Experimental Analysis of Freehand Multi-object Selection Techniques in Virtual Reality Head-Mounted Displays

RONGKAI SHI*, The Hong Kong University of Science and Technology (Guangzhou), China
YUSHI WEI*, The Hong Kong University of Science and Technology (Guangzhou), China
XUNING HU, Xi'an Jiaotong-Liverpool University, China
YU LIU, Xi'an Jiaotong-Liverpool University, China
YONG YUE, Xi'an Jiaotong-Liverpool University, China
LINGYUN YU, Xi'an Jiaotong-Liverpool University, China
HAI-NING LIANG†, The Hong Kong University of Science and Technology (Guangzhou), China

Object selection is essential in virtual reality (VR) head-mounted displays (HMDs). Prior work mainly focuses on enhancing and evaluating techniques for selecting a single object in VR, leaving a gap in the techniques for multi-object selection, a more complex but common selection scenario. To enable multi-object selection, the interaction technique should support group selection in addition to the default pointing selection mode for acquiring a single target. This composite interaction could be particularly challenging when using freehand gestural input. In this work, we present an empirical comparison of six freehand techniques, which are comprised of three mode-switching gestures (Finger Segment, Multi-Finger, and Wrist Orientation) and two group selection techniques (Cone-casting Selection and Crossing Selection) derived from prior work. Our results demonstrate the performance, user experience, and preference of each technique. The findings derive three design implications that can guide the design of freehand techniques for multi-object selection in VR HMDs.

CCS Concepts: • **Human-centered computing → Virtual reality**; **Gestural input**; **Empirical studies in interaction design**.

Additional Key Words and Phrases: Virtual Reality, Object Selection, Target Acquisition, Multi-object Selection, Mid-Air Interaction, Freehand Interaction, Gestural Input, Head-Mounted Display

ACM Reference Format:
Rongkai Shi, Yushi Wei, Xuning Hu, Yu Liu, Yong Yue, Lingyun Yu, and Hai-Ning Liang. 2024. Experimental Analysis of Freehand Multi-object Selection Techniques in Virtual Reality Head-Mounted Displays. *Proc. ACM Hum.-Comput. Interact.* 8, ISS, Article 529 (December 2024), 19 pages. https://doi.org/10.1145/3698129

*Part of this work was conducted when both authors were with Xi'an Jiaotong-Liverpool University.
†Corresponding author.

Authors' Contact Information: Rongkai Shi, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, rongkaishi@hkust-gz.edu.cn; Yushi Wei, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, ywei662@connect.hkust-gz.edu.cn; Xuning Hu, Xi'an Jiaotong-Liverpool University, Suzhou, China, xuning.hu22@student.xjtlu.edu.cn; Yu Liu, Xi'an Jiaotong-Liverpool University, Suzhou, China, yu.liu02@xjtlu.edu.cn; Yong Yue, Xi'an Jiaotong-Liverpool University, Suzhou, China, yong.yue@xjtlu.edu.cn; Lingyun Yu, Xi'an Jiaotong-Liverpool University, Suzhou, China, lingyun.yu@xjtlu.edu.cn; Hai-Ning Liang, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China, hainingliang@hkust-gz.edu.cn.

ACM 2573-0142/2024/12-ART529
https://doi.org/10.1145/3698129

1 Introduction

Object selection (or target acquisition) is a fundamental interaction in virtual reality (VR). It is typically the initial step to complete other canonical manipulation tasks, including positioning, rotation, and scaling, and is indispensable for composite workflows [1, 15]. Numerous prior work has proposed interaction techniques to improve object selection and enable their use in complex VR scenarios [1, 2, 15], such as selecting a target that is small, out-of-reach, (e.g., [20, 28]), or occluded by others (e.g., [18, 49]). However, most of these interaction techniques share the same goal—to acquire a single target, leaving fewer discussions on multi-object selection, which is common in VR applications involving large amounts of selectable objects, such as astronomical data exploration [50] and 3D modeling design [33, 47].

When a user has the same operation intention for multiple objects, selecting them first and manipulating them as a group would be less time-consuming and tedious than repetitively working with each object, especially when there are many intended targets or the manipulation requires high precision [17, 33, 46]. Furthermore, performing a group selection could be effortless if the intended targets are located in a particular area [32]. Compared to single-object selection, multi-object selection can be relatively more challenging because it requires additional iterations for refining the selection, like deselecting unwanted objects or appending other non-selected targets, and in the meantime, extra effort for holding the selection state for selected targets while completing the refinement. Moreover, as a special case of single-object selection, enabling multi-object selection introduces an additional mode to the interaction scenario, which may make users more prone to make mistakes when performing or transitioning between the two types of selection.

The multi-object selection techniques proposed in prior work are largely based on handheld devices, like pen-tablet combinations [17] or controllers [46, 47], while limited supports freehand gestural input, which is becoming popular. Freehand gestural input is controller-free and has the potential to facilitate effective, natural, immersive interaction and communication [6]. It is supported by current VR head-mounted displays (HMDs) without extra devices and has become an alternative to handheld controllers. However, there are several challenges for freehand gestural interaction, such as relatively high learning costs, limited available delimiters/triggers, lack of tactile feedback, and absence of critical clues for interaction [23, 30]. Moreover, prior work mainly focused on near-field virtual-hand-based selection for multiple objects, leaving a gap in far-field virtual-pointing-based selection, which is also common and essential in the immersive VR environment to overcome physical constraints [1, 21]. Given these challenges from both the selection task and the gestural input, enabling precise and efficient freehand selection for multiple objects in addition to pointing selection of a single object requires careful designs.

Thus, our research goal is to **design and evaluate freehand techniques for multi-object selection in VR HMDs**. To achieve this goal, we first frame three design goals in the context of related work to guide the interaction design. We then analyzed the interaction process of selecting multiple objects and derived three pivotal actions for this process: default single-object selection, group-based multi-object selection, and mode switching. The thumb-to-index pinch gesture, the most widely adopted hand gesture for freehand pointing selection, has been chosen for single-object selection. Building upon this, six potential techniques were proposed and selected for evaluation. These techniques are the combinations of three mode-switching gestures (Finger Segment, Multi-Finger, and Wrist Orientation) and two group selection techniques (Cone-casting Selection and Crossing Selection) derived from prior work. They were compared empirically via a user study with eighteen participants in randomized scenarios. We found Crossing Selection outperformed Cone-casting Selection while the latter was not disliked by participants. The three mode-switching gestures led to similar performance and user experience. Participants tended to like Multi-Finger

and to dislike Wrist Orientation. Our findings are useful for the future design of freehand interaction techniques for multiple objects in VR environments.

This work makes the following three main contributions:

- We articulate three design goals for freehand multi-object selection in VR based on a synthesis of previous work (Section 2).
- Based on the design goals, we propose six freehand multi-object selection techniques that combine three mode-switching gestures and two group selection techniques (Section 3).
- We empirically compare the six techniques via a user study and derive insights for future design and development of freehand techniques for multi-object selection in VR (Sections 4, 5, and 6).

2 Design Goals and Related Work

We identified three design goals for a freehand interaction technique supporting precise and efficient multi-object selection in VR. In this section, we frame these design goals in the context of related work.

2.1 Design Goal 1: Build upon General Interaction Metaphors for Freehand Object Selection

There are two major interaction metaphors for object selection in VR environments [1]—*virtual hand* [22] and *virtual pointing* [21]. The interaction provided by virtual hand techniques is widely considered natural and intuitive because users interact with virtual objects in a similar way as they do in the real world. However, this mapping also limits the use of the original virtual hand technique because the interaction only happens in users' reachable areas. Two common solutions are identified for letting users select out-of-reach objects, 1) using a technique that sends the virtual hand out and controls it remotely, such as Go-Go [28], and its extensions (e.g., [10, 44]); and 2) using virtual pointing techniques. The most common virtual pointing technique is ray-casting [21]. With a ray-casting technique, the user casts a ray emitted from the controller or the hand, allowing the user to select an object at a distance. However, its performance also suffers from difficulties in selecting small objects and hand jitter issues when triggering the selection (i.e., the Heisenberg effect [4]). The pointing accuracy has been improved via correction models from both the human side [8] and the system side [19]. To summarize, the two major interaction metaphors have their advantages and disadvantages.

Currently, top VR headset and hand-tracking accessory companies suggest a combined use of virtual hand and virtual pointing techniques for their hand-tracking solutions (see, for example, Meta Quest[1], HTC VIVE[2], and Ultraleap[3]). A user can tap on a target positioning within their reach for direct selection. On the other hand, the user can point the ray to a target and then pinch their thumb and index finger together for distant selection. This work considers this general metaphor group the foundation of enhanced technique, though there are also other hand gestures feasible for object selection (e.g., bending the thumb for selection [14]). We discuss the design of selection techniques in detail in Section 3.

[1] https://www.meta.com/en-gb/help/quest/articles/headsets-and-accessories/controllers-and-hand-tracking/hand-tracking-quest-2/
[2] https://www.vive.com/au/support/focus3/category_howto/hand-tracking.html
[3] https://docs.ultraleap.com/xr-guidelines/Getting%20started/design-principles.html

2.2 Design Goal 2: Facilitate Effective Multi-object Selection

There are two main approaches to selecting multiple objects: selecting objects serially or selecting by group; that is, selecting one object versus one or more objects per selection operation [15, 17, 46].

The single-object selection techniques can be used for serial selection. Prior empirical results have shown that serial selection is necessary for certain scenarios because adding or removing an object that is more challenging using group selection metaphors (like a distractor surrounded by targets) is unavoidable [17, 46]. We identified two ways to provide the serial selection mode. In the daily use of a desktop computer, users can activate the serial selection mode by pressing the shift key. While prior work focusing on multi-object selection in VR did not distinguish between these two operations—a new selection will not change the selection states of others [17, 46]. This work followed the way of a desktop computer to enrich VR interactions (single-object selection and serial selection as two operations), which is still underexplored.

Four existing metaphors are possibly suitable for group selection. One such metaphor is goal crossing, which has been introduced to various selection scenarios in HMDs [13, 40, 41]. Simply put, it works like a brush and can select an object by interacting with its boundary. With the selection activated and maintained, users can select multiple objects. Notably, the ray-casting crossing has been verified as a feasible complement to the ray-casting pointing [40]. In the desktop and tablet interfaces, users can 'click and drag' to complete a rectangle selection for selecting multiple targets. Rectangle selection has been adapted to the 3D world by Shi et al. [32], who explored gaze-assisted and hand-based region selection methods in AR HMD. In their hand-only region selection method, users can pinch and drag to formulate a rectangular region, which can potentially be used to select the objects cast by this region. Except for a 2D region, prior studies have also explored selecting objects via a 3D volume in VR environments. For example, Lucas [17] allowed multi-object selection by creating a cuboid region via tablet and stylus input. Wu et al. [46] adapted this approach to VR controllers, enabling group selection/deselection in the near-field with the simulated virtual hands. Another 3D volume proposed in the literature is a spherical container, see for example, Poros [27], SpaceTime [47], and BodyOn [48]. These works also focused on the interaction within arm's reach. For far-field techniques, cone-casting, which replaced the ray with a cone or a spotlight, may be suitable for group selection. It is a widely investigated pointing technique for assisting target selection in dense environments or small objects at a distance (e.g., [20, 34, 49]). Cone-casting makes the single-target acquisition easier and more precise by enabling users to select the target from a small group of objects, which is pre-selected via the cone and rearranged to a structured layout. Its first step is an obvious multi-object selection process but yet to be examined.

As seen above, the prior work has provided a few interaction approaches and metaphors with good potential for multi-object selection. This work proposes effective freehand techniques for multi-object selection based on their exploration.

2.3 Design Goal 3: Enable Rapid and Seamless Mode Switching

Mode switching is the transition between different modes, which allows users to achieve different outcomes with the same input [29]. Researchers have highlighted the importance of designing suitable mode-switching methods for freehand selection and manipulation [7, 34, 48] because the same action is inevitably needed for multiple purposes. Furthermore, rapid and seamless mode switching is necessary and particularly important for freehand multi-object selection, given the challenges we identified in Section 1.

There are three common mode-switching mechanisms. First, a mode can be sustained by the system (i.e., *system-maintained* [31]). The system persistently activates the selected mode until the user switches to others (e.g., the Caps Lock key). It is useful when several operations are

pending in the selected mode. Second, a mode can be maintained by users (*user-maintained* [31] or *quasi modes* [12, 29]), which requires users to 'hold' the switching action as long as the mode is needed. A user-maintained mechanism can help reduce mode errors [29, 31] and is suitable for temporary use [12]. Third, a mode is manually activated for one use only and then automatically returns to the previous mode (*Once* [12]), which can be ideal for certain cases, such as typing the first letter of a sentence. This work mainly focused on the user-maintained mechanism as we consider multi-object selection a special and temporarily performed task in addition to the regular single-object selection. User-maintained mode-switching methods are also widely studied for different VR tasks and reported to be usable in prior work [35, 38, 39, 43].

In freehand interaction, hand postures can be used to distinguish between different modes. When both hands are available, the dominant hand is assigned to the primary task, leaving the non-dominant hand to control the mode naturally. Using non-dominant hand posture for mode switching has been proven to be efficient and accurate [24, 36, 39]. For one-hand cases, Surale et al. [39] compared hand gestures empirically and suggested subtle dominant hand postures, such as rotating the wrist or using the middle finger to distinguish from the default thumb-to-index pinch. Similarly, Song et al. [38] enabled efficient keyboard switching for freehand text entry by rotating the wrist or extending the middle finger while performing the finger-touch selection. On the other hand, Yu et al. [48] used tapping the thumb on different fingertips or finger segments to achieve various manipulations with three levels of control-display ratios. These small adjustments to the hand gesture for mode switching were considered in this work (see Section 3). Introducing a secondary input modality for mode switching to support the interaction is a huge branch of related research (e.g., voice [7, 37], eye gaze [25, 26], or head movement [7, 35, 42]), but it is out of the scope of this paper.

3 Design of Techniques

We aim to investigate user performance and experience of possible freehand techniques for multi-object selection in VR HMDs. In this section, we first propose a set of considerations to formulate the problem (Section 3.1), followed by an analysis of the interaction process (Section 3.2). Based on this, we present 12 potential technique combos consisting of 3 mode-switching techniques and 4 group selection techniques, and 6 were selected for evaluation (Sections 3.3 and 3.4).

3.1 Considerations

To begin with, we describe several considerations that narrow down the problem space of this work. First, our domain selection metaphors are based on an egocentric point of view (first-person view), which is the most common for immersive environments and receives much attention from prior work on object selection. The interaction design of multi-object selection is built upon virtual-pointing-based selection for acquiring a single object. Second, we focus on selection-intensive scenarios that do not involve navigation. Though navigate-to-select approaches are interesting and important, they may generate potential issues beyond selection [1], which are outside this paper's scope. Third, we do not consider intelligent grouping facilitated by the system (e.g., [50]). In other words, the selection can only be driven and completed by users' intentions. Finally, we define *freehand* as input performed entirely by a hand gesture or movement. In this work, we utilized the self-built hand-tracking modules on VR HMDs to track freehand interaction, which we believe is low-cost and will be commonly available on future HMDs. Even so, we admitted this solution was imperfect, and as such, we focused on and ensured that the proposed interaction was doable for current commercial HMDs.

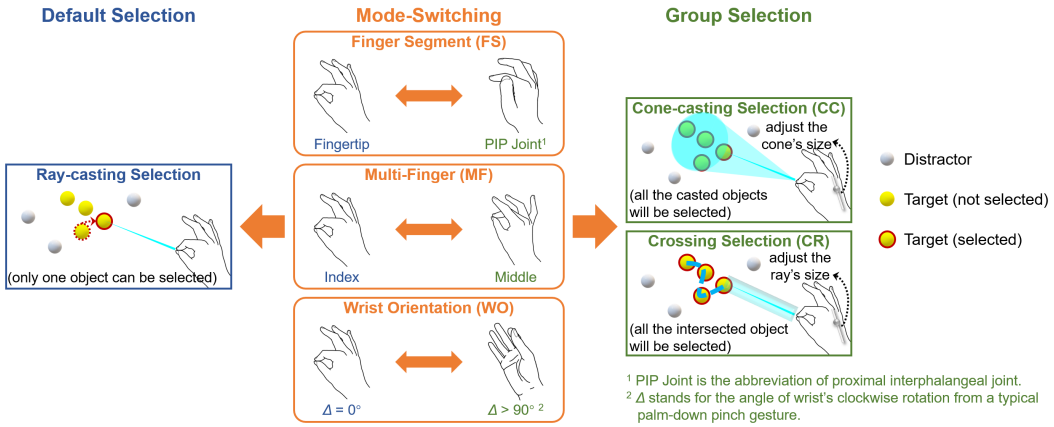

Fig. 1. The freehand techniques in each part of the multi-object selection process: a ray-casting technique is used for the default selection, Cone-casting Selection or Crossing Selection can do group selection, and the two modes can be transited using Finger Segment, Multi-Finger, or Wrist Orientation gestures.

3.2 Interaction Process

We view multi-object selection as an auxiliary case of single-object selection, not a standalone and individual task. Thus, single-object selection is still the default mode for interaction, while users can switch to the multi-object selection mode seamlessly and have a fluid workflow for the whole process. On the other hand, as shown in Section 2.2, prior work has shown empirical evidence suggesting serial selection for randomized target layout [46]. Thus, both default single-object and serial selection were included in the multi-object selection workflow. For simplicity and clarity, we call them **Default Selection** and **Serial Selection** to describe the object-of-interest cases in which only one object can be selected (like click on a PC mouse) and multiple objects can be selected ('shift + click'), respectively. Additionally, we use the term **Group Selection** to describe the action of acquiring several targets at a time. Users can switch between them via **Mode-Switching** methods. When a user performs serial selection or group selection to objects that have been selected, the selection states of those objects would be canceled (i.e., deselection).

However, in our pilot tests, we faced similar hand-tracking issues as reported in prior work (e.g., [34]). We found it difficult for current VR HMDs to track accurately the micro-gestures for switching between three modes.[4] We were also concerned about the high learning cost of gestures for people new to VR and mid-air interaction [23, 30]. Thus, we improved the interaction process by **making the serial selection an implicit part of group selection**, which does not require an explicit mode switch to distinguish between them. This approach has a lower hand-tracking requirement for HMDs but also eases users' learning process in remembering gestures.

3.3 Potential Techniques

Based on the considerations and the analysis of the interaction process, we derived three parts for multi-object selection: default selection, group selection, and mode switching. As shown in Section 2.1 (Design Goal 1), a ray-casting technique is the most widely used pointing selection technique in VR HMDs. Thus, we use it for default selection. Following the design goals, we propose potential mode-switching and group selection techniques in the next sections. The whole design is illustrated in Figure 1.

[4]The tested HMDs included Quest 2, 3, and Pro.

3.3.1 Mode-Switching Techniques. The default selection is achieved via a freehand ray-casting technique, wherein bringing the dominant hand's fingertips of the thumb and index finger together with the palm facing down (a standard pinch gesture) confirms the selection. We made small adjustments to this pinch gesture to enable a smooth transition to the group selection for our Design Goal 3 (see Section 2.3). In addition, we concentrated on one-handed mode-switching gestures to reduce the occupation of the non-dominant hand, which may introduce extra fatigue or be used for other interactions.

- **Finger Segment (FS)**: Instead of the fingertip of the index finger, the user uses her thumb to tap on the proximal interphalangeal (PIP) joint of the index finger to confirm a group selection.
- **Multi-Finger (MF)**: The user pinches her thumb and middle finger to trigger group selection.
- **Wrist Orientation (WO)**: When the user's wrist is rotated clockwise from her perspective for more than 90°, group selection will be executed if the pinch gesture is performed.

3.3.2 Group Selection Techniques.

- **Cone-casting Selection (CC)**: The user controls a cone emitted from hand for group selection. Once the user confirms the selection, all the projected objects are selected (or deselected if one was under selection previously). The user can adjust the size of the cone using their non-dominant index finger to drag the slider positioned on their dominant hand. The cone becomes a ray when adjusting the cone to its minimal size (the top of the slider or the closest to the finger). The selection process is discrete; that is, the user needs to release the trigger and do it again for another selection/deselection.
- **Crossing Selection (CR)**: Crossing selection is a continuous selection process. The user holds the group selection trigger, moves the ray to intersect the target for selection, and releases the trigger for confirmation. Deselection happens when the ray intersects a selected object and is allowed within one selection action. In this work, we leveraged the collision of the ray and the object as the criterion of selection. We also enable the user to adjust the size of the ray in the same way as described in CC. When the user increases the size, the ray looks like a cylinder and becomes easier to collide with the object.
- Rectangle Selection (discarded): The user holds the group selection trigger and formulates a rectangular region similar to the rectangle selection in a desktop interface. The rectangle is instantiated and remains in the X-Y plane where the user starts to draw the rectangle. The user's hand movement in the Z-axis during the group selection is projected to that X-Y plane. When the user releases the trigger, a perspective projection is cast from the user's head position (the origin) to the rectangle and projects to farther planes. All the projected objects are selected/deselected. However, during the pilot testing, we found this 3D viewing process may confuse the users. Furthermore, it was not easy to integrate serial selection and to track accurately using this technique (sometimes the hand moves out of the headset's accurate tracking area). Thus, we discarded this technique from the evaluation.
- Volume Selection (discarded): With Volume Selection, the user points to an object and sends out a pre-defined 3D spherical volume with the pointed object as the center. All the objects within this volume would be selected/deselected. Like CC and CR, the user slides on the dominant hand to adjust the volume's size. We also visualize a replica of the volume (visually equal-sized) above the user's hand to assist her in estimating the target area. However, as the user can only decide the center and define the volume's size by expanding or shrinking it from its center, it is hard to have targets in the volume from all directions accurately. Due to this, Volume Selection was significantly more inefficient than others in our pilot test and was not included in our formal comparisons.

3.4 Summary

The refined interaction process and selected techniques are visualized in Figure 1. We also illustrate the initial design of the interaction process and techniques in the appendix for reference (Figure 5). We tried to optimize the parameters within each technique through informal testing. Although enhancing them by adding exclusive features for each specific technique is feasible, we focused more on experimental analysis of these techniques with shared features for controlled comparisons.

Finally, six combinations of mode-switching and group selection techniques have been selected for evaluation. We aimed to investigate user performance and experience of these potential techniques in multi-object selection scenarios in VR. To understand how potential techniques could be incorporated with default selection to provide a smooth workflow, we varied two independent variables: MSTech and GSTech. MSTech represents mode-switching techniques (FS, MF, and WO) while GSTech represents group selection techniques (CC and CR). In the following sections, we call their combinations FS+CC, FS+CR, MF+CC, MF+CR, WO+CC, and WO+CR.

4 User Study

In this user study, we compared the potential techniques in a controlled, simplified test environment with two task complexities (Section 4.3). We followed the guidelines outlined by Bergström et al. [2] to design and report this object selection study.

4.1 Participants

We recruited eighteen participants (5 women and 13 men) aged between 19 and 26 years ($M = 23$, $SD = 2.376$). All of them are right-handed. They have either normal vision ($N = 3$) or corrected-to-normal vision ($N = 15$). None had claimed they could not see the test environment clearly in the experiment. Sixteen participants reported they were familiar or very familiar with VR/AR/MR HMDs. Ten identified themselves as being familiar or very familiar with mid-air interaction, while two identified their unfamiliarity.

4.2 Apparatus

The study used a Meta Quest Pro VR HMD. Quest Pro has a 106° horizontal field-of-view, an 1800×1920 per eye resolution, and a 90Hz refresh rate. Its inside-out cameras enable 6 degrees of freedom hand tracking. It was connected to a high-performance desktop computer to run the experimental program. The computer was equipped with a Windows 11 system, an Intel Core i9-11900K processor, an NVIDIA GeForce RTX 3090 GPU, and 64GB of RAM. The program was implemented using C# in Unity (version 2022.3.0f1) with Oculus XR Plugin (version 4.0.0).

During the experiment, participants sat in front of a table to complete the experiment to minimize fatigue. The experimenter can observe participants' actions in the test environments (the Game view in the Unity interface) through the computer monitor. Figure 2(A) illustrates this setup.

4.3 Test Environment and Task

We used randomized scenarios as test environments to cover more general use cases while applying a few constraints to ensure the given task was controlled to meet our research goal. There were two types of spherical objects in the test environments: *targets* in yellow and *distractors* in grey. They have the same size with a radius of 0.1m. They were all located within a cuboid area of 1.4m×1.4m×0.5m, which was 2.5m in front of a participant's vision. Distractors were randomly positioned in this area. To simulate target groups for group selection, we defined four 0.4m×0.4m×0.5m areas within the outer rectangular area and let targets be generated randomly within two out of these four areas. Figure 2(B) demonstrates this task setting. Objects could not be manipulated

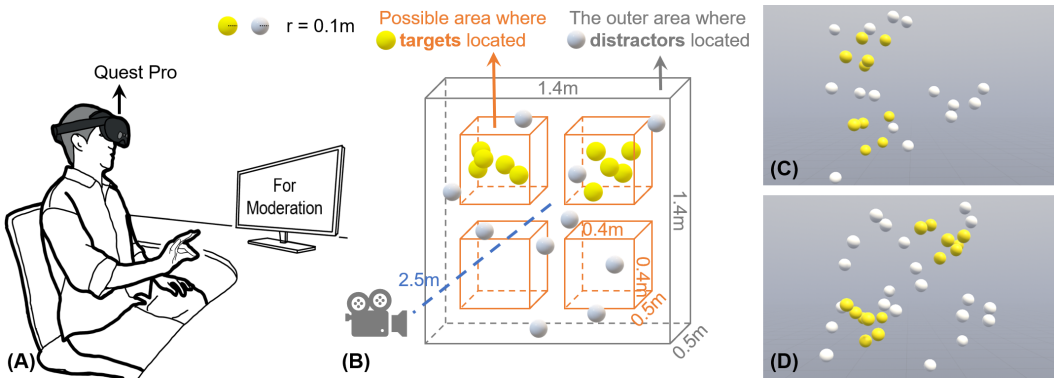

Fig. 2. Illustrations for (A) experimental setup, (B) experimental task, (C) the Low Complexity condition, (D) the High Complexity condition.

(like translation or rotation) and only had two states, either selected or not being selected. When an object had been selected, its outline would turn red. Participants could deselect an object by selecting it again, and the deselected object's outline would disappear.

The test environment involved two levels of task complexity. In the *Low Complexity* condition, 5 targets were randomly generated in each of the target areas (10 targets in total), and 10 distractors were randomly placed in the outer region, as shown in Figure 2(C). In the *High Complexity* condition, 8 targets were in each of the target areas (16 targets in total), and 16 distractors were in the outer region, as shown in Figure 2(D). That said, the task complexity was controlled by setting the number of objects in the test environment. The Low Complexity condition could be considered as testing in a sparse environment with fewer targets and distractors, and the High Complexity condition was a dense environment with more targets and distractors but kept the same ratio of targets to distractors (1:1).

Participants were asked to acquire all the targets and avoid selecting distractors as accurately and fast as possible. We explicitly mentioned to them that the priority was accuracy over speed. Typically, participants first performed a group selection and then refined their selection (deselect the distractors or select the missed targets).

4.4 Experimental Design

We used a 3×2 within-subjects design with MSTech (FS, MF, and WO) and GSTech (CC and CR) as two independent variables, as described in Section 3.4. To minimize the carry-over effect, the order of MSTech $\times$ GSTech conditions was counterbalanced via a balanced Latin Square approach [5]. Within each condition, participants completed ten randomized formal trials, five for each task complexity. Thus, we collected 18 participants $\times$ 3 mode-switching techniques $\times$ 2 group selection techniques $\times$ 2 task complexities $\times$ 5 repetitions = 1080 data trials in total.

4.5 Procedure

Participants first completed a demographic questionnaire and were introduced to the study purpose, design, VR device, and tasks. They were also briefed about the techniques and their controls. Participants then went through the conditions following the given order. Each condition could be divided into three phases. First, participants received a training session to familiarize themselves with the technique. In the training session, they were asked to get used to the given technique in the same task setting as described above. The experimenter explained the technique to participants

and then guided them to try all possible controls, including default selection, group selection, deselection, and size adjustment of the ray/cone. The training session included five trials and lasted at least one minute. Second, they completed the ten formal trials. The formal trials were given in a discrete form, where participants needed to click on the button above the object area via a default selection to continue with the next trial. Participants were informed explicitly to complete the ten formal trials in a condition carefully and continuously without rest. Third, they completed questionnaires about their feelings using the given technique (more details in the following section). A short break was given between two conditions. Once participants completed all conditions, they received a semi-structured interview to collect their feedback. The experiment lasted approximately 50 minutes for each participant.

4.6 Evaluation Metrics

We have a set of dependent variables involving both objective and subjective measurements.

4.6.1 Objective Measurements. For the objective measurements listed below, we used the average results per condition and participant for statistical analysis.

- *Completion Time*: We recorded the time (in seconds) taken to complete each trial.
- *Number of Errors*: We analyzed the number of missed targets, selected distractors, and total errors (the sum of the prior two).
- *Number of Actions*: The number of actions performed to complete the task. The actions counted included default selection, group selection, and ray/cone adjustment. We counted these actions once they were triggered (i.e., as a discrete action), regardless of how long they have been maintained.
- *Hand Movements*: The total hand movements in meters performed in each trial. It was calculated by aggregating the hand movement distance made in each frame.

4.6.2 Subjective Measurements. We also compared the techniques based on subjective measurements, including perceived workload, usability, arm fatigue, and preference rankings.

- *NASA-Task Load Index (NASA-TLX)* [11]: A raw NASA-TLX questionnaire was used to measure subjective workload when using the proposed techniques to complete the given task in terms of six dimensions: *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort*, and *frustration*. Further, these six scales derived a weighted *overall score*.
- *System Usability Scale (SUS)* [16]: We used a positive version of the SUS questionnaire to measure the usability of the proposed techniques. The ratings from 10 items were converted to an overall *SUS score* for statistical analysis.
- *Borg CR10 Scale* [3]: *Borg CR10 scale* is a categorical rating with scores ranging from 0 to 10 and corresponding verbal descriptions for assessing perceived arm exertion/fatigue.
- *Ranking*: At the end of the experiment, participants were asked to rank all six techniques according to their overall preference.

Except for the questionnaire for overall preference ranking, all other measurements were collected after participants completed a condition. The performance of the technique in both complexities was taken into consideration. We also interviewed the participants at the end of the experiment. We asked participants to reflect on their experience and share their opinions about the strengths and weaknesses or any other comments about the techniques.

4.7 Hypotheses

Based on our design process and pilot trials, we formulated the following two hypotheses that we were particularly interested in testing in the study:

- **H1.** Regarding the group selection techniques, CR will outperform CC and will provide a better experience because it involves a continuous selection mechanism, reducing the effort to trigger repeatedly the selection for multiple objects in CC.
- **H2.** The mode-switching gestures (FS, MF, and WO) will not lead to significantly varying performances and preferences because they are small and easy-to-perform gestures modified from the thumb-to-index pinch gesture.

5 Results

5.1 Objective Results

We removed trials in which the completion time was over three standard deviations from the mean ($> M + 3SD$) in each condition (15 trials, or 1.39% of total trials), the number of default selections was more than four times (9 trials, 0.83%), or the participants skipped by mistake (1 trial, 0.09%). In total, we removed 25 trials (2.31%). These trials were treated as outliers and removed because they implied unusual completion of trials (e.g., due to the hand tracking issue). We checked the normality of the data using both Shapiro-Wilk tests and QQ plots. The completion time and number of actions were normally distributed. We then performed repeated-measure (RM-) ANOVA to analyze the effects of the variables and conducted pairwise comparisons with Bonferroni adjustment to the p values. The number of errors (including the number of selected distractors, missed targets, and total errors) and hand movements were not normally distributed. We transformed them via aligned rank transform (ART) [45] before conducting RM-ANOVA tests and applied the ART-C procedure [9] for post-hoc analysis (p values are also Bonferroni-adjusted).

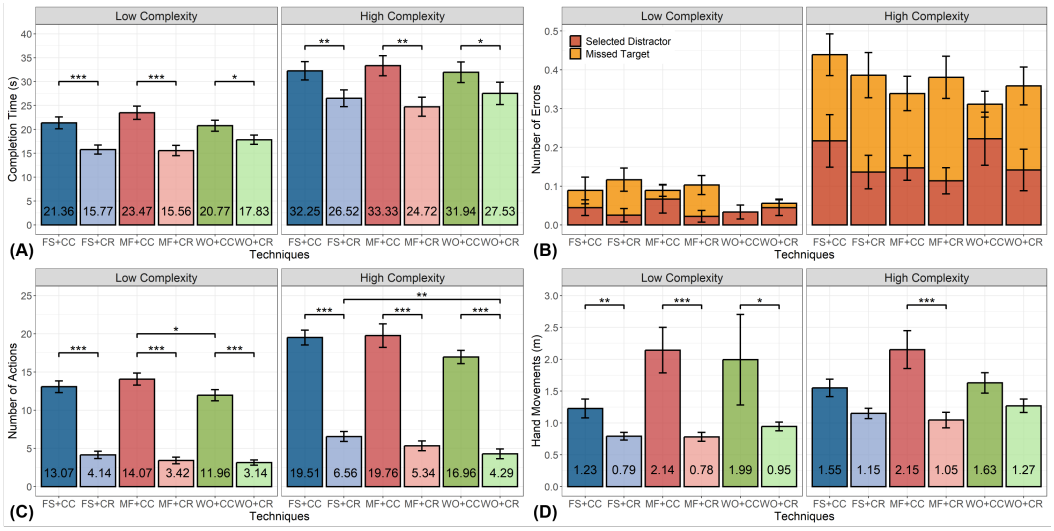

Fig. 3. Plots of average (±1SE) performance under two task complexities. (A) Completion time. (B) Number of errors. (C) Number of actions. (D) Hand movements.

5.1.1 Completion Time. In the Low Complexity condition, there was a significant effect of GSTech ($F_{1,17} = 40.394, p < 0.001, \eta_p^2 = 0.253$) and a significant interaction effect between MSTech and GSTech ($F_{2,34} = 4.528, p = 0.018, \eta_p^2 = 0.045$) on the mean completion time. MSTech did not have a significant effect ($F_{2,34} = 0.671, p = 0.518, \eta_p^2 = 0.007$). Pairwise comparisons showed that the mean

completion time was significantly longer with CC than CR when the mode-switching technique was FS ($p < 0.001$), MF ($p < 0.001$), and WO ($p = 0.036$).

In the High Complexity condition, there was a significant effect of GSTech ($F_{1,17} = 20.105, p < 0.001, \eta_p^2 = 0.120$) on mean completion time, while the effect of MSTech ($F_{2,34} = 0.102, p = 0.903, \eta_p^2 = 0.001$) and the interaction effect ($F_{2,34} = 1.728, p = 0.193, \eta_p^2 = 0.011$) were not significant. Similar to the Low Complexity condition, pairwise comparisons showed that CC had a longer completion time than CR when the mode-switching technique was FS ($p = 0.003$), MF ($p = 0.003$), and WO ($p = 0.011$). The results are illustrated in Figure 3(A).

5.1.2 Number of Errors. RM-ANOVA showed that the number of missed targets in the Low Complexity condition was significantly influenced by MSTech ($F_{2,85} = 6.424, p = 0.003, \eta_p^2 = 0.131$) and GSTech ($F_{1,85} = 5.736, p = 0.019, \eta_p^2 = 0.063$). On the other hand, there was a significant effect of GSTech ($F_{1,85} = 5.888, p = 0.017, \eta_p^2 = 0.065$) on the number of missed targets in the High Complexity condition. Except for this, RM-ANOVA did not indicate any other other significant effects. Pairwise comparisons did not show any significant differences among the techniques either. Figure 3(B) visualizes the results. As can be seen, the number of errors was very low (less than one time), regardless of the techniques.

5.1.3 Number of Actions. In the Low Complexity condition, there was a significant effect of MSTech ($F_{2,34} = 3.586, p = 0.039, \eta_p^2 = 0.042$) and a significant effect of GSTech ($F_{1,17} = 229.268, p < 0.001, \eta_p^2 = 0.777$) on the number of actions. The interaction effect between MSTech and GSTech on the number of actions was not significant ($F_{2,34} = 2.221, p = 0.124, \eta_p^2 = 0.027$). Pairwise comparisons showed that CC required significantly more actions to complete the task than CR when the mode-switching technique was FS, MF, and WO (all $p < 0.001$). Additionally, within the CC group selection technique, MF required significantly more actions than WO ($p = 0.022$).

There was a significant effect of MSTech ($F_{2,34} = 5.155, p = 0.011, \eta_p^2 = 0.067$) and a significant effect of GSTech ($F_{1,17} = 305.758, p < 0.001, \eta_p^2 = 0.748$) on the number of actions in the High Complexity condition. The interaction effect between MSTech and GSTech was not significant ($F_{2,34} = 0.795, p = 0.460, \eta_p^2 = 0.010$). Same as in the Low Complexity condition, FS+CC, MF+CC, and WO+CC involved a significantly higher number of actions than FS+CR, MF+CR, and WO+CR, respectively (all $p < 0.001$). Furthermore, we also found that FS+CR required significantly more actions than WO+CR ($p = 0.003$). Figure 3(C) summarizes the results regarding the number of actions.

5.1.4 Hand Movements. Figure 3(D) shows the results regarding the hand movements. RM-ANOVA showed that the hand movements in the Low Complexity condition were significantly influenced by MSTech ($F_{2,85} = 4.084, p = 0.020, \eta_p^2 = 0.088$), GSTech ($F_{1,85} = 43.879, p < 0.001, \eta_p^2 = 0.340$), and their interaction ($F_{2,85} = 5.101, p = 0.008, \eta_p^2 = 0.107$). Results from pairwise comparisons showed that FS+CC, MF+CC, and WO+CC involved a significantly more hand movements than FS+CR ($p = 0.004$), MF+CR ($p < 0.001$), and WO+CR ($p = 0.024$), respectively.

In the High Complexity condition, there was a significant effect of GSTech ($F_{1,85} = 44.501, p < 0.001, \eta_p^2 = 0.344$) and a significant interaction effect ($F_{2,85} = 4.519, p = 0.014, \eta_p^2 = 0.096$) on the hand movements, while there was no significant effect of MSTech ($F_{2,85} = 0.626, p = 0.537, \eta_p^2 = 0.015$). Pairwise comparisons only showed a significant difference between MF+CC and MF+CR ($p < 0.001$), with MF+CC having significantly more hand movements.

5.2 Subjective Results

We performed RM-ANOVA and pairwise comparison with Bonferroni adjustments to ART- transformed [9, 45] questionnaire results, including NASA-TLX scores, SUS scores, and Borg CR 10 scores. The descriptive analyses of these measurements are visualized in Figure 4(A-C).

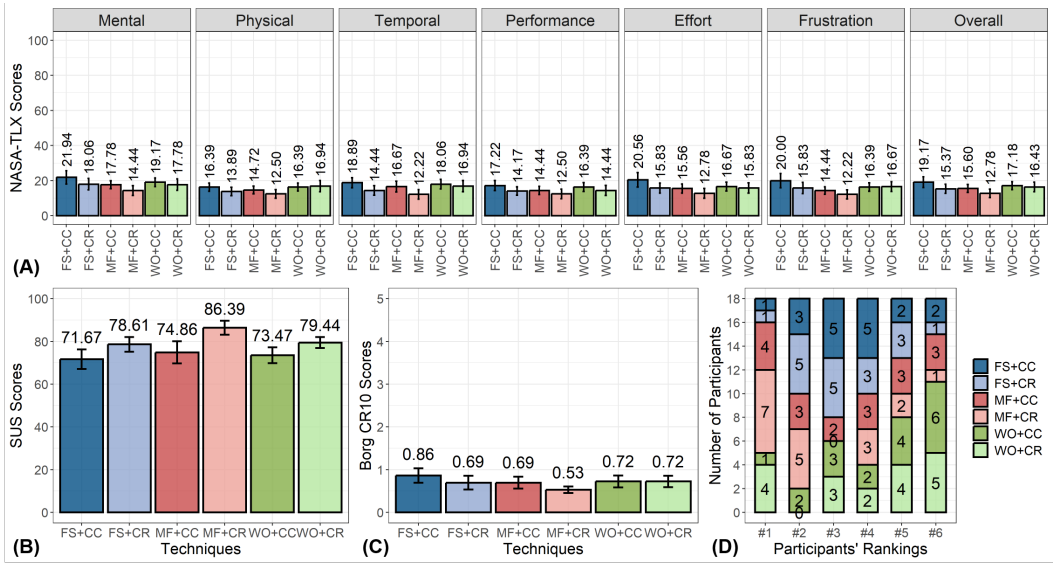

Fig. 4. (A) Average (±1*SE*) NASA scores of the techniques. The lower the score is, the lower the perceived workload of the technique (i.e., the better). (B) Average (±1*SE*) SUS scores of the techniques. The higher the score is, the higher the usability of the technique (i.e., the better). (C) Average (±1*SE*) Borg CR10 scores of the techniques. The lower the score is, the lower the perceived arm fatigue (i.e., the better). (D) Participants' ranking of each technique.

Regarding perceived workload, RM-ANOVA revealed significant effects of GSTech on mental demand ($F_{1,85} = 4.485, p = 0.037, \eta_p^2 = 0.050$), temporal demand ($F_{1,85} = 5.161, p = 0.026, \eta_p^2 = 0.057$), performance ($F_{1,85} = 4.058, p = 0.047, \eta_p^2 = 0.046$), and overall score ($F_{1,85} = 4.082, p = 0.046, \eta_p^2 = 0.046$). No other significant effect, interaction effect, or post hoc differences were found. As shown in Figure 4(A), the workload of using each technique to complete the selection task is perceived to be low.

Both MSTech ($F_{2,85} = 5.167, p = 0.008, \eta_p^2 = 0.108$) and GSTech ($F_{1,85} = 12.640, p < 0.001, \eta_p^2 = 0.129$) had a significant effect on SUS scores. However, we did not find a significant interaction effect between MSTech and GSTech ($F_{2,85} = 0.846, p = 0.433, \eta_p^2 = 0.020$) on SUS scores. Moreover, results from pairwise comparisons showed no significant differences among the techniques. Overall, the proposed techniques were rated with high usability. The average SUS scores for each technique are over 70 points (see Figure 4(B)).

In terms of Borg CR 10 scores, no significant differences were found. All the techniques were rated with low arm fatigue to complete the multi-object selection tasks, as Figure 4(C) shows.

Figure 4(D) shows participants' ranking of the six techniques. 7 participants (38.89%) ranked MF+CR the most favored technique, followed by MF+CC ($N = 4$, 22.22%) and WO+CR ($N = 4$, 22.22%). In terms of MSTech (FS vs. MF vs. WO), the figure shows a clear tendency to favor MF-based techniques. Twelve participants (66.67%) ranked them as the most favored technique. In

contrast, WO-based techniques were the least favored because 11 participants (61.11%) ranked them the last. On the other hand, if we group the ranking according to GSTech (CC vs. CR), participants' preferences were scattered. CR-based techniques were ranked first place by 12 participants (66.67%) and second place by 10 participants (55.56%). They were also ranked fifth place by 9 participants (50%), and sixth place by 7 participants (38.89%). We present and discuss the interview responses in the Discussion section.

6 Discussion

Our results demonstrate how the proposed mode-switching techniques and group selection techniques fit into a freehand multi-object selection workflow in randomized VR scenarios. In this section, we discuss the results and evaluation of the techniques and provide design implications that can help the design and development of object selection techniques in VR in the future.

6.1 Technique Evaluation

H1 was about the performance and experience of group selection techniques. It was partially supported—the results supported our assumption regarding group selection techniques' performance but contradicted our expectations of their user experience. Many significant performance differences were observed when comparing the two group selection techniques and they showed consistent patterns. CR was faster than CC in completing the multi-object selection tasks in randomized scenarios, regardless of the incorporated mode-switching gesture or the task complexity. In addition, it required fewer actions than CC to complete the task. This is an expected result because CR involves a continuous selection procedure so that users can refine their selection within one selection event. Furthermore, we found participants rarely resized the ray when using CR during the experiments, which also reduced the number of actions. Although CR led to a better performance, some participants raised negative comments on it. They described CR as "difficult to control" (P1, P6) or "unstable" (P4), especially when the target was surrounded by distractors. P1, P4, and P13 reported that they felt CR was more prone to select an unwanted object, while P13 also mentioned such an error was easy to fix. This might be the reason for participants not to increase the ray size. Though CC required more time and more actions to complete the task, it was preferred by some participants. The primary driver was its selection mechanism—it was consistent with the default pointing selection. P4 and P9 mentioned that they did not need to hold the pinch gesture, which was more relaxed compared to using CR. We also found that CC needed more hand movements in the Low Complexity conditions, which is out of our expectations. We speculate that participants barely adjusted the ray's size when using CR but had to do so with CC. As the ray-casting-based technique, either by crossing or pointing, did not require a large interaction space, the adjustment action involved a greater range of motion by contrast.

In summary, although CR outperformed CC in the speed, number of actions, and hand movements, CR and CC both led to a good user experience and preferred by participants to be integrated into the freehand multi-object selection workflow. The ray adjustment feature may not needed for CR in randomized scenarios as it behaves sensibly for some users. On the other hand, to further improve the CC, four participants (P4, P6, P11, P18) suggested giving visual cues when the objects are illuminated before the selection, such as showing them with an outline of another color. P11 also suggested allowing users to customize the direction of the slider.

H2 regarding the performances and preferences in mode-switching gestures was also partially supported. We did not observe much significant performance differences when comparing the mode-switching gestures. These gestures only have minimal modifications from the standard pinch gestures, and did not impact the performance much. However, the participants' ranking and their feedback revealed that they have different preferences toward these gestures (which

contradicted to **H2**). Overall, MF is the most favored mode-switching gesture and was acceptable to most participants (see Figure 4(D)). On the contrary, most criticism was given to WO. P10 said, "*when using WO, I pay extra attention to how much has been or still needs to rotate.*" After switching to the group selection mode, maintaining the gesture rotated was also more challenging (P6, P9, P10). P6 mentioned he felt controlling the ray/cone with palm facing up or close to up (WO) was more difficult and unnatural than with palm facing down (in a standard pinch, FS, or MF). As for FS, P6 and P17 thought this gesture was not natural and not commonly used even in daily life. On the other hand, P1 felt that it was too close to the fingertip pinch (from the perspective of distance and interaction), taking her some time to distinguish between them and remember.

In this user study, we used randomized testing environments with several constraints to compare the proposed techniques. Based on the results and our observations, completing the High Complexity tasks was clearly tougher than the Low Complexity tasks during the experiments. P1 and P15 expressed their concerns: "*I think I cannot select the targets accurately anymore if more objects crowded there.*" When the VR environment becomes more complex, such as having small or occluded objects, a disambiguation technique may be necessary to resolve the ambiguity. Future studies may explore how to insert a suitable disambiguation technique into the workflow to acquire multiple targets with more ease and precision. On the other hand, we forced the targets to be generated in two clusters to investigate the group selection technique. In the actual applications, the targets may be placed in a structured layout, which should make the group selection techniques more effective. Considering the participants' workload, we could not test the techniques in these scenarios in this study, but we mark the importance of this evaluation for future studies.

6.2 Design Implications

Based on the study results, we distilled the following three design implications.

- Both Crossing Selection (CR) and Cone-casting Selection (CC) are suitable for group-based selection. If the selection performance is critical to the application scenario, CR is superior to CC.
- Using the middle finger (MF) for switching between the single-object selection and multi-object selection modes is suggested.
- Rotating the wrist (WO) for switching the selection mode is not preferred by users and thus not suggested.

7 Limitations and Future Work

There are three limitations to our work. First and foremost, we only compared the techniques in randomized scenarios due to the size of the study. We chose to start with a random arrangement of multiple objects because the derived results and findings can be relatively more universal and generalizable. It is worth evaluating the techniques in more scenarios, including randomized scenarios with varying constraints and different types of structured layouts (see also Section 6), and collecting more types of user performance for deriving user behavior patterns (e.g., the most frequently used cone's size in Cone-Casting selection). Further, we also want to investigate their use in practical applications involving multi-object selection, such as data visualization, 3D modeling, and building design applications. Second, our participants were all right-handed. For future studies, we would like to investigate the performance of left-handed users. We also want to invite more participants with more diverse backgrounds (in terms of gender, age group, past experience, and degrees of motor challenges) to collect their feedback and invite them to use the techniques in the long term to investigate their prolonged use. By doing so, we could also provide further insights into the individual and group differences in the performance and user experiences of the techniques.

Third, we focused on the interaction techniques that were applicable to current available HMDs, and due to this, we used their built-in hand-tracking modules. We acknowledged that the precision and stability of hand-tracking may affect the results. In the future, we want to compare the proposed techniques with more micro-interactions that are not usable now but become feasible with the advancements in tracking technologies.

8 Conclusion

In this work, we present an analysis of freehand multi-object selection techniques in virtual reality (VR). Specifically, we investigate how different group selection techniques and mode-switching gestures impact the performance and user experience in multi-object selection tasks. With the results from a user study with eighteen participants, we found crossing selection to be fast and required fewer actions and hand movements compared to cone-casting selection, while both techniques led to a good user experience and gained acceptance by participants. For transitioning between the default single-object selection mode and multi-object selection mode, the three proposed mode-switching gestures showed comparable performance. Based on the participants' feedback, using the middle finger pinch gesture was recommended while oriented the pinch gesture was not. We hope these results and findings can be useful to the practitioners of the VR community in designing and developing more usable gestural interactions.

Acknowledgments

The authors thank the participants for their time and the reviewers for the helpful comments and suggestions. This work was funded in part by the Suzhou Municipal Key Laboratory for Intelligent Virtual Engineering (#SZS2022004) and the National Natural Science Foundation of China (#62272396).

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] Joanna Bergström, Tor-Salve Dalsgaard, Jason Alexander, and Kasper Hornbæk. 2021. 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 533, 20 pages. https://doi.org/10.1145/3411764.3445193

[3] Gunnar A Borg. 1982. Psychophysical bases of perceived exertion. *Medicine and science in sports and exercise* 14, 5 (1982), 377–381.

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

[5] James V. Bradley. 1958. Complete Counterbalancing of Immediate Sequential Effects in a Latin Square Design. *J. Amer. Statist. Assoc.* 53, 282 (1958), 525–528. https://doi.org/10.1080/01621459.1958.10501456 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/01621459.1958.10501456

[6] Gavin Buckingham. 2021. Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. *Frontiers in Virtual Reality* 2 (2021). https://doi.org/10.3389/frvir.2021.728461

[7] Di Laura Chen, Ravin Balakrishnan, and Tovi Grossman. 2020. Disambiguation Techniques for Freehand Object Manipulations in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 285–292. https://doi.org/10.1109/VR46266.2020.00048

[8] Tor-Salve Dalsgaard, Jarrod Knibbe, and Joanna Bergström. 2021. Modeling Pointing for 3D Target Selection in VR. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (Osaka, Japan) *(VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 42, 10 pages. https://doi.org/10.1145/3489849.3489853

[9] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '21)*. Association for Computing Machinery, New York, NY, USA, 754–768. https://doi.org/10.1145/3472749.3474784

[10] 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

[11] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[12] Ken Hinckley, Francois Guimbretiere, Patrick Baudisch, Raman Sarin, Maneesh Agrawala, and Ed Cutrell. 2006. The Springboard: Multiple Modes in One Spring-Loaded Control. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 181–190. https://doi.org/10.1145/1124772.1124801

[13] Susu HUANG, Daqing QI, Jiabin YUAN, and Huawei TU. 2019. Review of studies on target acquisition in virtual reality based on the crossing paradigm. *Virtual Reality & Intelligent Hardware* 1, 3 (2019), 251–264. https://doi.org/10.3724/SP.J.2096-5796.2019.0006 Human-computer interactions for virtual reality.

[14] Akira Ishii, Takuya Adachi, Keigo Shima, Shuta Nakamae, Buntarou Shizuki, and Shin Takahashi. 2017. FistPointer: Target Selection Technique using Mid-air Interaction for Mobile VR Environment. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 474. https://doi.org/10.1145/3027063.3049795

[15] 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, Boston, MA, USA.

[16] James R. Lewis. 2018. The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction* 34, 7 (2018), 577–590. https://doi.org/10.1080/10447318.2018.1455307 arXiv:https://doi.org/10.1080/10447318.2018.1455307

[17] John Finley Lucas. 2005. *Design and evaluation of 3D multiple object selection techniques*. Ph. D. Dissertation. Virginia Tech.

[18] Mykola Maslych, Yahya Hmaiti, Ryan Ghamandi, Paige Leber, Ravi Kiran Kattoju, Jacob Belga, and Joseph J. LaViola. 2023. Toward Intuitive Acquisition of Occluded VR Objects Through an Interactive Disocclusion Mini-map. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 460–470. https://doi.org/10.1109/VR55154.2023.00061

[19] Sven Mayer, Valentin Schwind, Robin Schweigert, and Niels Henze. 2018. The Effect of Offset Correction and Cursor on Mid-Air Pointing in Real and Virtual Environments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Montreal QC</city>, <country>Canada</country>, </conf-loc>) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174227

[20] Daniel Mendes, Daniel Medeiros, Maurício Sousa, Eduardo Cordeiro, Alfredo Ferreira, and Joaquim A. Jorge. 2017. Design and evaluation of a novel out-of-reach selection technique for VR using iterative refinement. *Computers & Graphics* 67 (2017), 95–102. https://doi.org/10.1016/j.cag.2017.06.003

[21] Mark R Mine. 1995. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept* (1995).

[22] Mark R. Mine, Frederick P. Brooks, and Carlo H. Sequin. 1997. 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)*. ACM Press/Addison-Wesley Publishing Co., USA, 19–26. https://doi.org/10.1145/258734.258747

[23] Donald A. Norman. 2010. Natural User Interfaces Are Not Natural. *Interactions* 17, 3 (may 2010), 6–10. https://doi.org/10.1145/1744161.1744163

[24] Chanho Park, Hyunwoo Cho, Sangheon Park, Young-Suk Yoon, and Sung-Uk Jung. 2019. HandPoseMenu: Hand Posture-Based Virtual Menus for Changing Interaction Mode in 3D Space. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces* (Daejeon, Republic of Korea) *(ISS '19)*. Association for Computing Machinery, New York, NY, USA, 361–366. https://doi.org/10.1145/3343055.3360752

[25] 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

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

[27] Henning Pohl, Klemen Lilija, Jess McIntosh, and Kasper Hornbæk. 2021. Poros: Configurable Proxies for Distant Interactions in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 532, 12 pages. https://doi.org/10.1145/3411764.3445685

[28] 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

[29] Jef Raskin. 2000. *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional.

[30] Gang Ren and Eamonn O'Neill. 2013. 3D selection with freehand gesture. *Computers & Graphics* 37, 3 (2013), 101–120. https://doi.org/10.1016/j.cag.2012.12.006

[31] Abigail J. Sellen, Gordon P. Kurtenbach, and William A.S. Buxton. 1992. The Prevention of Mode Errors Through Sensory Feedback. *Human–Computer Interaction* 7, 2 (1992), 141–164. https://doi.org/10.1207/s15327051hci0702_1

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

[33] Rongkai Shi, Jialin Zhang, Wolfgang Stuerzlinger, and Hai-Ning Liang. 2022. Group-Based Object Alignment in Virtual Reality Environments. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction* (Online, CA, USA) *(SUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 2, 11 pages. https://doi.org/10.1145/3565970.3567682

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

[35] Rongkai Shi, Nan Zhu, Hai-Ning Liang, and Shengdong Zhao. 2021. Exploring Head-based Mode-Switching in Virtual Reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 118–127. https://doi.org/10.1109/ISMAR52148.2021.00026

[36] Jesse Smith, Isaac Wang, Winston Wei, Julia Woodward, and Jaime Ruiz. 2020. Evaluating the Scalability of Non-Preferred Hand Mode Switching in Augmented Reality. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) *(AVI '20)*. Association for Computing Machinery, New York, NY, USA, Article 19, 9 pages. https://doi.org/10.1145/3399715.3399850

[37] Jesse Smith, Isaac Wang, Julia Woodward, and Jaime Ruiz. 2019. Experimental Analysis of Single Mode Switching Techniques in Augmented Reality. In *Proceedings of the 45th Graphics Interface Conference on Proceedings of Graphics Interface 2019* (Kingston, Canada) *(GI'19)*. Canadian Human-Computer Communications Society, Waterloo, CAN, Article 20, 8 pages. https://doi.org/10.20380/GI2019.20

[38] Zhaomou Song, John J. Dudley, and Per Ola Kristensson. 2022. Efficient Special Character Entry on a Virtual Keyboard by Hand Gesture-Based Mode Switching. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 864–871. https://doi.org/10.1109/ISMAR55827.2022.00105

[39] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental Analysis of Barehand Mid-Air Mode-Switching Techniques in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300426

[40] Huawei Tu, Susu Huang, Jiabin Yuan, Xiangshi Ren, and Feng Tian. 2019. Crossing-Based Selection with Virtual Reality Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300848

[41] Stephen Uzor and Per Ola Kristensson. 2021. An Exploration of Freehand Crossing Selection in Head-Mounted Augmented Reality. *ACM Trans. Comput.-Hum. Interact.* 28, 5, Article 33 (aug 2021), 27 pages. https://doi.org/10.1145/3462546

[42] Tingjie Wan, Rongkai Shi, Wenge Xu, Yue Li, Katie Atkinson, Lingyun Yu, and Hai-Ning Liang. 2024. Hands-free multi-type character text entry in virtual reality. *Virtual Reality* 28, 8 (2024), 1–19. https://doi.org/10.1007/s10055-023-00902-z

[43] Tingjie Wan, Yushi Wei, Rongkai Shi, Junxiao Shen, Per Ola Kristensson, Katie Atkinson, and Hai-Ning Liang. 2024. Design and Evaluation of Controller-Based Raycasting Methods for Efficient Alphanumeric and Special Character Entry in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 30, 9 (2024), 6493–6506. https://doi.org/10.1109/TVCG.2024.3349428

[44] 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

[45] 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

[46] Zhiqing Wu, Difeng Yu, and Jorge Goncalves. 2023. Point- and Volume-Based Multi-object Acquisition in VR. In *Human-Computer Interaction – INTERACT 2023*, José Abdelnour Nocera, Marta Kristín Lárusdóttir, Helen Petrie,

Antonio Piccinno, and Marco Winckler (Eds.). Springer Nature Switzerland, Cham, 20–42.
[47] Haijun Xia, Sebastian Herscher, Ken Perlin, and Daniel Wigdor. 2018. Spacetime: Enabling Fluid Individual and Collaborative Editing in Virtual Reality. 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, 853–866. https://doi.org/10.1145/3242587.3242597
[48] Difeng Yu, Qiushi Zhou, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2022. Blending On-Body and Mid-Air Interaction in Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 637–646. https://doi.org/10.1109/ISMAR55827.2022.00081
[49] 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
[50] Lixiang Zhao, Tobias Isenberg, Fuqi Xie, Hai-Ning Liang, and Lingyun Yu. 2024. MeTACAST: Target- and Context-Aware Spatial Selection in VR. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 480–494. https://doi.org/10.1109/TVCG.2023.3326517

A Initial Design of Interaction Process and Techniques

Figure 5 illustrates the initial design of the interaction process and techniques for freehand multi-object selection in VR HMDs. In this version, the serial selection is a separate selection mode with a unique activation gesture, and Rectangle Selection and Volume Selection are included. After pilot tests, we optimized the whole process. More has been discussed in Section 3.

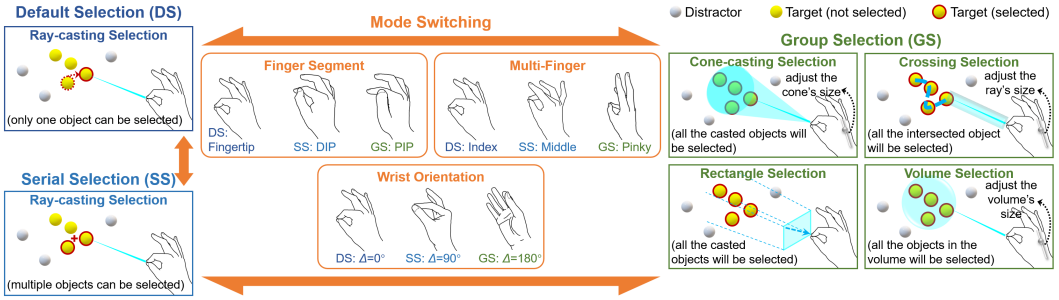

Fig. 5. The initial design of the multiple-object selection process and freehand techniques.

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