A Comparative Analysis of 3D User Interaction: How to Move Virtual Objects in Mixed Reality

Hyo Jeong Kang*
University of Florida

Jung-hye Shin†
University of Wisconsin-Madison

Kevin Ponto‡
University of Wisconsin-Madison

Figure 1: Example scenes of three interaction techniques

(a) Gaze and Pinch Interaction
(b) Direct Touch and Grab
(c) Worlds-in-Miniature (WIM)

ABSTRACT

Using one’s hands can be a natural and intuitive method for interacting with 3D objects in a mixed reality environment. This study explores three hand-interaction techniques, including the gaze and pinch, touch and grab, and worlds-in-miniature interaction for selecting and moving virtual furniture in the 3D scene. Overall, a comparative analysis reveals that the worlds-in-miniature provided the best usability and task performance than other studied techniques. We also conducted in-depth interviews and analyzed participants’ hand gestures in order to identify desired attributes for 3D hand interaction design. Findings from interviews suggest that, when it comes to enjoyment and discoverability, users prefer directly manipulating the virtual furniture to interacting with objects remotely or using in-direct interactions such as gaze. Another insight this study provides is the critical roles of the virtual object’s visual appearance in designing natural hand interaction. Gesture analysis reveals that shapes of furniture, as well as its perceived features such as weight, largely determined the participant’s instinctive form of hand interaction (i.e., lift, grab, push). Based on these findings, we present design suggestions that can aid 3D interaction designers to develop a natural and intuitive hand interaction for mixed reality.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

1 INTRODUCTION

The use of hand input has received increasing attention in the virtual reality (VR) community. There is a growing body of research that has begun to cast light on hand-interaction; yet, many prior VR research predominantly focused on 3D interaction using controllers. Using bare-hands offers a distinctive user experience as opposed to using controllers. Compared to the controller where a single button is often mapped to a specific action, human hands are capable of various gestures by nature. For example, to move a virtual object, the standard design practice for VR controllers is to hold a grip button. The dynamic nature of human hands, however, makes it challenging to find and pin down the most natural hand input among possible gestures—lift, grab, push, point, pinch, and more. The fundamental difference lies in our familiarity. Unlike a controller that typically requires extra efforts of learning, we are more likely to adopt our habitual actions of using hands when interacting with the virtual object. Consequently, identifying the most natural hand interaction is a complicated process since the idea of “natural” hand interaction can vary depends on the context, and it requires an understanding of how we interact with objects in everyday life.

Prior 3DUI research has developed the novel metaphors for hand gestures such as finger-pointing that works as a mouse cursor [10,15,27,39], a multimodal interaction that integrates finger-pointing and gaze [21,32,42], and multiple gestures such as grabbing and pinching for complex tasks [16,20]. While these streams of research provide valuable insights in designing innovative hand interactions, relatively little is known on tradeoff among existing technologies. Furthermore, the majority of prior works are focused on the usability aspects of hand input. For instance, 3DUI research abounds with discussions on efficient and accurate hand inputs that address technical concerns such as object occlusions and small field of view [13,17,21,32]. The question of which hand gesture would render a natural interaction, on the other hand, remains less explored.

To this end, this study aims to provide a comprehensive review of the tradeoff among the well-known 3D hand interaction designs. We compare three 3D interactions: 1) multimodal interaction using gaze and pinch gesture, 2) direct touch and grab interaction, and 3) worlds-in-miniature. In comparing these three interactions, we first observed participants’ behavioral responses without giving instructions. This approach helps us to examine the discoverability of each interaction design, thereby providing insights into natural hand interaction. We particularly focused on an interaction that involves selecting and arranging furniture items. Architecture and interior design are one of the most promising application areas in MR; thus findings from this study can provide immediate real-world implications. In comparing these three interactions, we address the following research questions.

• Discoverability: which interaction design requires a greater time investment to understand the mechanism without instructions? In addition to three studied interactions, are there other gestures participants try to use?

• Accuracy and Efficiency: which interaction mechanism offers
the user an increased efficiency and accuracy to complete the task?

2.1 Naturalism and Discoverability in 3D Interaction

The definition of the term naturalism varies widely depending on the field of study. In 3DUI research, the term describes user interaction that mimics the counterparts of the real-world interaction [5]. Discoverability, coined by Don Norman [25], is a measure of how easily users can find and understand the new interface with the minimum level of instruction. The previous human-computer interaction and VR research show that the major benefits of discoverability include minimizing a learning curve [5], increasing user satisfaction [33], and enhancing users’ intention to use the system [8].

Discoverability and naturalism are closely related concepts; however, the mere fact that interaction is easily discoverable would not always mean that the interaction is natural. For example, using a swipe gesture is not how we navigate pages in the real world, but most smartphone users can easily discover the swiping feature without instructions. In contrast, a user’s prior knowledge and mental model of how the system should work [25]. As such, examining the discoverability of existing interfaces can be a key first step toward understanding users’ expectation and belief system about how hand-based mixed reality interaction should work. In this paper, we investigate naturalism and discoverability of interaction design by observing users’ behavioral responses. An observational study with existing systems can help designers fill the gap between users’ mental models and existing interaction methods.

2.2 Three Existing Techniques

This study compares three 3D interaction techniques: 1) multimodal interaction using gaze and pinch gesture, 2) direct touch and grab interaction, and 3) worlds-in-miniature. These three interactions were chosen as the subject of the study based on their popularity as well as the comparative advantage each interaction provides.

First, a gaze and pinch interaction is the most widely used interaction for mixed reality headsets such as a Microsoft Hololens. This multimodal interaction integrates eye gaze or head pose with a gesture, enabling users to interact with the remote object. Using a gaze input allows users to avoid false occlusion cues when hand tracking is not available, and it also deals with the constraints of a limited field of view. With a gaze and gesture interaction, users can select the target object with their head movement or eye gaze and move the object by holding a pinch gesture.

Second, a direct touch and grab interaction is a default interaction mechanism for many VR interfaces. The classical design approach for VR controllers (e.g., Oculus Touch, Vive) is using grip button for grabbing the object. When using hands, users can utilize their natural hands’ movements by directly touching an object with their bare hands and physically moving the object with arms’ movement. The relative merits and demerits of these two interaction techniques make it worthwhile to compare a gaze and pinch and a direct touch and grab techniques. While being considered as the most natural, the primary issue with a direct touch-grab interaction is that users can only interact with objects within their arms’ reach. A gaze and pinch interaction can solve such issues as it enables users to interact with the remote object; however, a gaze-based interaction may seem less natural or intuitive.

Worlds-in-miniature (WIM) interaction entails the advantages of both a gaze-pinch and a direct grab-touch interaction [37]. While still allowing users to have a natural and direct interaction, the WIM provides users miniature copies of the actual objects so that users can interact with the remote object indirectly. There are other 3DUI techniques, such as go-go interaction [30] or ray-casting technique [3], that also support direct manipulation from a distance; yet, we examined the worlds-in-miniature given its growing popularity in VR home design application (e.g., TrueScale, Home VR).

2.3 Comparative Studies on 3D User Interaction

A 3D interaction technique refers to an interaction mechanism on how users can execute different types of tasks in the 3D environment [4]. In this study, we mainly address selection and manipulation tasks. Selection refers to the task of picking one or more objects from the environment and manipulation refers to the task of changing a virtual object’s orientation and position [5].

In evaluating 3D user interaction, accuracy and efficiency have been a fertile research topic in the previous HCI studies. As early as in 1999, Bowman and Hodges [2] compared six different 3D manipulation techniques to find a more accurate and efficient way to interact with a remote object in a VR environment. In their comparative study, a conventional raycasting offered more precise control than using a virtual arm. More recently, Kytö et al. [21] examined selection techniques for the head-worn AR device using a Microsoft Hololens. Their study findings indicated that using gaze was generally faster than a hand-gesture or a hand-held device. In a similar vein, research reported gaze input being faster than hand-pointing on a large stereoscopic projected display [35]. Perceived usability is another critical dimension in evaluating 3D user interaction techniques. Bernardos, Gómez, and Casar [1] compared the head-directed selection to the hand-based pointing on a projection screen. In their study, users found hand-based pointing far more intuitive and usable than head-directed selection. McMahan et al. [23] compared three interaction techniques on the projected wall and found that hand-based object manipulation more natural than using controllers.

In summary, much of prior research showed that more naturalistic interaction—using hands rather than gaze or ray-casting—provides better usability. With respect to accuracy and efficiency, however, several works indicated that perceived naturalism is not a necessary component for efficient and precise task performance. While these prior study findings provide insights into evaluating existing 3D user interactions, the differences across experiment settings preclude any generalizable conclusion. Prior works on 3D user interaction have predominantly focused on interaction with a projected wall and head-mounted VR displays, and relatively little is known for wearable Mixed Reality displays context. Furthermore, most of the previous comparative studies tend to focus on the specific interaction mechanism for a single task, for example, comparing head tracking to eye tracking for a selection task [21], different ray-casting mechanisms for a pointing task [7, 9], different finger-based projection techniques for a manipulation task [22].

In this study, we aim to provide a more comprehensive understanding of 3D user interaction by examining three interaction techniques in accomplishing series of tasks—selecting a menu, selecting an object, and positioning the object. We further examined the discoverability of three interaction design by observing participants’ hand gestures. Contribution of this study lies in providing a broader
In order to create a mixed reality environment, we used a Leap Motion for hand tracking. Leap Motion uses image processing with infrared stereoscopic video [36]. Unlike other gesture recognition devices such as Kinect, a Leap Motion does not return a complete depth map but only a set of relevant hand points and some hand pose features; thus, we used preset hand pose features of Leap motion—grabbing, pinching, and pointing. Virtual hand was always rendered in the scene letting users know their hands are tracked well. Leap Motion was attached to the user’s left hand and users could evoke the menu as they flip their left hand (seen in Figure 2, middle). We used the term “gaze” instead of head position, which is more accurate, because gaze is more generally accepted word in describing a multimodal interaction [21]. Three tasks are implemented in the following ways: 1) Menu: a menu was always displayed in front of the users at the distance of 2 meters. The gaze cursor moves along with the user’s head movement. When the cursor was pointing to the object, the object got highlighted with blue color. 2) Selection: selection are triggered explicitly using gaze. The gaze cursor moves along with the user’s head movement. When the cursor was pointing to the object, the object got highlighted with blue color. 2) Selection: selection are triggered explicitly using gaze. The gaze cursor moves along with the users’ head movement. When the gaze cursor is pointing to the object, the object gets highlighted with blue color. 3) Manipulation: users could move the furniture by holding a pinch gesture. The virtual hand moved along with the hand’s movement as seen in Figure 2 (top).

### 3.3 Direct Touch and Grab Interaction

The grab and touch interaction uses head-movement as a proxy to eye gaze. We used the term “gaze” instead of head position, which is more accurate, because gaze is more generally accepted word in describing a multimodal interaction [21]. Three tasks are implemented in the following ways: 1) Menu: the attached menu metaphor was used. A menu was attached to the user’s left hand and users could evoke the menu as they flip their left hand (seen in Figure 2, middle). We used this attached menu metaphor for two reasons: first, this menu selection mechanism corresponds well to direct touch and grab interaction because the menu should be directly touched to be invoked. Second, this attached menu minimizes unnecessary physical movement involving menu selection, such as walking toward the menu, thus enables better comparisons with other conditions. 2) Selection: user could select the menu by directly touching the button with their index finger. As users select the virtual furniture, the furniture item appeared in front of the users, within the range between 0.6 and 0.8 meter. This range reflects the average length of human arms. When it comes to object selection, users could select the object by directly touching it with their bare hands then could grab it by making a fist. 3) Manipulation: users could move the object with their arms’ movement. Users had to physically walked around the room in order to place the virtual furniture.

### 3.4 Worlds-in-Miniature Interaction

The last interaction is a world-in-miniature (WIM) metaphor [37]. Unlike the previous two interactions where users can directly select and move the actual-scale object, WIM interaction provides users the hand-held miniature representation of the objects. In order to create the miniature floor, users were first asked to scan the surrounding environment. Using ZED mini’s depth-camera, the scanned room environment was brought in to the miniature environment. After the generation of the miniature floor, three tasks are implemented in the following ways: 1) Menu: menu was displayed in front of the users at the distance of 0.6 - 0.8 meter, which reflects the reachable length for average human. 2) Selection: As users select the menu by pressing the button with their index finger, the miniature virtual objects appeared. 3) Manipulation: By moving the miniature object, users can move the corresponding actual-size object as seen in Figure 2 (bottom). Users were able to select and move the miniature object by directly manipulating it with touch and grab interaction.

### 4 USER STUDY

#### 4.1 Participants

21 participants were recruited through mailing lists and flyers in the campus. Participants’ age ranged from 18 to 44 with a mean of 24.2 (SD = 7.8) and the majority of them were university students. We specifically sought participants who had no prior experience with head-worn mixed reality devices and hand tracking devices. None of the participants had previous experience with mixed reality headsets or hand tracking device, but nine participants had prior experience with head-worn VR devices such as an Oculus Rift or HTC Vive. On average, participants spend a significant amount of time working on a computer on a daily basis (3-8 hours a day: 47%, more than 8 hours: 28%).

![Figure 2: Example scenes: Gaze and pinch (top), Direct touch and grab (middle), and worlds-in-miniature (bottom).](image)
4.2 Tasks

With each interaction condition, participants were asked to complete three tasks: 1) select the menu to choose a virtual object; 2) move the object to the designated area, and 3) arrange multiple furniture items. Figure 3 shows an overview of the three tasks and the experiment procedure. A virtual object always appeared within a reachable area from the user, a range between 0.6 to 0.8 meter, which is average length of human arm.

The primary purpose of task 1 and task 2 is to examine discoverability each interaction technique. Thus, a minimum level of instruction was provided to participants. For example, for gaze-pinch interaction, participants were told to "use a hand-gesture to move the object," and no further information about what gesture users should use was provided. For touch-grab interaction, participants were told to "use hands to move objects." No further information about how participants can move the virtual furniture was provided. For a worlds-in-miniature interaction, participants had a same instruction with touch-grab interaction--"use hands to move objects."

After completing task 1 and 2, detailed instructions of how to interact with the virtual objects were provided before the task 3. The goal of task 3 is to measure the accuracy and efficiency of each interaction. After a brief demonstration and trials, participants were asked to select two red stools, two black stools, and one table, then arrange the four chairs around the table so that each chair is perpendicular to each other. The total time it took to complete the task was measured as an indicator of efficiency. The angle between each chair was measured as an indicator of accuracy. After the completion of all three tasks, participants evaluated the interaction design. These three tasks are repeated for 3 interactions.

4.3 Experiment Procedure

In order to examine participants' preference and compare three techniques, this study employed a within-subject study design. Each interaction condition was fully counter-balanced to minimize the learning effect. Participants experienced all three interactions in a randomized order and were asked to pick one interaction design they preferred the most. Participants evaluated the interaction by answering questionnaires. Finally, they were asked to complete the final survey that includes questions on demographic information and their average computer usages. After the completion of the survey, participants were invited for a follow-up interview. The entire process approximately took 45 minutes.

4.4 Measures

4.4.1 Evaluation of three interaction designs

Discoverability: Discoverability was examined by measuring the time it took to complete the task 1 and task 2 [29]. In our pilot study with 7 graduate students, we asked them to notify us when they could not figure out the interaction and wanted to give up the tasks. The pilot study indicated that people feel a great level of frustration when they failed to interact with the virtual objects after 1.5 minute. Based on this result, when participants could not figure out the interaction for longer than 2 minutes, we marked the task as "failed" then asked participants to proceed to the next task.

Efficiency and Accuracy: The efficiency of interaction design was examined by measuring the time it took to complete the task 3. The accuracy of interaction design was examined using position data of five furniture items: four stools, and one table. Participants were asked to arrange four stools so that they are perpendicular to each other. The average deviation from 90-degree was calculated to examine accuracy.

Usability: The perceived usability of each interaction design was evaluated using the System Usability Scale (SUS) [6]. The system usability scale includes 10 items with five responses that range from strongly agree to strongly disagree. The example questionnaire includes: "I found the system was easy to use," and "I would imagine that most people would learn to use this system very quickly." In order to examine perceived task loads, the NASA Task Load Index (NASA-TLX) questionnaire was also included in the survey [14].

Naturalism and Enjoyment: To measure the perceived level of naturalism, we used five-point Likert scale questionnaires. The questionnaire includes three items: "interaction with a virtual object was similar to how I would interact with an object in a real-world," "interaction with a virtual object felt familiar," "interaction with a virtual object seems intuitive and natural." The questionnaire for the enjoyment was adopted from previous literature on extended Technology Acceptance Model [11]. The enjoyment questionnaire includes two items: "It was enjoyable to use this system," "It was fun to use this system." Cronbach’s alpha for both variables was above 0.7 (αnaturalism = 0.78; αenjoyment = 0.81).

Control variables: Responsiveness of the system was included as a control variable. The questionnaire asked participants whether they experienced system lag when they experience each interaction design. Participants’ demographic information, as well as their previous experience with VR, video game, and average daily computer use, were also measured.

4.4.2 Gesture analysis and follow-up interview

Gestures: Studying gesture can help 3D UI designers to find suitable interface features that correspond well to users' expectation [31]. Following Kendon [19]'s gesture analysis method, the session was video-recorded and changes of gestures were encoded with a description of the movement change. The purpose of this analysis is to identify gestures participants use when they first encounter the virtual object, thus the video recordings of task 1 and task 2 sessions were mainly used for the analysis. The hand movement is described with three main elements: a number of hands used, specific gesture used, and a target object. The examples of description include: pressing the button with an index finger, lifting the table using both hands, pushing the chair with one hand, and gripping the top of the chair.

User experience: In order to better understand participants’ experience with three interaction designs, a follow-up interview was conducted. The interview questions mainly include five items: 1)
which interaction did you prefer the most? 2) why do you prefer interaction X over others? 3) what do you like or dislike about other interactions? 4) What’s the hardest part about the interaction you experienced? 5) How would you like to provide a solution to make an improvement? In addition to these five questions, we also asked questions based on our observation, for instance, what specific gesture participants used and why.

5 RESULTS

5.1 Evaluation of Three Interaction Designs

5.1.1 Discoverability

In order to examine how easy each interaction is to discover, the time it took for participants to complete the first task (menu selection) and the second task (object selection and manipulation) was compared. ANOVA analysis shows a significant main effect for both tasks (menu selection: $F(2, 51) = 11.13$, $p < .001$, $\eta^2_{p} = .303$; object selection and manipulation: $F(2, 50) = 9.67$, $p < .001$, $\eta^2_{p} = .283$).

Tukey’s HSD post hoc test was conducted for pair-wise comparisons. The result of the analysis is shown in Figure 4.

For the menu selection task, gaze-pinch interaction took the longest amount of time to complete the menu selection ($M_{gaze-pinch} = 54.88$, $SD_{gaze-pinch} = 20.47$; $M_{touch-grab} = 42.83$, $SD_{touch-grab} = 25.57$; $M_{WIM} = 23.69$, $SD_{WIM} = 12.27$). With worlds-in-miniature interaction, it took significantly less time to select the menu than gaze-pinch interaction ($p < .001$) and touch-grab interaction ($p < .05$).

For the object selection and manipulation task, gaze-pinch interaction took a significantly longer amount of time to complete the task ($M_{gaze-pinch} = 49.20$, $SD_{gaze-pinch} = 38.33$; $M_{touch-grab} = 21.57$, $SD_{touch-grab} = 14.50$; $M_{WIM} = 15.92$, $SD_{WIM} = 11.04$). Six participants failed to figure out gaze and pinch interaction within 2 minutes with the gaze-pinch interaction. There was no significant difference between grab-touch interaction and worlds-in-miniature interaction ($p = .76$).

Figure 4: Discoverability (top), Accuracy and Efficiency (bottom)

5.1.2 Accuracy and Efficiency

The object manipulation task involved arranging four stools around the table so that each stool is perpendicular to each other. Accuracy of manipulation was calculated by creating vectors of stools across from each other and comparing the angle between these two vectors to that of a right angle. Mean of absolute angle was used to represent the accuracy of each interaction design. ANOVA analysis shows a significant main effect on accuracy for different interaction designs ($F(2, 54) = 8.58$, $p < .001$, $\eta^2_{p} = .241$). The result of the analysis is shown in Figure 5.

The pair-wise comparison results show that mean manipulation error was significantly lower for worlds-in-miniature interaction design than other two interaction techniques ($M_{gaze-pinch} = 20.36$, $SD_{gaze-pinch} = 7.31$; $M_{touch-grab} = 18.42$, $SD_{touch-grab} = 6.85$; $M_{WIM} = 12.01$, $SD_{WIM} = 5.16$). The efficiency of each interaction was measured with the total time it took to finish the object selection and manipulation task (task3). ANOVA analysis shows a significant main effect on efficiency for different interaction designs ($F(2, 54) = 29.23$, $p < .001$, $\eta^2_{p} = .51$). The Tukey’s HSD pair-wise comparison test shows that worlds-in-miniature was significantly faster to complete the task than the other two interaction techniques ($M_{gaze-pinch} = 5.90$, $SD_{gaze-pinch} = 2.08$; $M_{touch-grab} = 3.98$, $SD_{touch-grab} = 1.56$; $M_{WIM} = 2.44$, $SD_{WIM} = 0.51$).

5.1.3 Usability, Naturalism, and Enjoyment

Usability of the interface was measured with two indexes: the NASA-Task Load Index (NASA-TLX) and System Usability Scale (SUS). The ANOVA results for NASA-TLX is shown in Figure 6. There was a significant main effect for mental load ($F(2, 53) = 10.04$, $p < .001$, $\eta^2_{p} = .27$), physical load ($F(2, 54) = 4.84$, $p < .05$, $\eta^2_{p} = .15$), performance ($F(2, 54) = 6.97$, $p < .01$, $\eta^2_{p} = .20$), effort
We analyzed video recordings of participants’ hand movements by systematically categorizing and labeling gesture they used. Due to the video recording error, videos of 17 participants were used for the analysis. Table 1 shows the summary of findings from gesture analysis. When participants first encountered the gaze-pinch interface, in which the menu was displayed far away from the user, the first action all participants tried was to press the menu button with an index finger. When pressing gesture failed to evoke the menu interaction (p < .001), and frustration (F(2, 54) = 23.09, p < .001; \(\tau_p^2 = .46\)). The results from pair-wise comparisons between three interaction designs indicated that the gaze-pinch interaction are significantly more demanding. The gaze and pinch interaction required a significantly higher level of effort than touch-grab interaction (\(p < .05\)) and the worlds-in-miniature interaction (\(p < .001\)).

The significant main effects were reported for system usability (F(2, 54) = 20.13, p < .001; \(\tau_p^2 = .43\) and perceived naturalism (F(2, 51) = 13.35, p < .001; \(\tau_p^2 = .34\), yet no significant effect was found for enjoyment (F(2, 54) = 2.41, p = .09; \(\tau_p^2 = .08\)). The pairwise Tukey HSD test shows that the worlds-in-miniature interaction provides significantly better usability than other two interactions (\(M_{gaze-pinch} = 45.97, SD_{gaze-pinch} = 18.39; M_{touch-grab} = 57.61, SD_{touch-grab} = 19.22; M_{WIM} = 81.31, SD_{WIM} = 13.92\)). In terms of perceived level of naturalism, both touch-grab and worlds-in-miniature interaction were rated to be significantly more natural than gaze-pinch interaction (\(M_{gaze-pinch} = 2.25, SD_{gaze-pinch} = 0.92; M_{touch-grab} = 3.77, SD_{touch-grab} = 0.97; M_{WIM} = 3.41, SD_{WIM} = 0.87\) as shown in Figure 6.

5.2 Gesture Analysis and Follow-up Interview

5.2.1 Gesture Analysis

We analyzed video recordings of participants’ hand movements by systematically categorizing and labeling gesture they used. Due to the video recording error, videos of 17 participants were used for the analysis. Table 1 shows the summary of findings from gesture analysis. When participants first encountered the gaze-pinch interface, in which the menu was displayed far away from the user, the first action all participants tried was to press the menu button with an index finger. When pressing gesture failed to evoke the menu interaction, 14 participants walked or leaned toward the menu and tried to press the button again using an entire hand. The next action participants tried was a grabbing gesture. In the follow-up interview, many participants noted that they would not be able to figure out the pinch gesture unless they were provided with the instruction.

When participants first encountered the large-scale furniture items, they preferred using both hands when moving a large object. Interview with participants reveal that perceived scale objects such as a table. On the other hand, for the miniature size object, participants frequently used the grabbing gesture to pick up the virtual object. The majority of participants used only one hand when they were interacting with a miniature size object.

5.2.2 Follow-Up Interview: User Experience

The follow-up interviews reveal the strength and weakness of each interaction techniques in greater depth. The majority of participants preferred worlds-in-miniature interaction over the other two for ease of use. Participants frequently described the worlds-in-miniature interaction as the “easiest” interaction. Although there was no statistically significant differences reported on the level of enjoyment, about one third of participants indicated that they would enjoy directly interacting with real-size object rather than manipulating a miniature object. Furthermore, more than half of interviewees said they preferred walking around the room to standing still when interacting with virtual objects.

When it comes to discoverability and ease of use, Gaze and pinch interaction was participants’ least favorite design amongst three. Many noted that gaze and pinch interaction was difficult to figure out without instruction. Three interviewees mentioned that they did not notice the gaze cursor nor understood the mechanism of gaze input. While majority of participants prefer other two interaction designs, two participants said that they would like to use gaze and pinch interaction in the future because it allows them easier interaction once they learn how to use it. These two participants indicated that they especially liked gaze-pinch interaction as it requires less physical movement and users can still directly interact with a life-scale object instead of moving miniature items.

“Even if it’s not real, you think it’s furniture. It feels like it should be heavy, and I should be using both hands”

More than 80 percent of participants used both hands trying to lift a table or chairs. Interview with participants reveal that perceived feature of furniture such as its weight influence their instinctive form of hand interactions. Even after we provided the instruction and participants learn that they can move the object using only one hand, they preferred using both hands when moving a large object.

Participants also shared their opinions on desired attributes for interaction design and made suggestions on how to make an improvement. Frequently mentioned topics include 1) dynamic interactions, 2) natural interaction, and 3) direct interaction. Based on these findings, we discuss design recommendation in the remaining of this paper. Figure 7 provides a summary of the follow-up interviews with sample verbatim.
6 Discussion

This study examines hand interaction techniques for a head-worn MR device. The primary goal of this study is to understand how to design 3D user interaction that is intuitive and easy to learn. Through a series of user studies, we compared three widely known interaction techniques: gaze and pinch interaction, touch and grab interaction, and worlds-in-miniature. The finding from a user study suggests two key implications.

First, worlds-in-miniature interaction stands out amongst three for providing better usability. Compared to other two studied interactions, worlds-in-miniature interaction allowed faster and more accurate interaction in manipulating multiple objects. Unlike the gaze-pinchoch and touch-grab interaction, that only provide perspective views of the objects, having an additional elevated view from above makes it easier to see the entire layout and understand spatial configuration [34].

Second, the findings from the gesture analysis suggest the critical role of the visual appearance of a virtual object or, more specifically, the affordance of an object in enhancing discoverability [25]. That is, a visual characteristic of a virtual menu which makes it looks "pushable" and the shape of a virtual table that affords "lifting" largely determined participants’ form of interaction. The perceptual property an object such as weight influences how users would interact with the object. This is well represented in one of the participant’s comment: "It’s furniture. It feels like it should be heavy."

7 Design Implications

7.1 Natural Interaction is Dynamic and Complex

Designers often strive to create a simple design. Especially for an innovative interface such as mixed reality system, one of the common design approach is to find a one single input system that can work across different platforms [26]. Assigning specific interaction for each input system—for example, simulating mouse movement with a gaze or using finger pointing as a mouse cursor—has been a classical design approach. This one-to-one mapping approach may be easier to implement; however, such an interaction mechanism is difficult to figure out without instructions.

Natural interaction is complex and dynamic. In our in-depth interviews, when we asked participants to choose a single interaction they like the most, many indicated that they want to combine all three interactions rather than picking one. The desire for dynamic interaction is well expressed in one of the participants comments—"I hope there are more you can do. Like you can pull, push, or lift, basically do everything you would do with the real chair and table.”

7.2 Natural Interaction is Direct and Physical

Natural interaction is direct and physical, which means it involves "touching" objects. The mid-air gestures has gained its popularity with the release of motion sensing devices [40]. Research defines mid-air gesture as an interaction that involves touchless manipulation of digital content [38]. The term itself presumes the in-direct interaction. In our user study, participants frequently noted that...
The focus of our study was discoverability. Discoverability can be virtually touch the object rather than using a mid-air gesture, can improve possible interaction with users. The bounding box and app bar in the Microsoft Hololens 2. The bounding box shown in the Hololens 2 informs the user that the object interaction. The great example of visual guidance comes from the games industry, as visually presenting the possible forms of gestures and interaction and computationally expensive. Providing interaction guidance such as making an object affordable can allow direct interaction with the object. Providing digital buttons and icons within the reachable distance, so that users can virtually touch the object rather than using a mid-air gesture, can simulate more realistic interaction.

7.3 Affordance and visual guidance
Coining the concept of affordance has become the core design principle in interaction design. Our gesture analysis confirmed that understanding a virtual object’s affordance is vital for developing a more natural and intuitive 3D user interaction. Depending on the virtual object’s visual appearance, the user’s form of interaction can largely vary. A button should be pushed, a table should be lifted, and a miniature object should afford to be picked up. Designers can take into account visual affordance and create more realistic simulations by developing physics-based interaction. However, creating an accurate mesh colliders for individual objects and simulating realistic collision detection can often be time-consuming and computationally expensive. Providing interaction guidance such as visually presenting the possible forms of gestures and interaction can be a simple solution that can help users to quickly learn the interaction. The great example of visual guidance comes from the bounding box and app bar in the Microsoft Hololens 2. The bounding box shown in the Hololens 2 informs the user that the object is currently adjustable while also responding to the user’s finger’s proximity. This interactive visual feedback can communicate the possible interaction with users.

8 Limitation
The focus of our study was discoverability. Discoverability can be an important concern for the users who have difficulties in learning a new interface; however, such a pure novice user is only a small part of the picture. The results of the study would likely have been very different when participants had enough time to learn each interaction. Future work should address the progressive learning effect and further examine the usability of varying interaction techniques in the longer term.

With the worlds-in-miniature interaction, it is challenging to pick up the virtual object that is smaller than their hand size. Prior 3DUI works [16, 17] had tackled these issues and examine selection techniques in cluttered or dense VR environments. It would be interesting to explore how to design a WIM interface with these selection techniques. The technical limitation is another important concern that has to be addressed in interpreting study results. In creating a mixed reality setting, we used a Zed Mini camera’s video see-through features rather than using an optical head-mounted display such as a Microsoft Hololens. A video see-through images can suffer from image lagging and delays in responses, causing a motion sickness to the users. Leap Motion sensor does not provide the most accurate hand tracking nor supports realistic object occlusions, which can result in negative user experience.

Furthermore, we did not consider real-virtual occlusion in a larger space. In our study, we scanned the environment using a ZED mini camera, but it was limited to the small room space. In the larger cluttered environment, gaze-punch might be able to outperform touch-grab and WIM. Compared to other two, it would be easier to navigate the larger space and move the objects in a cluttered environment with gaze-pinch interaction. Future studies can explore various scenarios to address such concerns.

9 Conclusion
This study compares three widely known interaction techniques: gaze and pinch interaction, touch and grab interaction, and worlds-in-miniature. Overall, the majority of participants found that the pinch and grab interaction most challenging to use, while the worlds-in-miniature interaction provided users the most accurate, efficient, and easiest interaction in arranging multiple furniture items. The results of gesture analysis and in-depth interviews reveal the critical roles of visual affordance of virtual objects in determining the participants’ form of interactions in mixed reality. We further provide design suggestions that can aid in developing more intuitive interaction for a wearable mixed reality headset.
