-grained transformation and making the oversized object semi-transparent to minimize its intrusiveness.

6.5.3 Implicit vs. Explicit Eye-Hand Transitions. Participants' opinions differed in whether it would be more beneficial to apply explicit or implicit transitions between eye and hand input. The advocates of the explicit transition mechanism most appreciated its robustness; the rapid eye movement would not frequently bring the object to the user's facing direction. Although the dynamically-resized safe region was reported as being useful (*ImplicitGaze* did not produce random gaze-like movement for objects), it was not able to handle rapid, long-distance searching actions, and could occlude participants' view by snapping the target to the gaze location. As mentioned previously, making the target semi-transparent would mitigate this issue.

On the other hand, some participants found that using hand movement to confirm the gaze action was somewhat redundant. Moreover, because of the separate nature of gaze and hand input, participants noticed that it was sometimes challenging to determine whether they were using hand or gaze input. Additionally, it can be confusing for users when they actually want to use gaze to translate an object, but because the gaze point is still located inside the safe region, only the hand movement (which was meant to be a trigger action) affected objects' location. In these scenarios, it would be helpful to provide a small widget to indicate which input modality is taking control of the object for explicit transition based techniques.

Also, as suggested by participants, it would be beneficial to provide hand amplification [69] for both *ImplicitGaze* and *3DMagicGaze* in the depth dimension to speed up the translation along the z-axis.

6.5.4 Gaze-Supported Techniques vs. Virtual Hand. As indicated in Table 2, the results from the user experiences questionnaire suggest that the current market-available solution (Virtual Hand) was not sufficient for target manipulation tasks in VR, while gaze-supported techniques lead to pragmatic and hedonic improvements. Despite being "natural" and "realistic", Virtual Hand was seen as not being an efficient, convenient, and comfortable solution for long-term object manipulation in VR.

7 DESIGN IMPLICATIONS

We derived a set of design implications for future gaze-supported manipulation techniques in VR. We do not advocate a one-size-fits-all technique, as different design features can be useful for different environments and task proposes. Instead, we summarize their strengths, possible applications, and provide potential compensation for their weaknesses.

- While embedding gaze input (like *ImplicitGaze* and *3DMagicGaze*) might not offer significant performance benefits for manipulating (translating) objects that are within the primary working space (that is, all targets are located in front of the user and within arm-reach distance), it can be useful for a larger environment with distant objects.
- If gaze input is used for object selection and only hand input is used for manipulation, consider adding hand amplification (e.g., [69]) when users need to manipulate objects that are outside of the primary working space. Otherwise, it can feel tiresome to manipulate remotely gaze-selected objects.
- The hand-eye coordination strategy which snaps the target to the hand position when selection is triggered is efficient for bringing distant objects to the user. However, this function may require the user to teleport to different places frequently when working in a large environment. Therefore, a complementary precise and convenient teleportation mechanism is needed. Additionally, we suggest disabling the snap-to-hand function for objects within arm-reach distance, as repetitive snapping close objects to hands can cause confusion and make the fine-grained adjustment difficult.
- While manipulating an object directly via hands is intuitive, it may lead to more arm fatigue as users need to hold their arms in the air. One could consider minimizing the duration of using such direct-mapping and use indirect-mapping techniques (like *RemoteHand*) which allow users to rest their arms under a comfortable position. Also, large objects can easily occlude users' view and pose difficulties for accurate manipulation. Therefore, providing an accompanying mini-map (e.g., [57, 73, 74]) as an overlay would provide an overview of the environment, while making the oversized object transparent to reduce intrusiveness.
- Providing an implicit transition between gaze and hand input (such as *ImplicitGaze*) can enable the smooth and concurrent transformation. It would be useful to consider applying a dynamically-resized safe region (as used in this research) to reduce the random movement of objects caused by eye saccades. Note there are also other design opportunities to enable implicit transitions. For example, designers may choose to use a probabilistic/heuristic model to implicitly determine whether gaze or hand should take control of the target. Also, we suggest making the object under manipulation semi-transparent to avoid visual occlusion while searching.
- Explicit transition (like *3DMagicGaze*) enables robust control over the effect of gaze on objects. However, some effort is required in performing the 'switch' command and users may be unsure about whether to make a 'switch' or not. We recommend adding a small widget to indicate which input modality is currently taking control of the manipulation.
- For techniques that use both gaze and hand for manipulation (e.g., *ImplicitGaze* and *3DMagicGaze*), hand amplification in the depth dimension would be beneficial to speed up the translation along the z-axis when interacting with objects outside of the primary working space.

8 LIMITATIONS AND FUTURE WORK

We have identified several limitations in this research. First, we did not embed techniques that enable non-linear mapping of hand input [69], as our primary focus was gaze input. Hand amplification can interplay with or enhance gaze input, and it would be interesting to investigate how they influence one another. Second, we did not explore the long-term usage of gaze-supported manipulation techniques. For instance, if 3D modelers used gaze input every day, they would probably find even more efficient ways of using them. Third, we did not test the methods alongside more complex sculpturing and modeling tools/functions (like smoothing and inflating an object). Further research can extend the gaze input modality to accompany more advanced manipulation functions. Fourth, we treated gaze as a 2 DoF modality and thus explored more of its usage for translating objects in the lateral direction. However, we acknowledge that there is a potential of using gaze for rotation and scaling with novel approaches. Lastly, as head gaze can be a cheaper solution than eye gaze for current VR systems, it is worth exploring if head gaze possesses similar features as eye gaze for object manipulation.

9 CONCLUSION

In this research, we explore gaze-supported 3D object manipulation in VR. Specifically, we investigate how different ways of integrating, coordinating, and transitioning gaze and hand input can aid the existing approach based on the virtual hand. Results from two user studies evaluating and comparing four techniques regarding their usability and efficiency show that gaze input does not offer significant performance benefits for object manipulation in the primary working space (when all targets are located in front of the user and within arm-reach distance), but can be useful for larger spaces with distant objects. Gaze input was also shown to mitigate the arm fatigue issue, and different integration, coordination, and transition strategies can provide benefits for building more usable and efficient object manipulation techniques. Our work contributes novel insights regarding multimodal interfaces with gaze and hand input that can enhance existing and future 3D object manipulation solutions in VR.

ACKNOWLEDGMENTS

This research is partially funded by the Melbourne Research Scholarship provided by The University of Melbourne. It is also supported in part by Xi'an Jiaotong-Liverpool University (XJTLU) Key Special Fund (KSF-A-03). Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Award (Project Number: DE180100315) funded by the Australian Government.

REFERENCES

[1] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136. https://doi.org/10.1016/j.cag.2012.12.003
[2] Benjamin Bach, Ronell Sicat, Johanna Beyer, Maxime Cordeil, and Hanspeter Pfister. 2017. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 457–467. https://doi.org/10.1109/TVCG.2017.2745941
[3] Costas Boletsis. 2017. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal Technologies and Interaction* 1, 4 (2017), 24. https://doi.org/10.3390/mti1040024
[4] Gunnar AV Borg. 1982. Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise* (1982).
[5] Doug A. Bowman and Larry F. Hodges. 1997. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. Association for Computing Machinery, New York, NY, USA, 35–ff. https://doi.org/10.1145/253284.253301
[6] Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. 2001. Testbed Evaluation of Virtual Environment Interaction Techniques. *Presence: Teleoperators and Virtual Environments* 10, 1 (2001), 75–95. https://doi.org/10.1162/105474601750182333
[7] Doug A. Bowman, Ryan P. McMahan, and Eric D. Ragan. 2012. Questioning Naturalism in 3D User Interfaces. *Commun. ACM* 55, 9 (Sept. 2012), 78–88. https://doi.org/10.1145/2330667.2330687
[8] Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. 2019. Investigating Smartphone-Based Pan and Zoom in 3D Data Spaces in Augmented Reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) *(MobileHCI '19)*. Association for Computing Machinery, New York, NY, USA, Article 2, 13 pages. https://doi.org/10.1145/3338286.3340113
[9] Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Seattle, Washington, USA) *(ICMI '15)*. Association for Computing Machinery, New York, NY, USA, 131–138. https://doi.org/10.1145/2818346.2820752
[10] Shujie Deng, Nan Jiang, Jian Chang, Shihui Guo, and Jian J Zhang. 2017. Understanding the impact of multimodal interaction using gaze informed mid-air gesture control in 3D virtual objects manipulation. *International Journal of Human-Computer Studies* 105 (2017), 68–80. https://doi.org/10.1016/j.ijhcs.2017.04.002
[11] Joachim Diepstraten, Daniel Weiskopf, and Thomas Ertl. 2002. Transparency in interactive technical illustrations. In *Computer Graphics Forum*, Vol. 21. Wiley Online Library, 317–325. https://doi.org/10.1111/1467-8659.t01-1-00591
[12] Niklas Elmqvist and Philippas Tsigas. 2007. A taxonomy of 3D occlusion management techniques. In *2007 IEEE Virtual Reality Conference*. IEEE, 51–58. https://doi.org/10.1109/VR.2007.352463
[13] Barrett M. Ens, Rory Finnegan, and Pourang P. Irani. 2014. The Personal Cockpit: A Spatial Interface for Effective Task Switching on Head-Worn Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3171–3180. https://doi.org/10.1145/2556288.2557058
[14] Facebook. [n.d.]. *Oculus Medium*. Retrieved September 6, 2020 from https://www.oculus.com/medium/
[15] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1118–1130. https://doi.org/10.1145/3025453.3025599
[16] Tiare Feuchtner and Jörg Müller. 2018. Ownershift: Facilitating Overhead Interaction in Virtual Reality with an Ownership-Preserving Hand Space Shift. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 31–43. https://doi.org/10.1145/3242587.3242594
[17] S. Frees and G. D. Kessler. 2005. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* 99–106. https://doi.org/10.1109/VR.2005.1492759
[18] Epic Games. [n.d.]. *Unreal Engine VR Mode*. Retrieved September 6, 2020 from https://docs.unrealengine.com/en-US/Engine/Editor/VR/index.html/
[19] Christoph Gebhardt, Brian Hecox, Bas van Opheusden, Daniel Wigdor, James Hillis, Otmar Hilliges, and Hrvoje Benko. 2019. Learning Cooperative Personalized Policies from Gaze Data. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 197–208. https://doi.org/10.1145/3332165.3347933
[20] Google. [n.d.]. *Introducing Blocks*. Retrieved September 6, 2020 from https://arvr.google.com/blocks/
[21] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908. https://doi.org/10.1177/154193120605000909
[22] Devamardeep Hayatpur, Seongkook Heo, Haijun Xia, Wolfgang Stuerzlinger, and Daniel Wigdor. 2019. 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* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1185–1195. https://doi.org/10.1145/3332165.3347916
[23] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130
[24] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, Article 625, 12 pages. https://doi.org/10.1145/3290605.3300855
[25] Mozilla Hubs. [n.d.]. *Hubs - Private social VR in your web browser*. Retrieved August 19, 2020 from https://hubs.mozilla.com/
[26] Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze Gestures or Dwell-Based Interaction?. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 229–232. https://doi.org/10.1145/2168556.2168602
[27] VRChat Inc. [n.d.]. *VRChat*. Retrieved September 6, 2020 from https://www.vrchat.com/
[28] Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-Based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246
[29] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3328–3339. https://doi.org/10.1145/3025453.3025523
[30] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article 81, 14 pages.

https://doi.org/10.1145/3173574.3173655

[31] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.

[32] Johnson Ching-Hong Li. 2016. Effect size measures in a two-independent-samples case with nonnormal and nonhomogeneous data. *Behavior research methods* 48, 4 (2016), 1560–1574. https://doi.org/10.3758/s13428-015-0667-z

[33] Zhenxing Li, Deepak Akkil, and Roope Raisamo. 2019. Gaze Augmented Hand-Based Kinesthetic Interaction: What You See is What You Feel. *IEEE transactions on haptics* 12, 2 (2019), 114–127. https://doi.org/10.1109/TOH.2019.2896027

[34] Chang Liu, Alexander Plopski, and Jason Orlosky. 2020. OrthoGaze: Gaze-based Three-dimensional Object Manipulation using Orthogonal Planes. *Computers & Graphics* (2020). https://doi.org/10.1016/j.cag.2020.04.005

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

[36] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. 2019. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum* 38, 1 (2019), 21–45. https://doi.org/10.1111/cgf.13390

[37] Daniel Mendes, Filipe Relvas, Alfredo Ferreira, and Joaquim Jorge. 2016. The Benefits of DOF Separation in Mid-Air 3D Object Manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich, Germany) *(VRST '16)*. Association for Computing Machinery, New York, NY, USA, 261–268. https://doi.org/10.1145/2993369.2993396

[38] Minecraft. [n.d.]. *Minecraft Official Site*. Retrieved Augest 19, 2020 from https://www.minecraft.net/

[39] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-Touch: Combining Gaze with Multi-Touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[40] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[41] Jeffrey S. Pierce, Brian C. Stearns, and Randy Pausch. 1999. Voodoo Dolls: Seamless Interaction at Multiple Scales in Virtual Environments. In *Proceedings of the 1999 Symposium on Interactive 3D Graphics* (Atlanta, Georgia, USA) *(I3D '99)*. Association for Computing Machinery, New York, NY, USA, 141–145. https://doi.org/10.1145/300523.300540

[42] Matti Pouke, Antti Karhu, Seamus Hickey, and Leena Arhippainen. 2012. Gaze Tracking and Non-Touch Gesture Based Interaction Method for Mobile 3D Virtual Spaces. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (Melbourne, Australia) *(OzCHI '12)*. Association for Computing Machinery, New York, NY, USA, 505–512. https://doi.org/10.1145/2414536.2414614

[43] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. 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* (Seattle, Washington, USA) *(UIST '96)*. Association for Computing Machinery, New York, NY, USA, 79–80. https://doi.org/10.1145/237091.237102

[44] Warren Robinett and Richard Holloway. 1992. Implementation of Flying, Scaling and Grabbing in Virtual Worlds. In *Proceedings of the 1992 Symposium on Interactive 3D Graphics* (Cambridge, Massachusetts, USA) *(I3D '92)*. Association for Computing Machinery, New York, NY, USA, 189–192. https://doi.org/10.1145/147156.147201

[45] Kunhee Ryu, Joong-Jae Lee, and Jung-Min Park. 2019. GG Interaction: a gaze–grasp pose interaction for 3D virtual object selection. *Journal on Multimodal User Interfaces* 13, 4 (2019), 383–393. https://doi.org/10.1007/s12193-019-00305-y

[46] Jeff Sauro and Joseph S. Dumas. 2009. Comparison of Three One-Question, Post-Task Usability Questionnaires. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1599–1608. https://doi.org/10.1145/1518701.1518946

[47] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *IJIMAI* 4, 6 (2017), 103–108. https://doi.org/10.9781/ijimai.2017.09.001

[48] Marcos Serrano, Barrett Ens, Xing-Dong Yang, and Pourang Irani. 2015. Gluey: Developing a Head-Worn Display Interface to Unify the Interaction Experience in Distributed Display Environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 161–171. https://doi.org/10.1145/2785830.2785838

[49] Ludwig Sidenmark and Hans Gellersen. 2019. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article 4 (Dec. 2019), 40 pages. https://doi.org/10.1145/3361218

[50] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

[51] Ludwig Sidenmark, Diako Mardanbegi, Argenis Ramirez Gomez, Christopher Clarke, and Hans Gellersen. 2020. BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 8, 9 pages. https://doi.org/10.1145/3379155.3391312

[52] Adalberto L. Simeone, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2016. Three-Point Interaction: Combining Bi-Manual Direct Touch with Gaze. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (Bari, Italy) *(AVI '16)*. Association for Computing Machinery, New York, NY, USA, 168–175. https://doi.org/10.1145/2909132.2909251

[53] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2018. Saliency in VR: How do people explore virtual environments? *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1633–1642. https://doi.org/10.1109/TVCG.2018.2793599

[54] Dana Slambekova, Reynold Bailey, and Joe Geigel. 2012. Gaze and Gesture Based Object Manipulation in Virtual Worlds. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology* (Toronto, Ontario, Canada) *(VRST '12)*. Association for Computing Machinery, New York, NY, USA, 203–204. https://doi.org/10.1145/2407336.2407380

[55] Sophie Stellmach and Raimund Dachselt. 2013. Still Looking: Investigating Seamless Gaze-Supported Selection, Positioning, and Manipulation of Distant Targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 285–294. https://doi.org/10.1145/2470654.2470695

[56] Sophie Stellmach, Sebastian Stober, Andreas Nürnberger, and Raimund Dachselt. 2011. Designing Gaze-Supported Multimodal Interactions for the Exploration of Large Image Collections. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications* (Karlskrona, Sweden) *(NGCA '11)*. Association for Computing Machinery, New York, NY, USA, Article 1, 8 pages. https://doi.org/10.1145/1983302.1983303

[57] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 265–272. https://doi.org/10.1145/223904.223938

[58] Amedeo Troiano, Francesco Naddeo, Erik Sosso, Gianfranco Camarota, Roberto Merletti, and Luca Mesin. 2008. Assessment of force and fatigue in isometric contractions of the upper trapezius muscle by surface EMG signal and perceived exertion scale. *Gait & Posture* 28, 2 (2008), 179 – 186. https://doi.org/10.1016/j.gaitpost.2008.04.002

[59] Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. 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* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 4179–4188. https://doi.org/10.1145/2702123.2702355

[60] Jayson Turner, Jason Alexander, Andreas Bulling, Dominik Schmidt, and Hans Gellersen. 2013. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *IFIP Conference on Human-Computer Interaction*. Springer, 170–186. https://doi.org/10.1007/978-3-642-40480-1_11

[61] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2013. Eye Drop: An Interaction Concept for Gaze-Supported Point-to-Point Content Transfer. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (Luleå, Sweden) *(MUM '13)*. Association for Computing Machinery, New York, NY, USA, Article 37, 4 pages. https://doi.org/10.1145/2541831.2541868

[62] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2014. Cross-Device Gaze-Supported Point-to-Point Content Transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Safety Harbor, Florida) *(ETRA '14)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/2578153.2578155

[63] Jayson Turner, Andreas Bulling, and Hans Gellersen. 2011. Combining Gaze with Manual Interaction to Extend Physical Reach. In *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction* (Beijing, China) *(PETMEI '11)*. Association for Computing Machinery, New York, NY, USA, 33–36. https://doi.org/10.1145/2029956.2029966

[64] András Vargha and Harold D Delaney. 2000. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics* 25, 2 (2000), 101–132. https://doi.org/10.3102/10769986025002101

[65] Manuel Veit, Antonio Capobianco, and Dominique Bechmann. 2009. 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* (Kyoto, Japan) *(VRST*

'09). Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/1643928.1643942

[66] Eduardo Velloso, Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. In *Human-Computer Interaction – INTERACT 2015*. Springer International Publishing, Cham, 315–330. https://doi.org/10.1007/978-3-319-22668-2_25

[67] Simon Voelker, Sebastian Hueber, Christian Holz, Christian Remy, and Nicolai Marquardt. 2020. GazeConduits: Calibration-Free Cross-Device Collaboration through Gaze and Touch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376578

[68] Tobii VR. [n.d.]. *Hand-Eye Coordination*. Retrieved Augest 12, 2020 from https://vr.tobii.com/sdk/develop/unity/unity-examples/hand-eye-coordination/

[69] Johann Wentzel, Greg d'Eon, and Daniel Vogel. 2020. Improving Virtual Reality Ergonomics Through Reach-Bounded Non-Linear Input Amplification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376687

[70] Curtis Wilkes and Doug A. Bowman. 2008. Advantages of Velocity-Based Scaling for Distant 3D Manipulation. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology* (Bordeaux, France) *(VRST '08)*. Association for Computing Machinery, New York, NY, USA, 23–29. https://doi.org/10.1145/1450579.1450585

[71] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[72] Yukang Yan, Chun Yu, Xiaojuan Ma, Shuai Huang, Hasan Iqbal, and Yuanchun Shi. 2018. Eyes-Free Target Acquisition in Interaction Space around the Body for Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173616

[73] Difeng Yu, Hai-Ning Liang, Kaixuan Fan, Heng Zhang, Charles Fleming, and Konstantinos Papangelis. 2020. Design and Evaluation of Visualization Techniques of Off-Screen and Occluded Targets in Virtual Reality Environments. *IEEE Transactions on Visualization and Computer Graphics* 26, 9 (2020), 2762–2774. https://doi.org/10.1109/TVCG.2019.2905580

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

[75] Difeng Yu, Qiushi Zhou, Benjamin Tag, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Engaging Participants during Selection Studies in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 500–509. https://doi.org/10.1109/VR46266.2020.00071

[76] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053