2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR) | 979-8-3315-3645-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/VR59515.2025.00071

May The Force be With You: Cloning Distant Objects to Improve Medium-Field Interactions in Augmented Reality

Danish Nisar Ahmed Tamboli University of Florida
*Rohith Venkatakrishnan† University of Florida
Roshan Venkatakrishnan‡ University of Florida
Balagopal Raveendranath§ Texas Tech University
Julia Woodward¶ University of South Florida
Isaac Wang ‖ James Madison University
Jesse Smith** University of Florida
Jaime Ruiz†† University of Florida

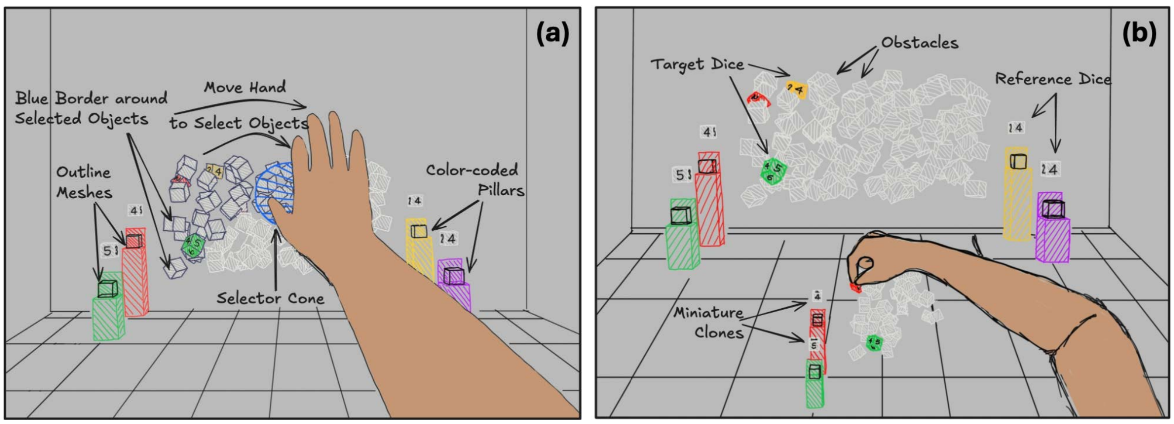


Figure 1: Sketches depicting The Force (section 3.4) being used to: (a) select objects to be cloned using the open palm gesture and (b) manipulate the cloned miniatures directly. These sketches illustrate the specifics of the task described in section 4.1.

Abstract

Augmented Reality (AR) interactions feature users interacting with virtual objects registered in the physical world. With contemporary AR experiences increasingly featuring interactions at distances, we conceptualized *The Force*, a technique that allows users to clone distant objects and manipulate their replicas. An empirical evaluation was conducted, comparing it against two well-established techniques including controller-based ray-casting and a gaze-based pinching technique in a pick-and-place task. We employed a within-subjects design, collecting data on both objective performance and subjective user experience. Results suggest that *The Force* allows for higher levels of accuracy and efficiency in medium-field tasks that require precision and fine motor control. Furthermore, we discovered avenues towards iteratively refining this technique. We go on to discuss the implications of our findings in an effort to facilitate better interactions in augmented reality.

Index Terms: Augmented Reality, Interaction Techniques, 3D User Interaction, Distant Object Manipulation, Cloning

*e-mail: danishtamboli@ufl.edu
†e-mail: rohith.venkatakr@ufl.edu
‡e-mail: rvenkatakrishnan@ufl.edu
§e-mail: braveend@ttu.edu
¶e-mail: juliaevewoodward@usf.edu
‖e-mail: wangid@jmu.edu
**e-mail: jd.smith@ufl.edu
††e-mail: jaime.ruiz@ufl.edu

1 Introduction

In recent years, Augmented and Virtual Reality (AR and VR) have gained prominence, evolving from passive experiences to interactive platforms used for gaming, training, entertainment, and productivity [38]. AR, which overlays digital elements onto the real world, offers unique advantages by situating such content in the physical environment, making it especially useful in scenarios requiring situational awareness [2]. A fundamental aspect of AR applications is user interaction with virtual elements integrated into the real world. These digital entities such as objects, menus, icons, interfaces, and virtual characters are registered precisely within the user's environment, enhancing immersion and interactivity [2, 33]. While many of these interactions occur in the near field (up to 1.5 meters) within the user's reach envelope, they can also occur in medium (1.5 - 30 meters) and far-field (beyond 30 meters) settings wherein users attempt to manipulate entities that are more distant but still visible [29, 62]. The manipulation of distant objects has been less explored, creating a gap in our understanding of how to better support interactions occurring in these ranges.

AR interactions can be facilitated through different methods, ranging from specialized hardware like controllers to more intuitive approaches like speech, eye tracking, and hand gestures [69]. In modern AR systems, eye gaze and hand gestures are commonly used to select and manipulate objects, with users directing their gaze to an object and confirming their selection via hand movements. While effective, users often prefer direct hand-based manipulation due to its intuitive feel and familiarity [31]. However, as objects move beyond a user's immediate physical reach, traditional hand-based techniques face limitations, such as reduced depth perception and increased hand jitter, which can make precise manipulation difficult [3]. To address such challenges, we developed *The Force*, an interaction technique inspired by the previously proposed Worlds in

*contributed equally to this work.

2642-5254/25/$31.00 ©2025 IEEE
DOI 10.1109/VR59515.2025.00071

Miniature (WIM) [61]. This technique allows users to selectively create miniature clones of distant objects and manipulate their replicas instead, thus enabling near-field manipulation of distant objects. By mirroring interactions performed on proximate replicas of distant objects, *The Force* aims to offer an improved sense of spatial relationships among objects and a more familiar experience, allowing for improved accuracy and control, since prior research suggests that impoverished vision and impaired depth perception is associated with manipulating distant objects [15]. In contrast to the previously used WIM techniques, however, *The Force* affords users with control over the objects they want cloned rather than only allowing cloning of the entire environment as in [61], predefined object lists as in [31], or predefined hierarchies of selected objects as in the Voodoo Dolls technique conceptualized in [52]. The technique hence affords users more fine-grained control over the choice of distant objects they want to replicate and interact with.

Raycasting and Gaze-based pinching are established interaction techniques often used for medium-field interactions. Both of these techniques rely on users manipulating objects through abstract concepts such as a ray from the controller or a gaze vector, often creating a disconnect from the natural feel of direct manipulation. Additionally, manipulating objects using these techniques may magnify hand manipulation errors due to human instability, increasing the error of object placement with distance [26]. Gaze-based techniques, while intuitive, can impose higher levels of mental workload, requiring users to fixate and maintain constant focus on the objects they wish to interact with [48]. This becomes more challenging when there are limitations with gaze tracking, necessitating the integration of high-fidelity eye trackers that can be expensive. Furthermore, these types of techniques become less efficient in densely cluttered environments where users may want to interact with objects that are substantially occluded. In such crowded settings, traditional techniques like those aforementioned will involve users having to move irrelevant objects out of the way before being able to interact with relevant objects, reducing efficiency and productivity. These techniques may also present challenges when the distant interactions require high degrees of precision, control, and vision of the objects from different perspectives like remote assembly tasks, architectural design tasks, and collaborative industrial tasks. These challenges arise because visual acuity decreases with depth of focus [21]. *The Force* attempts to address some of these challenges by transforming distant object interaction tasks into near-field manipulation tasks that inherently afford better viewing and control [3]. It allows users to create and manipulate miniaturized versions of objects in close proximity, providing a more natural, intuitive and less cognitively burdening interaction experience through direct hand-based manipulation. Working with scaled down proximate replicas will enable users to easily shift perspective, efficiently work on relevant objects, and more effectively perform manipulations that require spatial awareness and fine motor control [31]. To study its efficacy for medium-field interactions, we sought to compare it with these two aforementioned techniques. Apropos of these pursuits, we conducted a comparative evaluation of these techniques in a medium-field pick-and-place task involving precise perception-action coordination. We go on to discuss results from this evaluation, further detailing avenues for improvement and refinement of *The Force*.

2 Related Works

2.1 Interaction in Augmented and Virtual Reality

Interaction in AR/VR environments has been explored across various contexts, some of which involve the manipulation of virtual objects. Researchers often utilize Fitts' Law to evaluate object manipulation, demonstrating its applicability in modeling user performance in interactions [8]. Virtual experiences often involve the user interacting with virtual entities like objects, menus, and icons that are spatially registered in 3D and overlaid onto the real-world environment [1, 59]. Interaction methods range from natural techniques, such as hand gestures and voice commands, to hardware-driven methods, including handheld controllers and other devices [28, 43].

Prior research has shown that natural hand gestures are particularly effective for near-field interactions, where users directly manipulate virtual objects within arm's reach [16]. However, as the interaction distance increases, the precision of gesture-based interactions tends to diminish due to increased hand jitter caused by extending the arm, making controllers and raycasting techniques more suitable for medium and far-field interactions. Studies also suggest that hybrid approaches combining multiple interaction modalities, such as combining gaze with gesture or controller inputs, can enhance user experience and performance in AR/VR settings [11, 30, 49].

Interaction techniques are often evaluated in the context of docking tasks which involve the precise placement of objects in a target location, requiring both orientation and positioning of the object. Docking tasks involve gross motion and then fine-tuning once objects are near the target location [67]. Froehlich et al. [18] introduced two novel 6 DOF input devices namely GlobeFish and GlobeMouse, and compared them with SpaceMouse in terms of docking a tetrahedral cursor with a target tetrahedral. They found that both techniques outperformed the SpaceMouse technique in terms of task-completion times, times needed for translation and rotation, and subjective ratings of ease of use. Similarly, a comparative evaluation of input devices in terms of performing a docking task revealed that a two-surface multitouch outperformed a single-surface multitouch device, a 2D mouse, and a 6 DOF free-space device (Phantom Omni) in terms of efficiency [20]. As such, docking tasks often prove to be reliable means to perform evaluations due to the involvement of both gross-motion and fine-motor control.

2.2 Near Field Interaction

Near-field interactions encompass those that occur within 1.5 meters from a user, roughly denoting their reach envelope [29, 62]. Techniques for these kinds of interactions are predominantly characterized by their reliance on the user's hand movements and the headset's ability to track it, often augmented by gloves or other wearable devices, to manipulate virtual objects in close proximity [39]. These techniques leverage the natural dexterity and precision of the human hand, making them well-suited for tasks that require fine motor control [23, 39]. For example, glove-based systems have been developed to facilitate precise gesture recognition and interaction, enabling users to engage with virtual environments in an intuitive and immersive manner [39, 66].

Virtual hand and tangible interaction techniques further extend the capabilities of near-field interactions by allowing users to manipulate 3D objects as if they were real, providing a more natural and engaging experience [22, 65]. The use of multi-finger and precision grips in these environments has also been shown to enhance user control and accuracy, particularly in tasks that require the precise positioning or manipulation of small objects [37, 64]. While these approaches offer significant benefits in terms of user experience and interaction fidelity, they also highlight the challenges of achieving consistent and accurate gesture recognition, particularly in complex or cluttered environments [42, 53]. Moreover, users can often find wearing gloves or other such external apparatus clunky and disconnecting from the experience [39].

2.3 Medium & Far Field Interaction

Medium and far-field interactions are those that occur beyond 1.5 meters from the user and up to 30 meters or beyond depending on whether the former or latter is being referenced [29, 62]. Traditional direct manipulation techniques, which are effective in near-field scenarios, often become impractical or inefficient when applied to medium or far-field tasks due to increased hand jitter, possible poor vision, and impaired depth perception when manipulating these distant objects [15]. To address such challenges, novel interaction techniques have been developed and refined with the goal of offering distinct advantages depending on the task and context. These techniques include controller-based methods, gaze-based interactions, and Worlds in Miniature (WIM) approaches, all of which aim to enhance the user's ability to accurately and efficiently interact with distant virtual objects [11, 55, 61]. The effectiveness of these techniques is often influenced by factors like the complexity of the environment, the precision and efficiency required by the task, and the user's familiarity with the interaction modality [31]. For instance, integrating gaze with controller inputs or using a WIM techniques to transform far-field interactions into ones that are near-field can significantly reduce cognitive load and improve interaction efficiency [11, 30]. Such approaches reflect an ongoing effort to balance usability and precision in far-field interactions [50].

2.3.1 Controller-Based Techniques

Controller-based techniques remain fundamental in facilitating medium and far-field interactions within VR and AR. These methods typically incorporate raycasting as a primary interaction technique, where users point and select objects at a distance using a virtual ray extended from a handheld controller [69, 70]. Despite its intuitive nature, raycasting often struggles with precision, especially when interacting with small or distant objects, or when objects of interest are occluded. This has led to the development of various enhancements to these controller and raycast-based techniques, aimed at improving accuracy and reducing user frustration [54, 71]. Interaction techniques such as the Go-Go technique allow for dynamic adjustment of interaction distances, thereby bridging the gap between near- and far-field interactions and enhancing overall user immersion [54]. Additionally, hybrid approaches that combine raycasting with other input modalities, such as gaze or near-field manipulation, have shown promise in improving precision and reducing cognitive load [3, 5, 57]. Overall, research indicates that while controller-based techniques are effective for medium and far-field interactions, their performance can be significantly enhanced through the integration of complementary input methods [60].

2.3.2 Gaze-Based Techniques

Gaze-based interaction techniques have garnered significant attention for their potential to streamline medium-field interactions in AR and VR. By leveraging the natural movement of the eyes, these techniques enable users to select and manipulate virtual objects without the need for physical controllers or extensive hand movements [34, 63]. This approach not only reduces physical strain but also enhances interaction speed and efficiency, particularly in tasks that involve frequent or repetitive selections [9, 49]. Additionally, hybrid approaches that combine gaze with other input methods, such as gestures or voice commands, have been explored to improve accuracy and user satisfaction in medium-field interactions [45, 47]. Studies suggest that while gaze-based techniques are highly effective for quick selections and simple interactions, they are best used in conjunction with other modalities for more complex tasks [10, 27]. Integrating gaze with other interaction methods offers a balance, enhancing both the efficiency and precision of interactions [34].

2.3.3 Worlds in Miniature (WIMs)

Worlds in Miniature (WIM) techniques offer a novel approach to medium-field interactions by providing users with a manipulable, scaled-down representation of the virtual environment [31, 61]. This approach allows users to interact with distant objects or navigate large virtual spaces more effectively by bringing either the entire environment or a subset of the environment within easy reach [19]. The WIM concept, first introduced in 1995 [61], has been iteratively expanded towards improving user control and interaction fidelity [56]. World-in-Miniature (WIM) systems are effective in addressing occlusion issues and excel in tasks requiring spatial cognition and navigation [14]. Their performance can be further optimized when combined with bimanual input and occlusion-reducing strategies [4, 68]. Techniques like Voodoo Dolls demonstrate that manipulating clones of distant objects can significantly improve positional and rotational accuracy [51, 52]. This technique allows users to spawn clones of distant objects that are part of predefined object hierarchies (where parent objects cannot be spawned in isolation), allowing for avenues of improvement.

2.4 Comparative Studies across Near, Medium and Far Field Interaction

Studies across near, medium, and far-field interactions have revealed that different interaction techniques excel depending on the context and specific task requirements [24, 39]. Near-field interactions often benefit from the use of natural hand gestures and tangible interfaces, which offer high precision and an intuitive user experience [23, 39]. In contrast, medium-field interactions are generally more suited to controller-based techniques and gaze-based methods, which can efficiently manage interactions with objects at a distance [63, 70]. Research has also shown the effectiveness of hybrid approaches that combine multiple input modalities, such as gaze and gesture, to enhance both precision and ease of use across different interaction ranges [49, 53]. Research suggests that while each technique has its strengths, their limitations can often be mitigated through the use of complementary interaction methods, leading to a more seamless and effective user experience across varying interaction distances [11]. This multi-modal approach appears to be particularly valuable in complex virtual environments where users must frequently switch between interacting with objects at different distances [24, 30].

3 System Description

3.1 Physical Room Description

The room where the study was conducted measured 4.87 meters wide and 5.79 meters long. Participants stood within a rectangular area (1.21 meters wide and 0.91 meters long) marked with hazard tape. This area was strategically positioned at one end of the room to provide users with an optimal viewing experience. Participants remained within this area but were allowed to move freely within this space when performing the task described in section 4.1. They were given examples of permissible movements, including bending down, standing on their toes, and walking around within the rectangle. This flexibility was intended to enable natural interaction with the virtual environment while maintaining a controlled study area.

3.2 Hardware and Virtual Components (Holograms)

The simulation was built using Unity 2022.3.9f1 and deployed as an Android Application Package (APK) on a Magic Leap 2 Augmented Reality HMD. The development was conducted on a MacBook Pro 14" (2021) equipped with an Apple M1 Silicon chip and 16GB of RAM. The Magic Leap 2 features an AMD 7nm Quad-core Zen 2 processor with 8 threads, a 14-core computer vision processing engine (CVIP), 16GB LPDDR5 RAM, and an AMD RDNA 2 GPU. It has a 70°diagonal field of view and a frame refresh rate of 120 Hz. The built-in speakers provide a head-related transfer function (HRTF) for accurate spatial audio.

A set of virtual holograms were used for the pick-and-place task described in section 4.1. These holograms included four distinctly colored target dice, 100 non-target white obstacle cubes, placement pillars, reference dice and outline meshes. Figure 2 shows the layout of these holograms. The side length of the target dice was 0.18 meters and the obstacles' side lengths were 0.2 meters. The colors of the dice and obstacles were chosen, ensuring sufficient visibility, distinguishability from each other, and contrast with the background. The positions and orientations of these objects were randomized within a defined volume, ensuring that there were no intersections or overlapping geometries. The pillars atop which the target dice had to placed were positioned in pairs, flanking either side of this volume. Each pillar displayed a reference die above it with one of its sides facing the user. Furthermore, there was a placeholder-like black wireframed cubical outline mesh on each pillar.

The volume hosting the dice and all obstacles was preset to be 3 meters wide, 2 meters deep, and 1.5 meters high, representing a cluttered environment. The center of this volume was set 4.5 meters in front of the user. This positioning meant that objects were located at distances ranging from 3.5 to 5.5 meters along the depth axis, 1.5 meters on either side of the user along the horizontal axis, and from 0.5 meters above to 1 meter below the user's head along the vertical axis. The HMD served as the origin of the application's coordinate system. The height of the pillars was programmed, taking to consideration each user's height determined from the calibration routine outlines in section 3.3. Accordingly, both front pillars were a third of the user's height, while the two rear pillars were a sixth of the user's height. A submit button would appear at the end of each trial (section 4.1) at 9/16th of the user's height.

3.3 Calibration and Registration Routine

Prior to the start of the experiment, participants donned the HMD and completed a sequence of three calibration steps to ensure an optimal and individualized experience. The initial calibration focused on fit, verifying that participants wore the appropriate forehead and nose pieces tailored to their facial structures, a step crucial for both prolonged comfort and the accuracy of subsequent calibrations. Following this, a gaze calibration routine was conducted, ensuring that the system accurately tracked users' eye gaze. The final calibration step addressed floor level alignment; participants were instructed to gaze straight ahead at their perceived eye level while measurements were acquired using the HMD and controller to ascertain their eye height. This measurement was instrumental in generating virtual pillars that appeared to emanate naturally from the ground when viewed through the head-mounted display. Collectively, these calibration steps were essential in tailoring the virtual scene to each participant, spawning the holograms with respect their eye-heights.

3.4 Interaction Techniques

This study examined three distinct medium-field interaction techniques, detailed below. Each technique leveraged the HMD and its embedded cameras to track users' hand movements, eye gaze, and controller input as needed. All interaction techniques were implemented using real-time tracked pose drivers within the Unity engine, ensuring responsive tracking throughout the study. All techniques were used to perform the task described in section 4.1

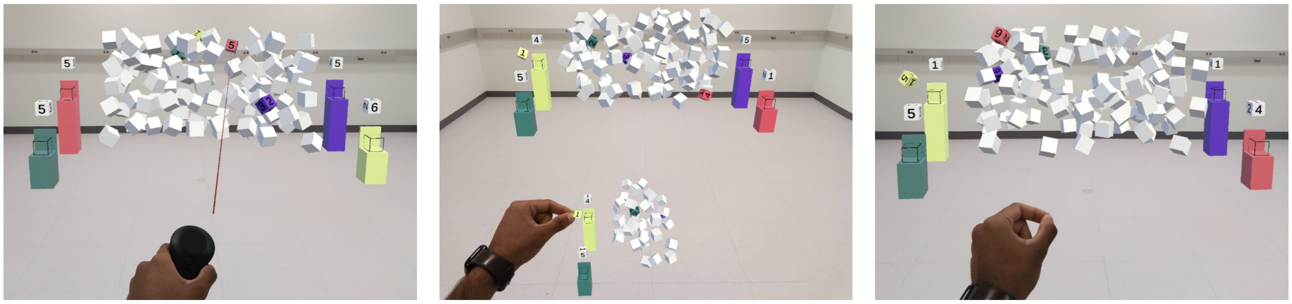
Raycast + Controller: In this technique, a line/ray is projected from the controller towards space. Users can point the controller in a direction of their choosing, effectively moving the ray towards objects of interest. When the ray intersects an object, the object gets highlighted with a blue outline, indicating that it can be grasped and then moved or rotated using the controller's trigger and bumper buttons respectively. When the trigger button remains pressed, a selected object can be moved by either translating or rotating the controller. Physical translation of the controller along the horizontal and vertical axes result in matched translation of the object along these axes, thus maintaining a one-to-one mapping. However, movement of a selected object along the depth axis undergoes a sixfold gain relative to the translation of the physical controller, thus amplifying the object's movement. This improves the ability to interact with distant objects without requiring large physical controller movements. This parameter was determined from the six pilot tests conducted for parameter tuning to make the task achievable within the constraints of controller-based interaction. Rotation of the controller when the trigger button remains pressed moves a selected object by an amount that is governed by a function of the angle of rotation and the distance of the object from the controller.

$$d_{traversed} = \sqrt{d_i^2 + d_f^2 - 2 d_i d_f \cdot \cos\theta}$$

Here, d_i represents the length of the line drawn from the controller to the object prior to its manipulation, while d_f denotes the length of the line drawn from the controller to the object after its manipulation. The angle θ is the angle formed between these two lines. The term $d_{traversed}$ corresponds to the distance between the initial and final positions of the object, indicating the extent of translation along the axis. When selected using the bumper button, objects' orientation can be manipulated by rotating the controller which maintains a one-to-one mapping to the object's orientation, based on [32]. For example, a 15° rotation of the controller on a given axis will result in a 15° rotation of the selected object on that same axis. The system supported concurrent rotation across all three axes, making for an intuitive user experience. The translation and orientational manipulations were deliberately designed to be isolated from each other rather than using a single button delimiter based on recommendations made in [44] and feedback from the pilot tests, thereby affording higher degrees of precision in both the movement and rotation of the distant objects.

Gaze + Pinch: This technique leverages eye gaze and a hand gesture to manipulate objects. Users move their eyes, directing their gaze toward a virtual object of interest. The system tracks users' gaze vector, highlighting with a blue outline, objects that said vector intersects, indicating that it is manipulatable. Once the object is highlighted, the user can perform a pinch gesture using their hand to interact with it. A pinch gesture is made by bringing the thumb and index fingertips together, similar to a pinching motion. This gesture acts as a "grab" command, allowing the user to pick up, translate, or rotate the object. Translation of selected objects is scaled along all three axes by a factor of six relative to movement of the user's phyical end-effector. The sixfold translational gain was added to make the task achievable without requiring excessive movements of users' arms. Manipulation of the orientation of a selected object is directly mapped to the rotation of the hand making the pinch gesture, with a one-to-one mapping. For instance, a 15°rotation of the hand on any given axis results in a 15°rotation of the object on that same axis. This technique enables simultaneous manipulation of all three axes, offering a fluid and intuitive user experience. It continues to be used in current state of the art XR systems including the Apple Vision Pro.

The Force: In this interaction technique, users can select multiple objects of interest and spawn miniature clones of them in their proximity. Manipulations performed on these clones are mirrored to the original objects, allowing for near-field manipulation of distant objects. Users can interact with these clones through a pinch gesture which allows to grab, translate, and rotate the objects. The clones were scaled down to a sixth of the original objects and a one-unit movement of the clones resulted in a six-unit movement of their original object counterparts, thus preserving the magnitude of translational gains across the three interaction techniques. To select objects to be miniaturized, the user extends their arm with an open palm, waving it across objects of interest. To provide feedback of which objects are going to be miniaturized, a blue cone extends from the user's wrist with its apex at the wrist and its base being visualized. Users can adjust the positioning of the cone by moving their hands accordingly. As the user performs the wave gesture, all objects' borders that ever collided with the cone are highlighted in blue and added to a list of selected objects. Once the selection of objects is complete, the user confirms it by closing their hand into a fist, which changes the objects' borders from blue to yellow,



(a) **Raycast + Controller** (b) **The Force** (c) **Gaze + Pinch**

Figure 2: First person perspective of the experience using the different interaction techniques in the study. Each sub-figure shows the user manipulating a target die with that technique. Note that sub-figure (b) involves the user manipulating the miniaturized clone of the target die.

signaling that selection is complete and that miniaturization can occur. The user now has two options after confirming their selection, either to discard or spawn miniatures. If the user feels that the selected objects do not include all objects of interest, they can open their fist with the palm facing downwards, deselecting all objects and turning their borders back to normal. This resets the selection process, allowing them to perform the wave gesture again. On the other hand, if the selected objects include all of their intended items, the user can open their fist with the palm facing upwards to spawn miniature clones of the selected objects at 1/6th the original scale. The position in 3D space where the fist is opened determines the placement of the nearest miniature, with other miniatures spawning relative to this position. This maintains the perspective from which the original objects were viewed and provides flexibility in orienting the workspace with the miniatures. Once spawned, the user can manipulate these miniatures, with all translational and rotational adjustments mirrored on the original objects, allowing for near-field manipulation of far-field objects. When finished, the clones can be destroyed by performing a clap-like gesture, allowing the selection process to be performed again.

In all three techniques, selected objects underwent a sixfold translational gain along the depth axis. With *The Force* and *Gaze+Pinch* techniques, this gain also occurred in the horizontal and vertical axes, ensuring consistency between these conditions. However, it is counterintuitive to provide the same translational gain along the horizontal and vertical axes for the *Raycast+Controller* technique because this introduces a mismatch between where the user points the controller and where the object has moved, making the ray now appear to point in direction different from the original forward vector used. This would imply that the controller no longer points to the object, but rather connects it through a misdirected ray, making for poor design. Unsurprisingly, pilot tests confirmed this.

4 Experiment

4.1 Task

For this experiment, a simple pick-and-place task was conceptualized wherein participants had to manipulate four distinctly colored target dice from a field of obstacles (non-target cubes), carefully placing each of the dice at designated locations over a number of trials. This task was performed across three blocks, one for each of the interaction techniques empirically evaluated in this study (section 3.4). Each block commenced with a practice trial, allowing users to familiarize themselves with the mechanics of the technique associated with that block. In every trial, each die had to be placed on one of four distinctly colored target pillars based on the die's corresponding color (e.g., red die on red pillar, green die on green pillar, etc.). These pillars were located on either side of the field of obstacles, each featuring a placeholder outline mesh that indicated the precise location and orientation with which the die had to be placed. Additionally, a reference die was displayed atop each pillar, requiring users to match the forward-facing face of the target die being placed to that of the reference (eg. reference die displaying number 5 facing the user meant that users had to place the target die with face 5 facing forward).

The system was designed to tolerate a predefined maximum level of inaccuracy in each die's positional and orientational accuracies. In other words, users had to place each of the targets, achieving a minimum level of positional and rotational accuracy for them to qualify as valid placements towards completing a trial. The tolerance threshold for positional accuracy required targets to be placed such that their centers had to be anywhere less than 0.25m from the center of their respective outline meshes. The tolerance threshold for rotational accuracy of targets permitted a maximum angular error of 90° aggregated across all axes. In each trial, the four target dice had to placed, meeting these constraints for participants to be able to proceed to the next trial. These threshold constraints were determined from pilots, taking into consideration the objectives of avoiding ceiling effects while also making the task achievable. These constraints also helped ensure that there were no avenues for simulation-based exploits that users could employ in attempts to game the system and simply proceed through the experiment, placing the targets arbitrarily. This task hence required precise perception-action coordination, requiring users to carefully place objects at designated locations with specific orientations.

At the start of every trial, the target dice, obstacles, and pillars along with their outline meshes and reference dice were all spawned based on content described in section 3.2. Users would then begin manipulating each of the four target dice, placing them on their respective pillars based on their color. When all four target dice met the threshold constraints discussed previously, a red-colored submit button would appear beside the user on their non dominant side. When satisfied with the placements of all targets, user could press the submit button, marking the end of that trial and initiating the next one. This continued until all four trials in that block were completed, seguing to the next block.

For all events during a trial, multi-modal feedback was provided to make for intuitive interaction design. Auditory feedback was delivered using the built-in headphones of the HMD. The system provided appropriate feedback for hover, selection, release, and trial completion events. At rest, all objects did not have any borders around them, be it target dice or non target cubes. When hovered over, the outline borders of targets/obstacles turned blue, signifying graspability. When grasped, the outlines turned yellow and a grasping sound was deployed. Similarly, when released, the object's yellow outline disappeared, and a release sound was deployed. Auditory feedback was intentionally provided for the grasping events given users' preference for it over visual feedback only [7]. With *The Force*, feedback was also provided for the events associated with creating and destroying clones. When selecting objects for cloning using the open palm gesture, a persistent selection sound was continuously deployed to indicate that *The Force* was in use. Objects with a history of being hovered over had blue outlines, indicating that they were a part of the selection. Once a fist was made to confirm the selection, the persistent sound stopped, and the blue outline of those selected objects turned yellow, effectively denoting confirmation of the selected set of objects. Subsequently, when users made an upward facing palm gesture, a distinct sound was deployed and the clones were spawned. When dismissing the created miniatures, a deletion sound was deployed, accompanying the disappearance of the miniatures from the user's field of view. The visual and auditory feedback pertaining to any event was presented simultaneously, providing rich multi-modal feedback that was clearly indicative of the different events that transpired during a trial.

4.2 Study Design

To empirically evaluate and compare how the interaction techniques described in section 3 affect performance and perceptions in a medium-field pick-and-place task in augmented reality, we employed a within-subjects study design. The study focused on a single independent variable: the Interaction Technique, which was manipulated across three experimental conditions: (1) *Raycast+Controller* (RC); (2) *Gaze+Pinch* (GP); and (3) *The Force* (TF). Participants performed the pick-and-place task described in section 4.1 over a number of trials across a series of blocks, each featuring one of the three interaction techniques. In each block, users performed four trials each of which involved picking and placing four dice at designated locations. Each participant hence performed a total of 12 trials across the three blocks, thus accruing up to a total of 48 placements (12 trials X 4 placements per trial). Participants experienced all three techniques, with the order of blocks counterbalanced using a Balanced Latin square design, ensuring equal representation in each of six possible condition orders. This approach helped control for potential order effects, allowing users to fully adapt to one technique before moving on to the next, similar to methods employed in [64].

4.3 Measures

4.3.1 Accuracy

Positional Error (Distance) - The euclidean distances between the centers of the dice and the centers of their corresponding outline meshes (placement locations) were computed across all trials in a block. This measure represents the average amount of positional error between users' actual placements of the dice and their ideal placements (error of 0). A lower positional error corresponds to a higher accuracy in correctly placing the dice.

Orientational Error (Angle) - The orientational differences between the dice's placements and their corresponding outline meshes were computed across all trials in a block. These differences were quantified as the least amount of rotation required for the placed dice to perfectly coincide with their outline meshes. This measure represents the average amount of orientational error between users' actual placements of the dice and their ideal placements (error of 0). A lower orientational error implies higher accuracy.

4.3.2 Efficiency

Time on trial - The interval between the start and end of each trial (pressing the submit button) was computed across all trials in a block. This measure represents the average time required to complete a trial, placing the four dice at their designated locations. The less time on trial, the more efficient users are.

Working-Time with Target Dice - The total handling-time of the target dice was computed and then averaged across all trials in a block. It represents the average duration specifically spent interacting the target dice while excluding time spent on interactions with irrelevant objects (obstacles) during a trial. The lower the working time with targets, the more optimal and efficient users are.

Number of Interactions with Obstacles - In each trial, the number of times users interacted with obstacles was computed and then averaged across all trials in a block. This measure represents the average number of obstacle-interactions in each trial. A smaller number of obstacle interactions corresponds to a higher efficiency.

4.3.3 User Experience

Perceived Usability - Users' perceived level of usability associated with interaction techniques was measured using the PSSUQ inventory [36]. A lower score corresponds to lower perceived usability.

4.4 Research Questions & Hypotheses

The overarching aim of this study was to address the following research question: **"How does *The Force* compare to contemporary interaction techniques with respect to medium-field interactions in augmented reality?"** Specifically, we were interested in ascertaining how well this technique facilitates accuracy and efficiency in medium-field pick-and-place tasks that require fine motor control and precision. We hence conceptualized a task (section 4.1), operationalizing accuracy and efficiency using measures described in section 4.3. We formulated the following hypotheses, reflecting work discussed in section 2 and subsequent portions of this section.

H1: The Accuracy will be highest when using *The Force*.
H2: The Efficiency will be highest when using *The Force*.
H3: *The Force* will rate highest on usability (PSSUQ)

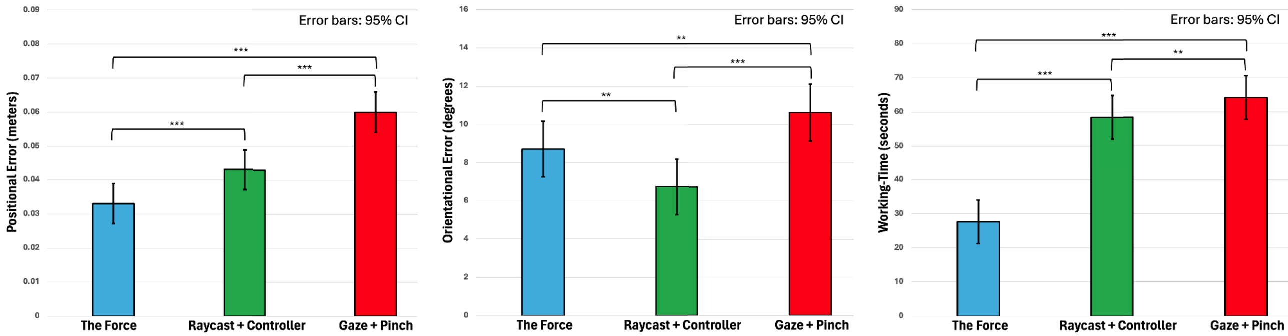
The rationale behind these hypotheses stems from characteristics and affordances that are inherent in each interaction technique. *The Force* is expected to yield the highest accuracy by allowing users to perform the medium-field task using near-field clones that provide users with a better sense of spatial relationships among the distant objects, improved vision of the objects and their placement locations, and a more natural experience through direct proximate interactions. For these reasons, it is expected that this technique yields the least positional and orientational error (highest accuracy). The ability to see different perspectives when working with the clones is likely to allow users to avoid having to interact with many obstacles, resultin in lesser obstacle-interactions. Furthermore, the clones being proximate with improved vision will allow users more quickly work on them and place them at their locations, thereby resulting in lesser trial and working-time. For these reasons, it is reasonable to hypothesize that *The Force* will produce the highest efficiency. Given the naturalness of being able to directly manipulate clones in users' proximity, it is expected that *The Force* be perceived as more usable than the other techniques, based on prior work showing the same.

4.5 Participants

An apriori power analysis using G*Power revealed that for a study of three techniques tested per participant, considering $\alpha = 0.05$, Power (1-β) of 0.8, a medium effect size f= 0.25, correlation among repeated measures of 0.5, and assuming that sphericity is met, the estimated sample size was 28. Thus, a total of 30 participants were recruited for this study that was conducted under the oversight of an Institutional Review Board (University of Florida IRB, Protocol #ET00041813). Participant ages ranged from 18 to 54 years old (M = 24.23, SD = 9.13), 10 of whom identified as female, one as non-binary, and the rest male. All participants reported themselves as not color blind and had normal or corrected-to-normal (20/20) vision, along with normal motor function of their upper and lower body. They rated their prior experience with VR (M = 4.37, SD = 1.65), AR (M = 3.80, SD = 1.58), and video games (M = 5.70, SD = 1.32), With 1 being "No experience at all", and 7 being "Highly experienced". Participants' Laterality Index (as measured using [46]) ranged from -80 to 100 (M = 68.00, SD = 41.87).

4.6 Procedure

Upon arrival to the laboratory, participants were greeted and requested to read and sign a consent form. After consenting, participants filled out a demographics questionnaire that included information about their backgrounds, experience with AR, and handedness [46]. The experimenter then explained the task to be performed



(a) Positional Error (distance) (b) Orientational Error (degrees) (c) Working-Time with Target Dice

Figure 3: Comparison of Positional Error, Orientational Error, and Working-Time with Target Dice across the three Interaction Techniques.

in the study (see section 4.1), detailing the logistics involved with progressing through the three blocks of the experiment. Participants then underwent a set of calibrative steps detailed in section 3.3 to ensure consistent and optimal viewing experiences. They then proceeded to perform the task in each block, one after the other. Users were encouraged to perform the task to the best of their abilities, trying to achieve high levels of both accuracy as well as efficiency.

At the start of every block, the experimenter presented a video demonstrating the mechanics of the interaction technique specific to that block. Participants then donned the Magic Leap 2 HMD and engaged in a practice trial (comprised of 4 dice pick and place) to familiarize themselves with the interaction technique. After each block, participants filled out the PSSUQ questionnaire and were allowed to take a break before proceeding to the next. Upon completion of all three blocks, participants removed the HMD, and filled out a brief post experiment questionnaire. They then engaged in a short semi-structured interview with the experimenter to discuss their experience with the different interaction techniques, strategies they used, and aspects they felt warranted improvement with each technique. Participants were then debriefed and compensated. On average,the experiment took about 90 minutes to complete.

5 Results

The positional error, oritentational error, working-time with target dice, time on trial, and the number of interactions with obstacles, were the dependent variables considered for analyses. All dependent variables were initially screened for possible outliers using Mahalanobis distance [17, 41]. Since five dependent variables were analyzed, the distances were compared to a critical x^2 value with five degrees of freedom. Nine cases were identified as potential outliers. However, since it was found that the inclusion of outliers in the analysis did not affect the results, we report the results of the analysis, including the outliers. Since repeated measures of each dependent variable was considered for each participant, variables had considerable nesting. This means that a portion of the variance in each dependent variable can be attributed to a common source – the participant themselves. To properly account for such variance, linear mixed effects modeling was used instead of repeated measures ANOVA [24, 35]. Prior to conducting analyses, the extent of nesting in the data was assessed by computing the intraclass correlation coefficient (ICC) from the null model for each dependent variable separately. The ICC was computed as 0.26 for the positional error, indicating that approximately 26% of the variance in this dependent variable was associated with the participant and that the assumption of independence was violated. Similarly, the ICC was calculated to be 0.37 for average orientational error, 0.33 for working-time with target dice, 0.59 for time on trial, and 0.12 for the number of interactions with obstacles. Using linear mixed effects modeling is ideal in this case. For each model, the interaction technique was included as the fixed effect. The effect size for each fixed effect is presented as the change in R^2 (proportion of variance explained) comparing the model that includes the effect and the same model with the effect removed. The resulting sr^2 (semi-partial r^2) is the percentage of variance uniquely accounted for by the fixed effect [58]. For all the models in the analyses reported, the only random effect computed was the intercept based on the Participant ID.

5.1 Positional Error (Distance)

There was a significant effect of interaction technique on the positional error, F (2, 342) = 76.4, $p < 0.001$, $sr^2 = 0.21$. The error was significantly smaller in *The Force* technique (*M* = 0.033, *SE* = 0.003) as compared to the *Raycast+Controller* (*M* = 0.043, *SE* = 0.003), t = 4.59, $p < 0.001$, as well as the *Gaze+Pinch* technique (*M* = 0.06, *SE* = 0.003), t = 12.26, $p < 0.001$. The error was also significantly smaller in the *Raycast+Controller* as compared to the *Gaze+Pinch* technique, t = 7.77, $p < 0.001$. Figure 3a depicts these results.

5.2 Orientational Error (Angle)

There was a significant effect of interaction technique on the orientational error, F (2, 342) = 23.86, $p < 0.001$, $sr^2 = 0.072$. The error was significantly smaller in the *Raycast+Controller* (*M* = 6.73, *SE* = 0.75) as compared to *The Force* technique (*M* = 8.71, *SE* = 0.75), t = 3.60, $p = 0.001$, as well as the *Gaze+Pinch* technique (*M* = 10.62, *SE* = 0.76), t = 6.9, $p < 0.001$. The error was also significantly smaller in *The Force* technique as compared to the *Gaze+Pinch* technique, t = 3.39, $p = 0.002$. Figure 3b depicts these results.

5.3 Working-Time with Target Dice

There was a significant effect of interaction technique on the working-time with target dice, F(2, 342) = 215.25, $p < 0.001$, $sr^2 = 0.34$. The time spent was significantly less in *The Force* technique (*M* = 27.6, *SE* = 3.26) as compared to the *Raycast+Controller* (*M* = 58.4, *SE* = 3.26), t = 16.43, $p < 0.001$, as well as the *Gaze+Pinch* technique (*M* = 64.2, *SE* = 3.28), t = 19.08, $p < 0.001$. The time spent was also significantly less in the *Raycast+Controller* as compared to the *Gaze+Pinch* technique, t = 3.03, $p = 0.007$. Figure 3c depicts these results.

5.4 Time on Trial

There was a significant effect of interaction technique on the time on trial, F (2, 342) = 9.14, $p < 0.001$, $sr^2 = 0.02$. The time spent was significantly less in the *Raycast+Controller* (*M* = 107, *SE* = 6.86) than *The Force* technique (*M* = 123, *SE* = 6.86), t = 4.26, $p < 0.001$, as well as the *Gaze+Pinch* technique (*M* = 116, *SE* = 6.91), t = 2.37, $p = 0.049$. However, there was no significant difference in the time spent when participants used the *Gaze+Pinch* technique as compared to *The Force* technique. Figure 4 depicts these results.

5.5 Number of Interactions with Obstacles

Since the number of interactions with obstacles is a count variable, performing a poisson regression is ideal. However, for poisson

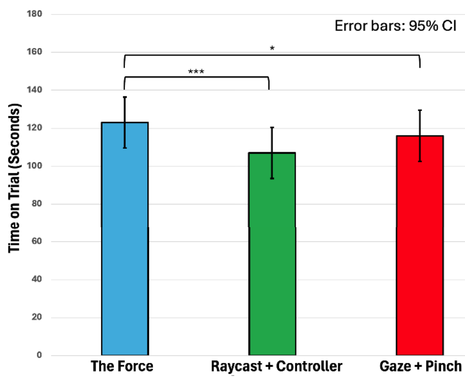


Figure 4: Total time spent on trial

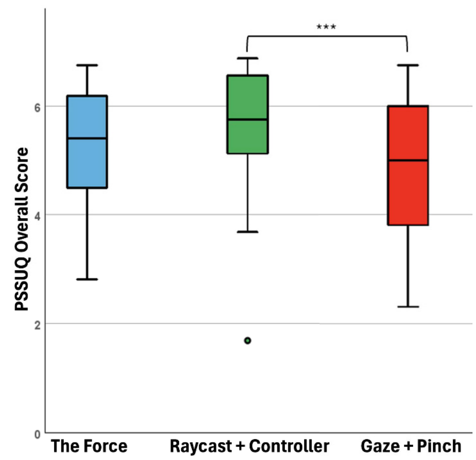


Figure 5: PSSUQ Overall Score across Interaction Techniques

regression, if the dispersion parameter ϕ is greater than 1 (overdispersion), a negative binomial regression can be used [13]. Since ϕ was found to be 2.27 for the model with only the fixed effects, negative binomial regression was used to fix overdispersion.

There was a significant effect of interaction technique on the number of interactions with obstacles participants had, χ^2 (2, N = 376) = 142.29, $p < 0.001$, $sr^2 = 0.31$. The number of obstacles interacted with was significantly fewer in *The Force* technique (M = -0.17, SE = 0.17) as compared to the *Raycast+Controller* (M = 1.41, SE = 0.14), t = 9.18, $p < 0.001$, as well as the *Gaze+Pinch* technique (M = 1.85, SE = 0.14), t = 11.67, $p < 0.001$. The number was also significantly fewer in the *Raycast+Controller* as compared to the *Gaze+Pinch* technique, t = 2.95, $p = 0.009$.

5.6 Perceived Usability

There was a statistically significant difference in usability across the three techniques, $\chi^2(2) = 16.974, p < .001$. Post hoc analyses with Wilcoxon signed-rank tests was conducted with a Bonferroni correction, resulting in a significance level set at $p < 0.017$. This meant that statistically significant differences between the techniques exist only when $p < .017$ because the significance level initially used (in this case, 0.05) divided by the number of comparison tests run reflects the Bonferroni adjustments. Median (IQR) perceived usability levels for the Raycast + Controller, *The Force*, and *Gaze+Pinch* techniques were 5.86(5.14 to 6.56), 5.47(4.55 to 6.12), and 5.0(3.78 to 6.03) respectively. There were no significant differences between *The Force* and *Raycast+Controller* techniques (Z= -1.814, $p = .07$) or between *The Force* and *Gaze+Pinch* techniques (Z= -2.131,$p = .033$), despite an overall reduction in usability in the *Gaze+Pinch* and *The Force* techniques compared to the *Raycast+Controller* technique. However, there was a statistically significant reduction in perceived usability in the *Gaze+Pinch* technique vs the *Raycast+Controller* technique (Z= -3.53, $p < .001$). Figure 5 depicts these results.

5.7 Interview Feedback

User feedback on the techniques highlighted the need for improved precision, control, and tracking. With respect to the *The Force*, users expressed the need for improvement in the selection component, suggesting to create a more stable and refined selector cone, afford the ability to better add and remove objects from selection, and consider two-handed controls with one hand being used to add and the other to remove objects from the selection. To address the selection inadequacies, one user commented "I think maybe just make it a little smoother" and another said "having more controlled, like a smaller beam, I guess. It would be easier". A hybrid approach was even recommended wherein *The Force* and *Gaze+Pinch* techniques could be combined by using eye-gaze as the selection metaphor for the clone-based interaction technique. Some users also suggested refining the technique to afford the ability to scale the miniature view on demand. When asked about utility, several users felt it would excel when used in scenarios involving occlusion like having to inspect and work on complex 3D models involving architecture and automobiles. For the *Gaze+Pinch* technique, users consistently pointed to challenges with eye-tracking precision, especially for small or distant objects, suggesting the need for accurate tracking. When asked about how to improve this technique, one user said "but I guess just better tracking or being able to like stay consistent with an object". Other suggestions included increasing the transparency of objects outside the user's direct view to reduce peripheral distractions. This would also help prevent accidental interactions with objects present in the user's peripheral vision. Consistent and accurate hand tracking was also emphasized as essential for maintaining precision across both of these techniques.

6 Discussion

The statistical analyses on the measures of accuracy revealed that the average positional error was significantly lower in *The Force* than the *Raycast+Controller* and *Gaze+Pinch* techniques (figure 3a). This meant that in terms of correctly centering the placements of the dice at their designated locations, users were most accurate when using *The Force*. As expected, performing the task based on the miniature clones allows users to be more accurate because the task becomes a near-field direct manipulation task. Consequently, users have improved vision and perspective of the task-relevant elements (targets, their locations, etc.), enabling higher degrees of accuracy [3]. Given these benefits, it was also expected that the orientational accuracy would be the highest in *The Force* technique. Interestingly, however, we found that *The Force* while outperforming the *Gaze+Pinch* technique, produced higher orientational errors than the Raycast technique. This can be explained by prior work showing that controller-based manipulations of objects can facilitate higher levels of accuracy than gesture-based manipulations due to the precise hand-tracking required for the latter [6, 42]. Note that according to [12], the effect sizes were large for the positional error and small for the orientational error, indicating that the interaction technique was more impactful on the former than on the latter. Nonetheless, the relatively superior performance of *The Force* on our measures of accuracy offer partial support for hypothesis **H1**, indicating that cloning and manipulating miniatures of distant objects can enable higher levels of accuracy for medium-field tasks that require high degrees of precision and accuracy. These results generally align with prior research, showing that near-field manipulation and viewing offer benefits for interacting with objects located further away [3].

Analyses on the measures of efficiency gave us some useful insights. In terms of the average time taken to complete the trials, we found a main effect of the interaction technique. However, the effect size for this measure was small, indicating a relatively modest influence of the factor on the time taken. We found that users took longer when using *The Force* than both The *Raycast+Controller* and *Gaze+Pinch* techniques (figure 4). In contrast to what we expected, we discovered that *The Force* exhibited limitations in terms of its ability to support faster completion of the trials. This becomes particularly insightful when considering the analyses of the working-time with the target dice and the number of times users interacted with obstacles which feature large effect sizes that indicate a stronger impact. In these aspects, we found that *The Force* required the least working-time with the targets and the fewest number of interactions with obstacles (figure 3c and section, 5.5). The former can be explained by virtue of the fact that the large amounts of movements of the distant dice (to their designated locations) could be quickly achieved, looking at their proximate miniature clone counterparts, thus requiring less working-time. Taking these smaller target handling-times observed in conjunction with the accuracy results discussed previously, we find that users were able to achieve higher levels of accuracy with *The Force* even though they spent less time working with the dice. This serves to indicate that this technique can enable increased accuracy without compromising efficiency, making it valuable in scenarios where both these aspects are desirable. The trends on the number of obstacle-interactions are understandable because working with the proximate clones allowed users to easily shift their perspective, directly grabbing and manipulating the targets from the cluttered environment without needing to clear obstacles out of the way. Such line of sight requirements explain why both the *Raycast+Controller* and *Gaze+Pinch* techniques involved a larger number of obstacle-interactions than *The Force*. Collectively, these results offer considerable support for hypothesis **H2**, indicating that interactions performed with *The Force* can generally support increased efficiency. The ability to manipulate miniaturized clones of cluttered environments seems to allow users to expeditiously work on relevant objects, optimally reducing irrelevant interactions and the working-time involved. More importantly, these results, when taken together, offer valuable insights into improving *The Force*. Given that it required lesser working-time with the targets and fewer number of irrelevant interactions (interactions with obstacles), it appears that the increased time taken to complete the task with this technique stemmed from inadequacies in facilitating swift selection and cloning, the initial gestures involved. These findings were corroborated by comments made in the debriefing interviews where users reported dissatisfaction with the selection metaphor and the miniature clone-spawning aspects of *The Force*, suggesting ways to make improvements (section 5.7). It follows that iteratively incorporating such suggestions will help refine this technique.

In terms of usability, we found that participants perceived the *Raycast+Controller* as significantly more usable than the *Gaze+Pinch* technique (figure 5). These trends may be related to the fidelity of these interaction metaphors and their abilities to support effective interactions. Along these lines, high degrees of precision is often a requisite for gaze-based interactions to be seamless and effectual. Research has shown that both selection and smooth pursuit are more accurately performed with controllers than eye-gaze [25, 40], highlighting the superiority of controller-tracking over gaze-tracking. Our results on the measures of accuracy and efficiency directly align with these findings, suggesting that the controller-based raycasting technique allowed users to be more effective and efficient at performing the task than the gaze-based pinching technique. These tracking limitations associated with the *Gaze+Pinch* technique was common theme that emerged from our post-study interviews and help explain why it was outperformed and perceived as less usable. While hypothesis **H3** was not supported, we found that *The Force* was not significantly different from either of the other techniques in terms of perceived usability. Notwithstanding its relatively superior performance in supporting accuracy and efficiency, participants were keen to suggest avenues to improve the usability of this technique. Some suggestions included the ability to deselect selected items, dynamically modify the geometry of the selector cone, use gaze a method for selection of objects to clone, etc. Such feedback is useful in the refinement of *The Force*, a technique that users deemed useful for scenarios involving clutter and occlusion.

In the pick-and-place task employed in this study, *The Force* allowed users to selectively spawn and work on miniature clones of any objects in the environment, including the pillars and outline meshes. The improved performance observed with this technique is hence a natural consequence of the task now becoming proximate with improved vision of its components. While it can be argued that this ability to transform a medium-field task into one that is near-field is a limitation of our study design in ensuring consistency between the techniques, we consider it an inherent affordance offered by this technique. Along these lines, we believe that the utility-value of this interaction technique stems from being able better perform fine motor interactions on clones situated in one's promximate reach envelope than having to work on distant objects that naturally involve reduced visual acuity. With environmental clutter often posing a challenge for the efficient performance of medium-field interaction tasks, the potential to easily change perspective also makes it advantageous to work on scaled-down clones. Our study's results serve to indicate that techniques like *The Force* can be considered by AR developers in scenarios that require precision for interactions occurring at farther distances like inspecting and moving components in 3D architectural and automobile models. Along these lines, when interactions with such entities occur, creating proximate clones can allow for easy inspection and more precise manipulation of tightly packed 3D artifacts. This is not to say that cloning is the only solution to address such requirements but that it appears to be a relatively simple and useful method of facilitating such interactions.

7 Conclusion and Future Work

In this work, we conceptualized *The Force*, an augmented reality interaction technique that allows users to clone distant objects and manipulate their replicas instead. We empirically compared it against established interaction techniques including gaze-based pinching and controller-based raycasting in terms of performing a medium-field pick-and-place task involving precise perception-action coordination. Users were tasked with manipulating four dice from a volume of obstacles, carefully placing them at designated locations over a number of trials, attempting to be as accurate and efficient as possible. We employed a within-subjects design, counterbalancing the order of the techniques evaluated. Results indicated that *The Force* resulted in relatively higher levels of accuracy and efficiency as a consequence of being able to perform medium-field tasks using proximate clones that afford better viewing and manipulation.

In future work, we aim to incorporate our insights and user-feedback to iteratively refine *The Force*. Our immediate interests lie in conceptualizing improved object-selection metaphors, creating a more refined selector entity with adjustable geometry and hybrid solutions that combine gaze-based selection for cloning. We also intend to explore the addition of two-handed controls with the clones, allowing to scale them and add/remove more objects to the miniatures.We then plan to evaluate it against other techniques (including other WIM variants) and in different types of tasks like object assembly, path planning, and collaborative manipulation tasks. These evaluations will help provide insights into how task-complexity and context influence the efficacy of this technique. Lastly, we wish to obtain answers to questions like: How should *The Force* be implemented for moving objects in dynamic environments? Should users be able to create clones of clones to amplify movements of objects even further? Obtaining answers to these kinds of questions will help determine how to better evolve this technique for augmented reality interactions in the years to come.

REFERENCES

[1] F. Argelaguet and C. Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, May 2013. 2

[2] R. T. Azuma. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, Aug. 1997. 1

[3] S. V. Babu, W.-A. Hsieh, and J.-H. Chuang. The Benefits of Near-field Manipulation and Viewing to Distant Object Manipulation in VR. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 408–417, Mar. 2024. ISSN: 2642-5254. 1, 2, 3, 8

[4] R. Balakrishnan and G. Kurtenbach. Exploring bimanual camera control and object manipulation in 3D graphics interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pp. 56–62. Association for Computing Machinery, New York, NY, USA, May 1999. 3

[5] M. Baloup, T. Pietrzak, and G. Casiez. RayCursor: a 3D Pointing Facilitation Technique based on Raycasting. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2019)*. Glasgow, United Kingdom, May 2019. 3

[6] G. Caggianese, L. Gallo, and P. Neroni. The vive controllers vs. leap motion for interactions in virtual environments: a comparative evaluation. In *International Conference on Intelligent Interactive Multimedia Systems and Services*, pp. 24–33. Springer, 2018. 8

[7] R. Canales and S. Jörg. Performance Is Not Everything: Audio Feedback Preferred Over Visual Feedback for Grasping Task in Virtual Reality. In *Proceedings of the 13th ACM SIGGRAPH Conference on Motion, Interaction and Games*, MIG '20, pp. 1–6. Association for Computing Machinery, New York, NY, USA, Nov. 2020. 5

[8] Y. Cha and R. Myung. Extended fitts' law for 3d pointing tasks using 3d target arrangements. *International Journal of Industrial Ergonomics*, 43(4):350–355, 2013. 2

[9] I. Chatterjee, R. Xiao, and C. Harrison. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, pp. 131–138. Association for Computing Machinery, New York, NY, USA, Nov. 2015. 3

[10] D. L. Chen, M. Giordano, H. Benko, T. Grossman, and S. Santosa. Gazeraycursor: Facilitating virtual reality target selection by blending gaze and controller raycasting. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, VRST '23. Association for Computing Machinery, New York, NY, USA, 2023. 3

[11] Z. Chen, J. Li, Y. Hua, R. Shen, and A. Basu. Multimodal interaction in augmented reality. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 206–209, Oct. 2017. 2, 3

[12] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, 2013. 8

[13] S. Coxe, S. G. West, and L. S. Aiken. The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives. *Journal of Personality Assessment*, 91(2):121–136, Feb. 2009. 8

[14] K. Danyluk, B. Ens, B. Jenny, and W. Willett. A design space exploration of worlds in miniature. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. 3

[15] F. El Jamiy and R. Marsh. Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing*, 13(5):707–712, 2019. 2, 3

[16] T. Feuchtner and J. Müller. Extending the Body for Interaction with Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 5145–5157. Association for Computing Machinery, New York, NY, USA, May 2017. 2

[17] L. S. Fidell and B. G. Tabachnick. Preparatory Data Analysis. In *Handbook of Psychology*, pp. 115–141. John Wiley & Sons, Ltd, 2003. 7

[18] B. Froehlich, J. Hochstrate, V. Skuk, and A. Huckauf. The globefish and the globemouse: two new six degree of freedom input devices for graphics applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, p. 191–199. Association for Computing Machinery, New York, NY, USA, 2006. 2

[19] M. Frutos-Pascual, C. Creed, and I. Williams. Head Mounted Display Interaction Evaluation: Manipulating Virtual Objects in Augmented Reality. In *Human-Computer Interaction – INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part IV*, pp. 287–308. Springer-Verlag, Berlin, Heidelberg, Sept. 2019. 3

[20] D. Glesser, F. Bérard, and J. R. Cooperstock. Overcoming limitations of the trackpad for 3d docking operations. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, p. 1239–1244. Association for Computing Machinery, New York, NY, USA, 2013. 2

[21] D. G. Green, M. K. Powers, and M. S. Banks. Depth of focus, eye size and visual acuity. *Vision research*, 20(10):827–835, 1980. 2

[22] T. Ha and W. Woo. An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment. In *2010 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 91–98, Mar. 2010. 2

[23] K. Hinckley, R. Pausch, D. Proffitt, and N. F. Kassell. Two-handed virtual manipulation. *ACM Trans. Comput.-Hum. Interact.*, 5(3):260–302, Sept. 1998. 2, 3

[24] D. A. Hofmann. An overview of the logic and rationale of hierarchical linear models. *Journal of Management*, 23(6):723–744, Jan. 1997. 3, 7

[25] W.-j. Hou and X.-l. Chen. Comparison of eye-based and controller-based selection in virtual reality. *International Journal of Human–Computer Interaction*, 37(5):484–495, 2021. 9

[26] W.-A. Hsieh, H.-Y. Chien, D. Brickler, S. V. Babu, and J.-H. Chuang. Comparing and contrasting near-field, object space, and a novel hybrid interaction technique for distant object manipulation in vr. *Virtual Worlds*, 3(1):94–114, 2024. 2

[27] R. J. K. Jacob. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Trans. Inf. Syst.*, 9(2):152–169, Apr. 1991. 3

[28] J. Jankowski and M. Hachet. A Survey of Interaction Techniques for Interactive 3D Environments. In M. Sbert and L. Szirmay-Kalos, eds., *Eurographics 2013 - State of the Art Reports*. The Eurographics Association, 2013. 2

[29] J. A. Jones, J. E. Swan, G. Singh, and S. R. Ellis. Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, APGV '11, p. 29–36. Association for Computing Machinery, New York, NY, USA, 2011. 1, 2

[30] E. Kaiser, A. Olwal, D. McGee, H. Benko, A. Corradini, X. Li, P. Cohen, and S. Feiner. Mutual disambiguation of 3D multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th international conference on Multimodal interfaces*, ICMI '03, pp. 12–19. Association for Computing Machinery, New York, NY, USA, Nov. 2003. 2, 3

[31] H. J. Kang, J.-h. Shin, and K. Ponto. A Comparative Analysis of 3D User Interaction: How to Move Virtual Objects in Mixed Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 275–284, Mar. 2020. ISSN: 2642-5254. 1, 2, 3

[32] S. Kim and G. Lee. Evaluating an In-Hand Ball-Shaped Controller for Object Manipulation in Virtual Reality. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 10–19, Mar. 2024. ISSN: 2642-5254. 4

[33] R. Kumar, T. Oskiper, O. Naroditsky, S. Samarasekera, Z. Zhu, and J. Kim. System and method for generating a mixed reality environment, Mar. 2017. 1

[34] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–14. Association for Computing Machinery, New York, NY, USA, Apr. 2018. 3

[35] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, Dec. 1982. 7

[36] J. R. Lewis. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human–Computer Interaction*, 7(1):57–78, Jan. 1995. 6

[37] K. Li, R. Nataraj, T. L. Marquardt, and Z.-M. Li. Directional Coordination of Thumb and Finger Forces during Precision Pinch. *PLOS ONE*,

8(11):e79400, Nov. 2013. Publisher: Public Library of Science. 2

[38] M. J. Liberatore and W. P. Wagner. Virtual, mixed, and augmented reality: a systematic review for immersive systems research. *Virtual Reality*, 25(3):773–799, Sept. 2021. 1

[39] G. Lu, L.-K. Shark, G. Hall, and U. Zeshan. Immersive manipulation of virtual objects through glove-based hand gesture interaction. *Virtual Reality*, 16(3):243–252, Sept. 2012. 2, 3

[40] F. L. Luro and V. Sundstedt. A comparative study of eye tracking and hand controller for aiming tasks in virtual reality. In *Proceedings of the 11th ACM Symposium on eye tracking research & applications*, pp. 1–9, 2019. 9

[41] P. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55, 1936. 7

[42] A. Masurovsky, P. Chojecki, D. Runde, M. Lafci, D. Przewozny, and M. Gaebler. Controller-free hand tracking for grab-and-place tasks in immersive virtual reality: Design elements and their empirical study. *Multimodal Technologies and Interaction*, 4(4), 2020. 2, 8

[43] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. *Computer Graphics Forum*, 38(1):21–45, 2019. 2

[44] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge. The benefits of dof separation in mid-air 3d object manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, p. 261–268. Association for Computing Machinery, New York, NY, USA, 2016. 4

[45] A. K. Mutasim, A. U. Batmaz, and W. Stuerzlinger. Pinch, Click, or Dwell: Comparing Different Selection Techniques for Eye-Gaze-Based Pointing in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '21 Short Papers, pp. 1–7. Association for Computing Machinery, New York, NY, USA, May 2021. 3

[46] R. C. Oldfield. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1):97–113, Mar. 1971. 6

[47] J. Orlosky, C. Liu, K. Sakamoto, L. Sidenmark, and A. Mansour. EyeShadows: Peripheral Virtual Copies for Rapid Gaze Selection and Interaction. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 681–689, Mar. 2024. ISSN: 2642-5254. 3

[48] Y. S. Pai, T. Dingler, and K. Kunze. Assessing hands-free interactions for VR using eye gaze and electromyography. *Virtual Reality*, 23(2):119–131, June 2019. 2

[49] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze + pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, pp. 99–108. Association for Computing Machinery, New York, NY, USA, Oct. 2017. 2, 3

[50] W. Piekarski and B. H. Thomas. Interactive augmented reality techniques for construction at a distance of 3D geometry. In *Proceedings of the workshop on Virtual environments 2003*, EGVE '03, pp. 19–28. Association for Computing Machinery, New York, NY, USA, May 2003. 3

[51] J. S. Pierce and R. Pausch. Comparing voodoo dolls and homer: exploring the importance of feedback in virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, p. 105–112. Association for Computing Machinery, New York, NY, USA, 2002. 3

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

[53] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn. User-defined gestures for augmented reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pp. 955–960. Association for Computing Machinery, New York, NY, USA, Apr. 2013. 2, 3

[54] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, UIST '96, pp. 79–80. Association for Computing Machinery, New York, NY, USA, Nov. 1996. 3

[55] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst. Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. *Computer Graphics Forum*, 17(3):41–52, 1998. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8659.00252. 3

[56] A. Robb, K. Kohm, and J. Porter. Experience Matters: Longitudinal Changes in Sensitivity to Rotational Gains in Virtual Reality. *ACM Trans. Appl. Percept.*, 19(4):16:1–16:18, Nov. 2022. 3

[57] L. E. Sibert and R. J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '00, pp. 281–288. Association for Computing Machinery, New York, NY, USA, Apr. 2000. 3

[58] T. Snijders and R. Bosker. *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. SAGE Publications Inc., 2nd (1st edition 1999) ed., 2011. 7

[59] M. Speicher, B. D. Hall, and M. Nebeling. What is Mixed Reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 1–15. Association for Computing Machinery, New York, NY, USA, May 2019. 2

[60] S. Stellmach and R. Dachselt. Look & touch: gaze-supported target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 2981–2990. Association for Computing Machinery, New York, NY, USA, May 2012. 3

[61] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pp. 265–272. ACM Press/Addison-Wesley Publishing Co., USA, May 1995. 2, 3

[62] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 13(3):429–442, 2007. 1, 2

[63] E. Velloso, J. Turner, J. Alexander, A. Bulling, and H. Gellersen. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2015*, pp. 315–330. Springer International Publishing, Cham, 2015. 3

[64] R. Venkatakrishnan, R. Venkatakrishnan, R. Canales, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. Investigating the effects of avatarization and interaction techniques on near-field mixed reality interactions with physical components. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2, 6

[65] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. Give me a hand: Improving the effectiveness of near-field augmented reality interactions by avatarizing users' end effectors. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2412–2422, 2023. 2

[66] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. How virtual hand representations affect the perceptions of dynamic affordances in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2258–2268, 2023. 2

[67] V. Vuibert, W. Stuerzlinger, and J. R. Cooperstock. Evaluation of docking task performance using mid-air interaction techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, SUI '15, p. 44–52. Association for Computing Machinery, New York, NY, USA, 2015. 2

[68] L. Wang, J. Wu, X. Yang, and V. Popescu. VR Exploration Assistance through Automatic Occlusion Removal. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2083–2092, May 2019. Conference Name: IEEE Transactions on Visualization and Computer Graphics. 3

[69] M. Whitlock, E. Harnner, J. R. Brubaker, S. Kane, and D. A. Szafir. Interacting with Distant Objects in Augmented Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 41–48, Mar. 2018. 1, 3

[70] J. Wither and T. Hollerer. Evaluating techniques for interaction at a distance. In *Eighth International Symposium on Wearable Computers*, vol. 1, pp. 124–127, Oct. 2004. ISSN: 1530-0811. 3

[71] F. Zhu, L. Sidenmark, M. Sousa, and T. Grossman. Pinchlens: Applying spatial magnification and adaptive control-display gain for precise selection in virtual reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1221–1230, 2023. 3