Investigating Asymmetric Collaboration and Interaction in Immersive Environments

Daniel Enriquez

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science and Application

Yalong Yang, Chair

Chris L. North

Sang Won Lee

Brendan M. David-John

November 13, 2023

Blacksburg, Virginia

Keywords: Virtual Reality, Asymmetric Collaboration, Human-Computer Interaction, Collaborative Interaction, Mobile Devices: Phones/Tablets

Investigating Asymmetric Collaboration and Interaction in Immersive Environments

Daniel Enriquez

(ABSTRACT)

With the commercialization of virtual/augmented reality (VR/AR) devices, there is an increasing interest in combining immersive and non-immersive devices (e.g., desktop computers, mobile devices) for asymmetric collaborations. While such asymmetric settings have been examined in social platforms, questions surrounding collaborative view dimensionalities in data-driven decision-making and interaction from non-immersive devices remain underexplored. A crucial inquiry arises: although presenting a consistent 3D virtual world on both immersive and non-immersive platforms has been a common practice in social applications, does the same guideline apply to lay out data? Or should data placement be optimized locally according to each device's display capacity? To this effect, a user study was conducted to provide empirical insights into the user experience of asymmetric collaboration in data-driven decision-making. The user study tested practical dimensionality combinations between PC and VR, resulting in three conditions: PC2D+VR2D, PC2D+VR3D, and PC3D+VR3D. The results revealed a preference for PC2D+VR3D, and PC2D+VR2D led to the quickest task completion. Similarly, mobile devices have become an inclusive alternative to head-worn displays in virtual reality (VR) environments, enhancing accessibility and allowing cross-device collaboration. Object manipulation techniques in mobile Augmented Reality (AR) have been typically evaluated in table-top scale and we lack an understanding of how these techniques perform in room-scale environments. Two studies were conducted to analyze object translation tasks, each with 30 participants, to investigate how different

techniques impact usability and performance for room-scale mobile VR object translations. Results indicated that the Joystick technique, which allowed translation in relation to the user's perspective, was the fastest and most preferred, without difference in precision. These findings provide insight for designing collaborative, asymmetric VR environments.

Investigating Asymmetric Collaboration and Interaction in Immersive Environments

Daniel Enriquez

(GENERAL AUDIENCE ABSTRACT)

With the commercialization of virtual/augmented reality (VR/AR) devices, there is an increasing interest in combining immersive and non-immersive devices (e.g., desktop computers, mobile devices) for collaborations across different devices. While such asymmetric settings have been examined in social platforms, questions surrounding collaborative view differences in 2D views or 3D views affect data-driven decision-making and interaction remain under-explored. A crucial inquiry arises: although presenting a consistent 3D virtual world on both immersive and non-immersive platforms has been a common practice in social applications, does the same guideline apply to lay out data? Or should data placement be optimized on each device according to each device's display capacity? To this effect, a user study was conducted to provide insights into the user experience of collaboration across different devices in data-driven decision-making. The user study tested different combinations of 2D and 3D layouts between PC and VR, resulting in three conditions: PC2D+VR2D, PC2D+VR3D, and PC3D+VR3D. The results revealed a preference for PC2D+VR3D, and PC2D+VR2D led to the quickest task completion. Similarly, mobile devices have become an inclusive alternative to head-worn displays in virtual reality (VR) environments, enhancing accessibility and allowing cross-device collaboration. Object manipulation techniques in mobile Augmented Reality (AR) have been typically evaluated in table-top scale and we lack an understanding of how these techniques perform in room-scale environments. Two studies were conducted to analyze object translation tasks, each with 30 participants, to investigate

how different techniques impact usability and performance for room-scale mobile VR object translations. Results indicated that the Joystick technique, which allowed translation in relation to the user's perspective, was the fastest and most preferred, without difference in precision. These findings provide insight for designing collaborative, asymmetric VR environments.

Dedication

I would like to dedicate this thesis to my mother and father for their unrelenting support and encouragement.

Acknowledgments

This thesis includes content from two submitted papers, submitted to peer reviewed conferences. This thesis' theme is is that of cross-device collaboration between immersive and non-immersive environments Under the supervision of Professors Yalong Yang, Chris North, Sang Won Lee, Myounghoon Jeon, and Douglas Bowman, I iterated on the development, ideas, and designs of the work presented in this thesis. These two papers were made in collaboration with fellow student researchers, Hayoun Moon in the Grado Department of Industrial and Systems, Virginia Tech and Wai Tong in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. The two submitted papers are a collaborative work with these student researchers and professors and are pending review. D. Enriquez, H. Moon, D. Bowman, M. Jeon, S. W. Lee. Investigating Object Translation in Room-scale, Mobile Virtual Reality. *In Review* (2023). D. Enriquez, W. Tong, C. North, H. Qu, Y. Yang. Evaluating Layout Dimensionalities in PC+VR Asymmetric Collaborative Decision Making. *In Review* (2023).

Contents

List of Figures

List of Tables

List of Abbreviations

AR Augmented Reality

HWD Head-Worn Displays

VR Virtual Reality

VR is a 3D, interactive environment in which a person is immersed into a digitally created environment where movements of their head are tracked and images are displayed to their eyes to give the feeling of being in that environment.

AR is an alteration of the physical environment through a device, done through a mobile camera in the case of mobile AR or done through an immersive display.

HWDs are the displays of an immersive AR or VR environment, in which a display is held in place in front of the users eyes.

Chapter 1

Introduction

The desktop computing environment, commonly known as Personal Computers (PCs), is the most widely utilized computing environment. As a result, extensive research has been conducted on collaborative systems involving multiple PCs [76]. Today, a growing array of technologies has emerged, presenting new opportunities for beyond-desktop computing [7, 75]. Notably, the commercialization of virtual reality (VR) and augmented reality (AR) devices has accelerated at an astonishing pace, poised to become an integral part of our daily lives in unprecedented ways [12, 49]. The capacity to deliver interactive 2D and 3D graphics within VR/AR environments offers an opportunity to revolutionize our interaction with data and holds immense potential for facilitating collaboration [81, 82].

However, VR devices, more specifically, Head-Worn Displays (HWDs) are not accessible to all people, as people with certain vision impairments, those that experience VR sickness, children [44], and those with kinky or coily hair [60] may choose to not use VR. Alternatively, mobile devices are light and hand-held, allowing users to use their peripheral vision, permitting for many users in an environment simultaneously. Mobile devices such as mobile phones are widely adopted by the public, making them more accessible and available than HWD devices. Researchers have created systems that allow mobile device users to observe and participate in VR environments through a "Camera" metaphor in which the physical device position matches the position of the camera in the VR environment [30, 89]. This implementation of mobile VR provides a more accessible yet less immersive interface into

a VR Environment. This interface is akin to that of mobile AR experiences, with a 3D environment, viewed through the 6 degree-of-freedom (DoF) pose of a mobile device.

As the range of available computing environments continues to expand, it has become increasingly common for individuals to employ different devices to collaboratively accomplish tasks [14].

This thesis focus lies in the realm of asymmetrical collaboration between PCs and virtual reality (VR), given that PCs maintain their position as the mainstream computing environment, while VR/AR represents the emerging next-gen display and interaction platforms.

1.1 3D Immersive Environments

The integration of non-immersive and immersive environments presents exciting challenges and opportunities in interaction and user interface designs. Within PC+VR collaboration, one area of substantial interest is remote instruction, where PC users provide guidance for VR users. However, real-world decision-making processes often demand intensive two-way communication.

The other context for collaboration that requires unique design considerations is the perspective of mobile devices and VR devices. When interacting inside of a virtual environment, collaboration and interaction are necessary to study in relation towards one another. Interaction methods are well studied in immersive VR [49] and mobile AR [28], however, mobile VR interaction methods are not heavily investigated and more work is needed to determine how interaction occurs between these two paradigms.

As a basis, Mobile AR and mobile VR are similar in terms of input, more so than immersive VR, given that mobile AR and mobile VR both share similar means to interact with the

device and the environment. Other works such as Bambušek et al. [6] have shown the potential to integrate mobile AR UI into mobile VR. As a result, many mobile AR manipulation methods such as 3DTouch, HOMER-S, and SlidAR [63, 73] can be ported into mobile VR. Overall, mobile AR has typically been studied in smaller scale, table-top environments [28]. Given the emergence of using mobile devices in VR, it is questionable if the knowledge of object manipulation techniques studied in table-scale mobile AR would transfer to room-scale mobile VR. Therefore, extra design factors need to be accounted for as shifts in perspective and added distance effect the performance and usability of manipulation methods given these room-scale environments.

1.2 Research Questions

To understand these differences in cross-device collaboration, two research questions were defined.

RQ-1: How does layout dimensionality affect collaborators across different levels of immersion?

In light of the aforementioned trade-offs and design choices in PC+VR asymmetric collaboration, our objective is to investigate how individuals communicate, exchange information, and engage in discussions while dealing with different combinations of dimensionalities. Inspired by prior studies [39, 52, 53], we study this problem within a data-driven context, wherein multiple pieces of information are displayed in various windows, requiring collaborators to arrive at a consensus and make informed decisions based on the provided data. In terms of the information itself, we have employed a well-established data-driven decision-making task from literature [43], where pairs of participants collaborate to select the most suitable hotel from a given set.

RQ-2: What are the ways people understand and interact with a VR environment through mobile devices?

Prior literature has defined how to understand manipulation methods in 3D environments in both the physical world [15] and the virtual world [17] by extending the application of Fitts' law to accompany the added dimension, depth. These extended models incorporate angle-related predictors and contribute to our understanding of 3D object manipulation. However, in mobile VR, users interact with a 2D display to manipulate objects in a 3D virtual world. Therefore, it remains uncertain whether the adaptations of Fitts' law models for 3D environments could be applied to mobile VR. One goal of this work is to evaluate if room-scale object translations in mobile VR will comply with these existing Fitts' law models.

1.3 Research Methodology

In this thesis, three user studies were performed, one to evaluate the effects of layout dimensionality in cross-device collaboration (**RQ-1**) and the other to evaluate translation techniques in mobile VR (**RQ-2**).

For the first user study, there were three practical PC+VR designs, taking into account the dimensional aspects of PC and VR environments: `PC2D+VR2D`, `PC2D+VR3D`, and `PC3D+VR3D`. Specifically, `PC2D+VR3D` optimizes settings separately for PC and VR, facilitating individual works. Meanwhile, `PC2D+VR2D` and `PC3D+VR3D` provide user interfaces with the same dimensionality, either in 2D or 3D, for both PC and VR, enhancing collaborative possibilities.

For the other two user studies, an application was created to evaluate manipulation methods. A mobile device would load the environment as a 3D Unity game [88] that would have its

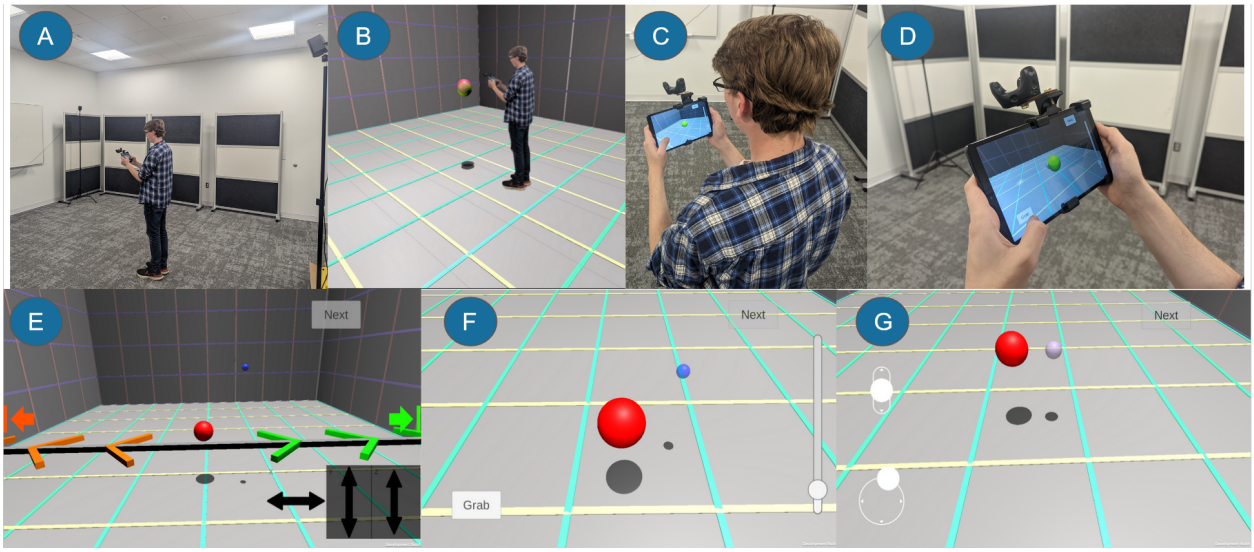

Figure 1.1: (a) The environment of the object translation user study with a user using the application. (b) A user superimposed into the VR environment (c-d) A closeup of the user's perspective with the user performing a translation. (e-f) showcase the object translation tasks that were evaluated in the user studies: (e) `3DSlide`, (f) `VirtualGrasp`, (g) `Joystick`.

Figure 1.2: A VR user and a PC user are collaborating to find the best hotel that suits their needs from the same set of hotels. This was the study task we tested in the controlled, layout collaboration study.

position synchronized from a PC running SteamVR [87] through a Vive tracker [18] on the mobile device. To evaluate manipulation methods in mobile VR, three translation methods were designed to cover a wide range of prior mobile AR interface design. `3DSlide` is a combination of 3DTouch [63] and SlidAR [73], in which independent of orientation, users can choose a world-coordinate axis along which to translate an object. This is an example of a touch-based interaction technique operated using global-coordinates. `VirtualGrasp` is partially derived from HOMER-S [63], in which users translate an object that mirrors the movement of their device, serving as a direct manipulation technique operating in the device reference frame. A slider was added to allow translation at further distances than originally grasped. `Joystick` is derived from virtual joysticks that are popular and common in mobile-phone gaming, serving as a local-coordinate, touch-based interaction technique.

1.4 Contribution

For the first user study, drawing insights from data collected from 18 pairs of participants, we have observed that optimizing user interfaces and interactions for individual environments positively influences the perceived user experience. Notably, participants generally preferred `PC2D+VR3D`. Additionally, maintaining a consistent view across different environments resulted in shorter completion times, with `PC2D+VR2D` emerging as the fastest overall. Furthermore, it is noteworthy that the common practice in asymmetric social applications (`PC3D+VR3D`) was frequently complained about by our participants, primarily due to the challenges posed by navigating a 3D environment on a PC.

From this work, asymmetric collaboration system was developed that facilitates collaboration between PC and VR environments. Secondly, a controlled user study was conducted to investigate the trade-offs between individual and collaborative effectiveness in PC+VR

collaboration, given the difference in layout dimensionalities across conditions.

Through two user studies, each with 30 participants, totalling 60 participants, we evaluated 3 factors to understand translation in mobile VR, target distance, target size, and user mobility conditions (stationary vs. movable). We found that optimizing the user interface for translation along a well-defined manipulation axis worked well, with participants favoring the `Joystick` method as it was an intuitive compromise between limited DoF control and usability. Results also indicated that device-based methods such as `VirtualGrasp` can be physical demanding in prolonged-use environments, yet also, `VirtualGrasp` was found to be significantly more intuitive when users were allowed to walk in the environment (movable) as opposed to remaining stationary.

For mobile VR, the contributions are as follows: First, we describe the design of how interaction methods can be transformed from a mobile AR translation into a mobile VR translation. Second, we provide findings and lessons learned from two user studies to understand how mobile VR translations are affected in larger scale environments according to their UI interface features and how this relates from traditional, table-scale mobile AR counterparts.

Chapter 2

Review of Literature

2.1 Immersive Analytics

Typically, data analysis tasks have been relegated towards PC environments; however, there is significant precedent to change this, given that immersive technologies provide tools that allow for sensemaking which provides spacial context towards thoughts that promote data interpretation [16]. As analysis tasks grow complex, other mediums should be leveraged to grow and facilitate understanding of analysis tasks [75]. Technologies such as tabletop displays allow users to understand data by providing more context for awareness [40]. The same benefit could be obtained from immersive head-worn devices (HWDs). Many visualizations leverage this extra dimension, and much prior work has gone into detail on the advantages and disadvantages Immersive Analytics provides [25, 48, 83]. As a result, a few of the relevant and key works will be discussed to provide insight into the direction to lead our work rather than providing an exhaustive list of Immersive Analytics.

Immersive analytics has a strong potential to facilitate analysis. However, less work was investigated in collaborating between immersive technology and traditional workflow with desktops. It is crucial for immersive Analytics as different users have different preferences, accessibility, and capabilities in using different devices for analytics, especially since immersive technology is relatively new and requires higher learning curves. Specifically, Ens et al. [22] define two grand challenges, i.e., *supporting behavior with collaborators* and *supporting*

cross-platform collaboration, which highlight the necessity to facilitate collaboration across different levels of the virtuality continuum [62]. With the increasing maturity of various computing environments, particularly immersive ones, there are now greater opportunities for collaboration using diverse devices. This has garnered significant interest from researchers and has led to the emergence of asymmetric collaboration [29] and cross-device collaboration [14]. Following this line of research, we aim to improve asymmetric collaboration with data visualization, especially in layout dimensionality.

One of the benefits of immersive analytics is the unlimited space, allowing multiple windows to be viewed at the same time in different locations in the virtual space. Layouting multiple windows in immersive analytics has been considered by researchers, and individual effectiveness could be altered as a result. One earlier example of this is Shupp et al.'s [85] analysis of viewport sizes in Large High-Resolution Displays in which greater viewport curvature decreases performance times and are preferred by users. More work has tried to analyze these benefits in immersive spaces, with works such as Ens et al. [23] simulating this effect of multiple windows in an immersive, VR environment. Lee et al. [50] discusses that 2D visualizations are often paired by users using walls and 3D visualizations are positioned using the space around them, with other works such as *Maps Around Me* [80] defining different patterns such as spherical, spherical cap, and planar orientations. Lisle et al. [51, 52] and Davidson et al. [19] explore how people understand and apply multiple windows in immersive environments. They discovered three of the main sensemaking arrangements that people employ in multi-window environments, semicircular, environment-based (those based on structures in the virtual environment), and planar. As a result, we decided to use semicircular (3D) and planar (2D) arrangements of layouts, as these prior works showcase how people make sense of space in immersive environments. We excluded environment-based layouts given the lack of environment encodings possible on a 2D PC screen.

The decision towards how information should be displayed in immersive environments has been one that has been long considered among other works. Liu et al. [53] discussed how given fewer multiples, a flat layout is preferred in cases where fewer multiples were present (4x3), in comparison to a larger set of multiples (12x3), where a semi-circular layout is preferred when interacting with 3D small multiples. In Liu et al.'s future work [54], performance between a flat layout, a semicircular wrap-around layout, and a circular wrap-around layout were all relatively the same, despite participants preferring the semicircular wrap-around layout more than the other layouts.

However, the effect of layout in asymmetric collaboration setting is still under-explored. We want to understand how layout and design between immersive and 2D desktop window management with small multiples affect users' performance, collaboration, and preferences.

2.2 Asymmetric Collaboration

The wide array of devices across the augmented-virtuality spectrum has led to a wide spectrum of papers encompassing different types of immersive devices and their mediums of collaboration [27]. Research has also developed a foundation for taxonomies to design applications allowing cross-device collaboration [14, 69].

Much work has shown how design considerations could lead to better collaboration in Immersive Analytics. Piumsomboon et al. [72] discussed that when it came to the sharing of awareness cues across users in immersive collaborative settings that the sharing of the head-gaze was considered the most useful and easy to use. From this, it was imperative that for cross-device collaboration to occur effectively, visual cues indicating user position would need to be implemented. Müller et al. [64, 65] discusses how the usage of shared virtual objects aided in user discussion over physical objects. To analyze so, they analyzed

their communication behavior and defined their speech for spatial expressions into different categories such as *"Physical object"*, *"Deictic speech"*, *"Person"*, etc. We derived the analysis of communication based on these works with minor modifications that would align with the chosen task.

Many platforms have been developed for asymmetric collaboration, with platforms such as Spatial [86], VirBela [93], and Meta's Horizon Workrooms [61] showcasing the variety of asymmetric collaboration and the necessity of incorporating other devices into the design of immersive analytic tasks. However, these commercial platforms primarily emphasize social interactions, offering a consistent 3D environment for both PC and VR. In contrast, our focus lies in productivity scenarios, allowing us to explore a wider array of design options across these two environments' dimensionalities. Kim et al. [47] introduced a framework enabling clients with various devices and immersion levels to interact within a common platform, specifically tabletop displays. More recently, Saffo et al. [77] argued that as the diversity of visualization types increases, a wider range of visualization metaphors becomes necessary to attain a higher degree of shared awareness. In a related vein, Tong et al. [90] conducted a controlled study to empirically evaluate the efficacy of asymmetric collaboration between immersive and non-immersive environments. They compared PC+VR asymmetric collaboration with PC+PC and VR+VR symmetric collaborations and discovered that asymmetric collaboration did not increase collaborative efforts, instead reducing the mental load associated with completing data analytic tasks. Furthermore, they recommended optimizing asymmetric systems for the collaborators' respective devices to enhance the user experience.

However, their recommendation lacked empirical validation when juxtaposed with alternative asymmetric designs. For instance, instead of employing distinct designs, one could opt for a consistent design that prioritizes either PC or VR affordances. Thus, one of the investigations of this work delves into an evaluation of various design alternatives in PC+VR asymmetric

collaboration, particularly concerning layout dimensionalities.

While HWDs offer high immersion with surrounded view and 6 DOF, mobile devices offer alternatives for broader populations to consume VR content due to HWDs' inaccessibility [44, 60]. Examples of this implementation are TranceiVR and WebTranceiVR [56, 89], in which the view of the HWD is transmitted for other devices. A similar work by Drey et al. [20] analyzed the effectiveness of symmetric (VR+VR) and asymmetric (VR+Tablet) learning, in which the users can see the VR environment using the mobile device through a spectator view with teaching instructions. Another work, XRDirector [67] incorporates a mobile device that serves as an AR camera into a VR environment. This mobile device has the capacity to interact with and modify the VR environment. Other works, such as ShareVR, VR Invite, and Owlchemy Lab's Mobile Spectator [1, 26, 30] integrate the user's position to provide embodied interaction inside a VR environment in relation to the user's view and provide experiences of such. These works also serve as good platforms to evaluate and explore interactions in mobile VR scenarios. However, the literature lacks understanding in how mobile-device users can interact with virtual objects, as the object manipulation in mobile VR has not been investigated. As a result, our work aims to analyze how input modalities function in mobile VR environments. The lens of object translation is a good avenue to analyze mobile device interaction and mobile VR input, given its capacity to effect objects of the VR environment through a variety of input methods.

2.3 3D Object Manipulation

Mobile VR offers the benefit of being applying interaction schemes to and from mobile AR. Examples of this can be seen with Bambušek et al. [6], in which a mobile AR environment is replicated in mobile VR because of mobile AR reachability limitations. The design im-

plications in the mobile VR environment in Bambušek et al. [6] provided insights to utilize mobile AR techniques, as our work extends design insights from mobile AR applications to create the interface for mobile VR systems.

A broad range of manipulation techniques exist in immersive environments, yet usually, translation techniques are fundamental in spatial interaction. Translation techniques allow us to the broad applicability of interaction while offering a foundation to understand how shifts in perspective and space affect input. With HWD users, VR manipulation methods are extensively well-defined [49], with device-based manipulations being common-place as the predominant means of manipulation for immersive VR [11, 42]. These insights help as the implications from the manipulations allow us to understand how device-based translations function in VR environments and how they may differ from mobile VR.

Similarly, AR manipulation methods are also well-defined, with 3 distinct classes of interaction techniques: touch-based, gesture-based, and device-based, being common-place in literature [28]. Given the abundance of new works that operate in a mobile VR context [1, 20, 26, 30, 56, 67, 89], it is clear that interaction needs to be better defined when analyzing mobile VR to understand how these techniques differ from their AR counterparts. Many methods for manipulation from mobile AR to mobile VR are relatively the same, given the "window-like" appearance to glance into another view. Insight from the extensive literature on mobile AR manipulation are applied to mobile VR because of their similarity, as VR has been an effective means to understand AR in prior work [13].

Mobile AR techniques have been closely examined through the lens of table-top interactions [28]. However, mobile AR techniques have typically not been examined in larger, room-scale environments. One exception to this is Hellmuth et al. [36], in which device-based and touch-based interactions were examined in relation to the creation and movements of anchor points in room-scale mobile AR. Hellmuth et al. examined the differences in time

and accuracy across manipulation and creation techniques at a larger scale. The results from Hellmuth et al. provide design considerations for larger scale environments. Motivated by this work, we decided to explore how shifts in environment scale could affect object translation. More specifically, given that VR environments tend to be larger in scale, typically encapsulating an entire room [26, 30, 67], our work examines mobile VR in room-scale environments.

There exist many AR manipulation techniques for handheld devices [28]. As a result, choosing broad techniques that encompass many scenarios is critical to the evaluation of mobile VR translation techniques. Oftentimes these methods derive from VR counterparts, such as HOMER-S [63], a derivative of HOMER [11] in which an object's position is mapped from the user's mobile device. We decided to adopt a similar implementation of HOMER-S to allow for device-based translations.

Alternatively, to adopt a touch-based interaction approach, systems such as SlidAR [73] and 3DTouch [63] integrate methods to translate objects along a singular axis, either the x, y, or z-axis. SlidAR translates along a fixed, epipolar line and 3DTouch translates according to the device pose and touch input. SlidAR benefits from having an epipolar line to position an object along while 3DTouch benefits from global object translation. Our work adopts a hybrid of these two methods to allow global-coordinate object translation.

To contrast global-coordinate object manipulation, our work also looks into local-coordinate object translation through the exploration of a directional joystick or gamepad, similar gaming consoles and controllers. One implementation is ChildAR [35], in which school-aged children could manipulate a car's position to drive the car around in a mobile AR game. Another implementation is Blaga and Gorgan [9], in which a drone is operated using two virtual joysticks.

Mobile AR gesture-based interaction was not evaluated because works such as Bai et al. [4] and Nor'a et al. [68] require users to determine hand position by looking at their own hand to manipulate the object [46]. The VR environment in mobile VR does not offer pass-through or hand tracking due to technological limitations as this functionality is not supported by mobile devices.

Other techniques such as Martinet et al. [59] and Guo et al. [31] were not considered because of their usage of multi-touch displays, oftentimes requiring two hands to perform manipulations. In our work, users would be required to hold the device as well. As such, these methods were not investigated, but rather, the implications from these works aid to understand how dividing DoF control function in 3D environments.

Chapter 3

Layout Dimensionalities in PC+VR Asymmetric Collaborative Decision Making

3.1 Introduction

To address **RQ-1**, differences in dimensionality have to be examined to understand the differences in collaboration. This chapter focuses on the implementation of a collaborative task into an asymmetric environment, the results from a user study to analyze the collaboration and the discussion of such.

As outlined in the literature review, certain commercial social platforms offer accessibility from both PC and VR. However, these platforms exclusively offer 3D environments in both PC and VR settings. Our aim is to systematically investigate design possibilities, specifically considering various combinations of dimensionalities, for PC+VR integration in a productivity context focused on data-driven decision-making.

We chose to focus on a collaborative decision-making task from the literature [43], where collaborators are deciding on hotel choice based on various factors like price, rating, and amenities. To allow users to collaboratively complete the task and based on previous work [77, 90],

we proposed the following design goals:

- **Facilitate hotel comparison.** The comparison of different hotels is a crucial aspect of the proposed task, and we aim to design user interfaces that effectively support this objective.
- **Support multi-scale navigation.** Viewing a large number of hotels' details simultaneously is impractical. Our design approach focuses on enabling users to seamlessly switch between different levels of detail based on the specific information they need at any given time.
- **Provide real-time visual awareness.** In collaborative settings, maintaining awareness of the collaborator's status, focus, and intent is crucial for effective communication and coordination. In line with this objective, we provided real-time visual awareness techniques to assist collaboration.

3.2 Design

3.2.1 Small Multiple Layouts in 2D and 3D

To enhance the process of *hotel comparison*, we implemented the small multiple design in our experimental conditions. A small multiple refers to a series of visually similar content, such as graphs or charts, that enables easy comparison [92]. In our task scenario, we utilized a small multiple to represent a group of hotels, where each hotel's information was presented in a consistent format. In a 2D environment, this small multiple was typically arranged in a grid format, as illustrated in Figure 3.1. In a 3D environment, the additional dimension allowed for different small multiple layouts. As mentioned, Liu et al. [53, 54] examined

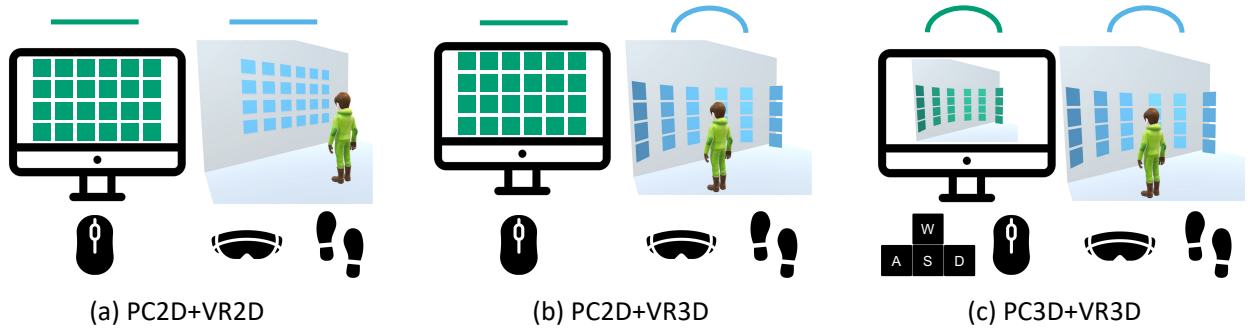

Figure 3.1: Three tested conditions in our user study. (a) PC2D+VR2D, where the layout of views is 2D in both PC and VR; (b) PC2D+VR3D, which involves a 2D layout for the PC collaborator and a curved 3D layout for the VR collaborator; and (c) PC3D+VR3D, where both PC and VR have a 3D layout. The PC collaborator uses pan&zoom to navigate in 2D environments (i.e., PC2D+VR2D and PC2D+VR3D), while employing a combination of WASD keys and the mouse to navigate in 3D environments (i.e., PC3D+VR3D). This navigation method is similar to playing a first-person shooter (FPS) game and is commonly provided by commercial PC+VR social platforms. The VR collaborator walks in the space for both 2D and 3D layouts.

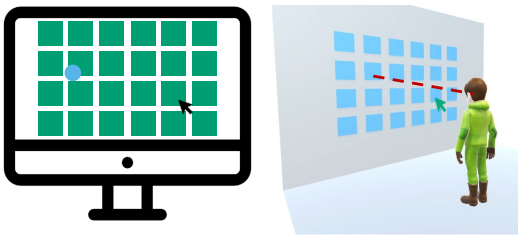

Figure 3.2: Demonstration of real-time awareness cues across PC and VR. Leveraging the depth-adaptive cursor technique [90, 97], we are able to provide real-time awareness cues across platforms and dimensions (i.e., in PC2D+VR3D). On PC (left), a moving dot shows which window the VR collaborator is looking at. At the same time, in VR (right), an icon is rendered to indicate the PC collaborator's cursor position.

Figure 3.3: Demonstration of multiple-level of the hotel information and the amount of information presented on PC2D at each level.

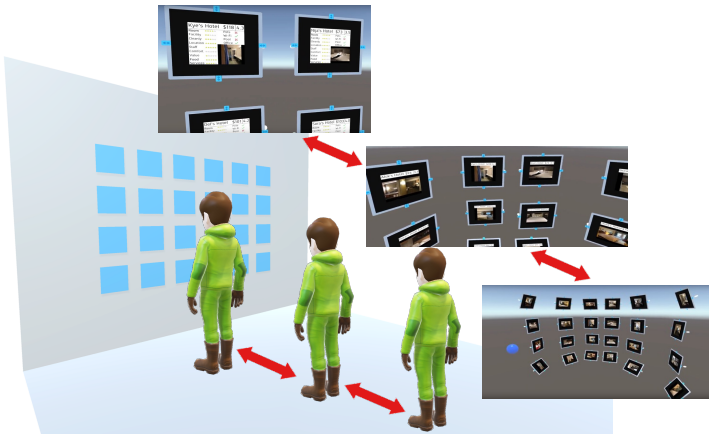

Figure 3.4: Illustration of bidirectional multiscale zooming in VR. Inspired by proxemic interaction [3, 5], the presented level of detail is determined by the distance between the user and the views. Three levels of detail were provided in our study: far distance results in an overview with hotel images; medium distance adds hotel names, prices, and ratings; close distance further provides breakdown ratings and amenities. The same levels of detail were provided on PC, and users used the mouse scroll to switch between them.

various small multiple layouts in VR, including a flat layout (similar to a 2D environment), a semi-circular layout, and a fully circular layout. The semi-circular layout demonstrated overall benefits compared to other alternatives, and therefore, we selected it as the layout for our 3D environment.

3.2.2 Multi-scale Navigation

Due to the extensive amount of information contained within the hotel dataset, it is impractical to present every detail comprehensively. This limitation arises from the restricted screen size of a PC, which cannot accommodate a large volume of information, and the fact that a VR user can only observe all hotel windows at a distance, with presented information being illegible. To tackle this issue, we incorporated multi-scale navigation into our tested conditions. This means that the level of detail presented depends on the available display space. Essentially, by reducing the number of hotels displayed on a PC screen or within

the VR user's field of view (FoV), we can increase the amount of detail provided for each individual hotel, and vice versa. On a PC, users have the ability to utilize the mouse scroll to zoom in and out, thereby controlling the number of hotels they prefer to see on their screen, see Figure 3.3. In VR, users can physically move closer or farther away from the view, allowing them to determine the number of hotels that fit into their FoV, see Figure 3.4. The VR multi-scale navigation was inspired by proxemic interaction [3, 5], which focuses on the relationship, particularly spatial relationship in our case, between people and the space around them and how it influences their interactions.

3.2.3 Real-time Visual Awareness

To enhance collaboration and communication, it is essential to provide users with contextual information about the content being referred to by their collaborators. To address this challenge, we have designed awareness cues that explicitly indicate the element under discussion. Providing awareness cues across different virtuality is challenging, as both the dimensionalities and interaction modalities are different. In order to bridge the gap between different virtuality levels, where dimensionalities and interaction modalities differ, we adopted the depth-adaptive cursor technique, building upon previous research [90, 97]. This technique allows us to project a 2D cursor into the 3D space and a 3D viewing direction into a 2D environment, as illustrated in Figure 3.2. By leveraging this approach, users can easily navigate and interact with elements across different dimensionalities while maintaining a clear understanding of their collaborators' focus. Additionally, we considered incorporating the view frustum into our conditions, as it has been found to be beneficial in VR/AR collaborative systems [72]. However, based on recent studies involving PC+VR setups, it was discovered that providing the frustum may not be necessary [90], as the frustum might be unsuitable for the PC environment.

3.2.4 Study Conditions and Implementation

In our study, we specifically focused on investigating how the different dimensionalities in small multiple layouts can impact collaboration in PC+VR settings. We conducted tests using three combinations: `PC2D+VR2D`, `PC2D+VR3D`, and `PC3D+VR3D`, as depicted in Figure 3.1. Across all conditions, we provided consistent multi-scale navigation and real-time visual awareness functionalities.

In `PC2D`, we utilized SAGE3[1] (Smart Amplified Group Environment), a successor of SAGE2 [74] for our implementation. It offers an infinite canvas as the working space, which is both zoomable and pannable, allowing users to overcome the limitations of a physical screen size. The concept of the infinite canvas has been adopted by various commercial tools, such as Miro and Google Jamboard. Within the SAGE3 workspace, we arranged the hotel information in a flat grid layout.

In `VR3D`, we adopted a semi-circular layout to organize the hotel information. This layout takes advantage of the additional dimension offered by the VR environment, providing users with a curved arrangement of hotel details. The semi-circular layout offers improved visibility and facilitates easy comparison between different hotels [53, 54]. Conversely, in `VR2D`, we used a flat layout instead. In all VR conditions, users employed natural locomotion, which involves physically walking to navigate in the virtual environment.

In `PC3D`, we utilized the semi-circular layout in a 3D environment, similar to the one used in `VR3D`. To navigate within this 3D environment on a PC, the user needs to use the WASD keys on the keyboard for movement (forward, backward, left, and right) and the mouse for controlling the camera viewpoint. This input method is widely familiar in gaming environments, and commercial social platforms, such as Spatial [86], VirBela [93], and Meta's

[1] https://sage3.sagecommons.org/

Horizon Workrooms [61].

We did not include PC2D+PC2D and VR3D+VR3D because these conditions are symmetric collaborative settings, i.e., collaborators use an identical device and digital tool for collaboration, and our focus is comparing the design alternatives in asymmetric collaborations. Additionally, these conditions were investigated by Tong et al. [90] by comparing them to an asymmetric condition (similar to our PC2D+VR3D) to identify its potential benefits. We also excluded PC3D+VR2D as this combination is likely to compromise both individual and collaborative effectiveness without evident benefits.

3.3 User Study

3.3.1 Task

The Hotel Search Task proposed by Jetter et al. [43] was slightly modified in that participants were given a searchable set of 24 hotels, displayed in a fixed 6x4 layout, and were also given unlimited time to look for a hotel. We did not want the time pressure to impact participants' behaviors. Participants were given a set of requirements that denoted their preferences when it came to which hotel to choose. These preferences took the form of a budget, overall rating, preferred amenities, or a preferred trait rating. Each preference was also binary in whether or not it could or could not be met, with each preference having equal weight. The PC participant was handed a list of these requirements on a piece of paper and the VR participant list existed in the virtual environment in a virtual canvas. This list differed across participants, training tasks, and conditions. The participants were not able to view the other person's preferences; however, they were allowed to communicate their preferences freely. The set of hotels and preferences was made to never allow for a hotel that fit all their

preferences, forcing participants to compromise. Tie-breakers were broken by price. This allowed for an optimal choice of hotel for each task.

This task was chosen because it was a collaborative task that necessitated the use of communication, collaborative strategies, as well as spatial referencing. Participants needed to communicate in order to describe preferences to one another. Participants needed to devise a strategy to examine the data. Participants needed to communicate location information when discussing hotels. This task was chosen over other tasks such as *"Stegosaurus"* [94] because of its relative simplicity, which allowed for straightforward derivation of collaborative actions. Other tasks might make it difficult to determine causality in collaborative behaviors, given the added complexity of the task. Other tasks also run the risk of low collaborative effort, given that participants may feel less comfortable to collaborate if another participant is actively doing more analysis.

3.3.2 Participants

The study had 18 pairs of participants, totaling 36 participants. Participants were asked to come in pairs with someone they were familiar with to promote natural conversations and to alleviate the confounding factor of personal relationships affecting study results. Participants were aged 19-30, with an average age of 24.47 years old and a standard deviation of 3.16. 21 participants identified as male, 12 participants identified as female, 1 participant identified as non-binary, and 2 participants decided not to disclose their gender. Participants were assigned to be VR or PC users based on preference and remained to use the assigned device throughout the study for all testing conditions. The VR users had an average VR experience rating of 2.94 out of 5 experience with a standard deviation of 1.43, with 1 having no experience and 5 having plentiful experience. All participants indicated that they had

an average collaborative shared workspace experience (Such as Miro, Google Jamboard, or SAGE) of 2.25 out of 5, with a standard deviation of 1.44.

3.3.3 Experimental Setup

For VR, a Meta Quest Pro headset was used, providing 1800×1920 pixel resolution per eye and a 90Hz refresh rate. The headset was wirelessly connected to a computer with an AMD Ryzen 7 5800X 8-core processor and NVIDIA GeForce RTX 3080 graphics card, enabling free movement within the $4 \times 4m^2$ space without cable impediments and leveraging the computer's powerful graphics processing. The windows were positioned $1meter$ in front of the participant's starting position, each $0.6 \times 0.4m^2$ in size. This arrangement allowed participants to view all 6×4 windows within their field of view. By walking through the space, participants could control how many windows were in their field of view at once, as well as the level of detail of the content (Figure 3.4).

For the PC setting, a 27-inch monitor with a 2560×1440 pixel resolution and 75Hz refresh rate was used. The monitor was connected to a computer featuring an AMD Ryzen 7 5800X 8-core processor and NVIDIA GeForce RTX 3080 graphics card. Initially, all 6×4 windows were displayed on the screen. Participants could scroll and zoom with the mouse to control how many windows fit on the screen at once, as well as the level of detail of the content (Figure 3.3).

All study conditions were able to be executed and interacted smoothly using the provided equipment and settings.

3.3.4 Design and Procedure

Our user study followed a full-factorial within-subjects design, with conditions balanced using a Complete Latin square to minimize the order effect. The study lasted for a total of around 90 minutes on average. Participants were initially welcomed and reviewed a consent form. Then, we briefly introduced the study's objectives and procedural steps. Following this introduction, participants proceeded to the various components of the study as follows:

Preparation: We asked participants to adjust the chair height to a comfortable level for PC and adjust the Quest Pro headset for VR before they started. We confirmed that all participants were in comfortable conditions and could see the text in all environments clearly.

Main Task: We followed the procedure below for each condition.

Training: Participants were asked to complete a simple version of the Hotel Search Task (Section 3.3.1) with a smaller data set (4×3 hotels) to get familiar with the new collaboration environment. Participants were free to inquire about interactions or tasks. The training concluded once participants were proficient with tasks and especially the interactions, generally taking 3-5 minutes.

Study Task: Upon completion of the training session, participants proceeded to the study task. The study task was the same as the training task but with a different and larger data set (6×4 hotels). Participants had no time limit for task completion but were encouraged to prioritize accuracy and efficiency. For the VR environment, we reset the participants' position to the center of the room and had them face the same initial direction before each study task started.

Break: Participants were given a mandatory 5-minute break to prevent stress, burden and to relieve potential physical demands of VR.

Table 3.1: The coding schema of the transcription for *Collaborative Effects*, derived from [58].

Category	Code	Definition	Example
Discuss Hypothesis	DH	A statement of noting a discussion of a hypothesis, either a claim or a comparison	*"I think John's is the best."*
Strategy Coordination	SC	A statement outlining coordinating strategies of what the pair should perform next	*"I get the left half, and you get the right half."*
Personal Information	PI	Sharing information pertinent that only they would know	*"I need a 4.2 overall. It needs to have a pool."*
Seeking Awareness	SA	Questioning and actively looking for knowledge of a location	*"Where was John's hotel?"*
Verbalize Findings	VF	Communicating a new piece of knowledge based on a recent finding of information	*"John's hotel has three compromises for me."*
Question Findings	QF	Questioning the findings	*"How many compromises does this have for you?"*

Table 3.2: The coding schema of the transcription for *Spatial Referencing*, derived from [65].

Category	Definition	Example
Exact Position	A statement noting a coordinate position of a panel	*"It's at Column 3, row 2."*
Relative Position	A statement noting the position of a panel relative to themself	*"It's to the right."*
Mix	A combination of the exact position and relative position	*"Third from the right."*
Deictic Speech	A statement that does not convey location information	*"It's over there"*

Ending: After the completion of all three tasks, the participants filled out questionnaires for the evaluation and ranking of all conditions, followed by a semi-structured interview.

3.3.5 Measures

The following measures are gathered from the perspective of *effectiveness*, *collaborative effects*, *group awareness*, and *preferences*.

Effectiveness. *Completion time* was measured from the moment the operator said to begin until the users indicated they were finished. *Choice accuracy* was also measured, given that each hotel had an individual count of compromises and issues to resolve ties. *Task load* and *collaborative engagement* were also evaluated with the 7-point Likert scale NASA TLX questionnaire [32].

Collaborative Effects. *Communication effectiveness, strategy, and coordination* were measured through a quantitative assessment of words spoken. To do this, two independent coders coded the transcripts of the audio recording. Each coder coded 12 sessions, with six overlapping sessions. The six overlapping sessions allowed us to evaluate the inter-coder reliability using Cohen's Kappa of > 0.7. The coding scheme was derived from Mahyar and Tory [58]

and was modified to fit with the task specified by this work. The coding scheme with examples is shown in Table 3.1. Other statements that are relevant but not categorized into these six categories become uncategorized.

Additionally, we were particularly interested in *spatial referencing* behaviors, as we anticipated it would be an important process in collaborative work. To measure spatial referencing, we utilized a coding scheme similar to Müller et al. [65] in which types of spatial expressions were grouped and the frequency was denoted for each individual. This coding scheme is shown in Table 3.2. Lastly, we were interested in seeing whether different layouts affect memorability during collaboration. Therefore, we counted the number of times when users mentioned that they forgot the position of the information, such as, "*Where is that?*" and "*Do you remember ...?*", as *forget*. We also counted the number of times when another participant directly responded to one of the questions asked above, such as "*So I remember the one is also three.*" as *recall*.

Group Awareness. To measure group awareness, we utilized the 7-point Likert scale behavior engagement questionnaire from Networked Minds Measure of Social Presence [10].

Preference. To assess preference, we requested participants to rank the conditions. Additionally, we gathered detailed rankings on user-*friendliness*, *productivity*, and *communication ease*.

3.3.6 Hypotheses

Five hypotheses have been made in order to assess the differences PC users and VR users have in asymmetric collaboration.

Effectiveness. We believe that `PC2D+VR3D` will have the best individual effectiveness (H_{eff}), given that the environment is suited to both meet the needs of the PC and VR users.

Specifically, users in PC2D+VR3D completed the task the fastest (H_{effA}), while the accuracy of the task should remain similar since we did not provide additional functionalities for specific conditions (H_{effB}). Lastly, users should perceive the least task load (H_{effC}) in PC2D+VR3D.

Collaborative Effects. We believe that PC2D+VR2D will have better collaborative effectiveness (H_{coll}), given that in the system, the VR user's environment appears extremely similar to the PC user's environment. Specifically, while communicating, participants in PC2D+VR2D should have the fewest instances of spatial references (H_{collA}) and more instances of discussing the hypothesis (H_{collB}). On the other hand, we expected the involvement of VR3D to leverage more spatial memory than VR2D which could help recall the forgotten information. Therefore, participants in PC2D+VR3D and PC3D+VR3D should have more instances of recall than PC2D+VR2D (H_{collC}).

Group Awareness. We derive that given the "eyes-and-shoes" principle defined by Saffo et al. [77], we can expect that given the diverse extreme of the virtuality continuum that the users' experience, greater group awareness will be required. As a result, PC2D+VR3D should have the most social behavior (H_{aware}).

Preferences. We expect that PC2D+VR3D will rank the best as both individuals are able to use the most comfortable representation for the working device [90] ($H_{preference}$). Specifically, PC2D+VR3D will be preferred in terms of user-friendliness and productivity but not communication ease as asymmetry will lead to harder spatial referencing.

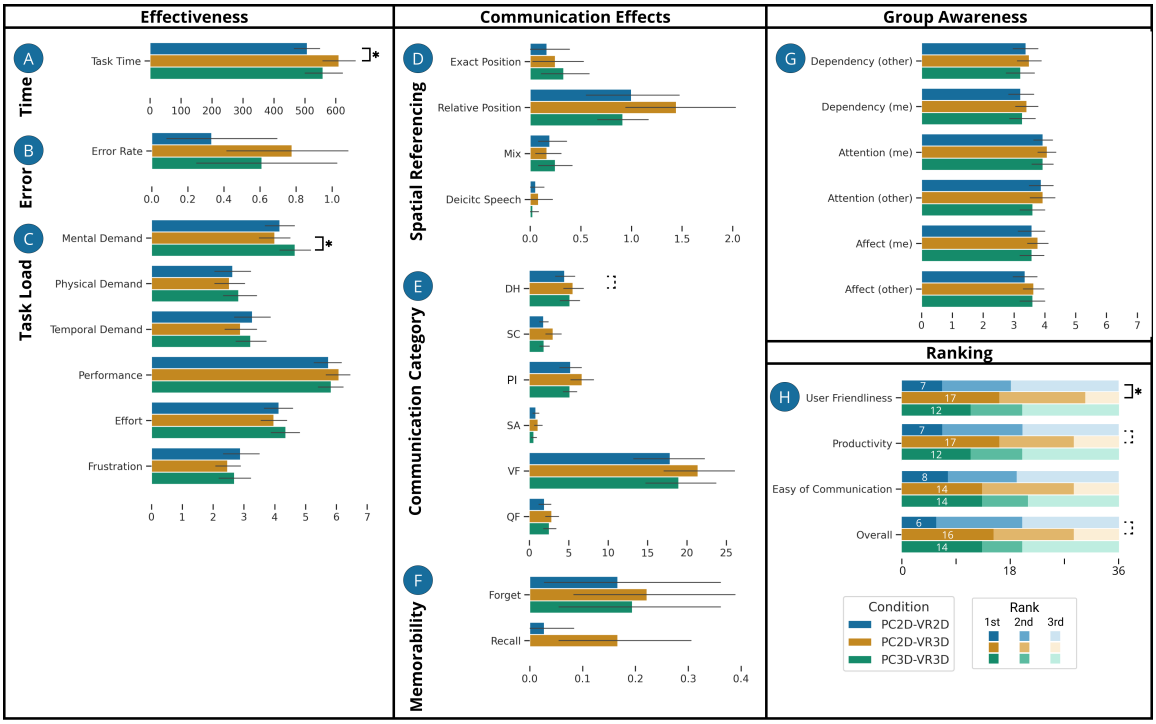

Figure 3.5: The figure shows the mean and 95% confidence intervals across all three conditions with the result of all measures, i.e., (a) task time in seconds, (b) error rate, with 0 being the correct choice, 1.0 being the second best choice, 2.0 being the third best, etc., (c) task load, (d) spatial reference frequency, (e) communication category frequency, (f) memorability frequency, and (g) group awareness. Additionally, a stacked bar chart displays (h) shows the distribution of ranking. Dotted lines indicate marginal significance, while solid lines with stars indicate the level of significance, with symbols denoting a p-value of less than .05 ($*$), .01 ($**$), and .001 ($***$).

3.4 Results

We applied different tests and analytics methods to analyze the data. For completion time, we first applied a log transformation to meet the normality assumption and then used *repeated-measure ANOVA* with *Tukey post-hoc* analysis. For all other quantitative measures, we conducted *Friedman tests* with *Nemenyi post-hoc* analysis to test if there are significant differences. For qualitative feedback, we adapted affinity diagramming [34] to analyze the subjective feedback of the individuals from the transcribed interview recordings. Significance values are reported for $p \leq .1(\cdot)$, $p < .05(*)$, $p < .01(**)$, and $p < .001(***)$. The results can be seen in Figure 3.5.

Effectiveness. For *completion time*, We found there is a significant difference in complete time ($F = 4.64$,$p = 0.01$, $*$). Figure 3.5(A) showed that participants in `PC2D+VR2D` completed the task significantly faster than `PC2D+VR3D` ($p = 0.02$, $*$). H_{effA} **is rejected.** In terms of *accuracy* (Figure 3.5(B)), we did not find a significant difference between the three conditions, and H_{effB} **is accepted.** Regarding *task load* (Figure 3.5(C)), no notable differences were observed except for mental demand ($\chi^2 = 6.34, p = 0.04$, $*$). The study revealed that `PC3D+VR3D` leads to considerably higher mental demand compared to `PC2D+VR3D` ($p = 0.04$, $*$). H_{effC} **is partially accepted.** As a result, we **partially rejected** H_{eff} given `PC2D+VR3D` did not provide the best effectiveness but could reduce the mental demand compared to `PC3D+VR3D`.

Collaborative Effects. For *spatial referencing* (Figure 3.5(D)), there is no significant difference between the three conditions. H_{collA} **is rejected.** For the number of instances in *communication categories* (Figure 3.5(E)), we have observed a marginal variance in the number of DH (discuss hypothesis) ($\chi^2 = 5.81, p = 0.05$, $\cdot$) instances between the three conditions. More specifically, we found that `PC2D+VR3D` exhibits a marginal significantly

higher number of DH instances compared to PC2D+VR2D ($p = 0.09$, $\cdot$). H_{collB} **is rejected.** For *memorability* (Figure 3.5(F)), there is a notable variance in recall between the three conditions based on the outcome ($\chi^2 = 8.85, p = 0.01$, $*$). However, we cannot identify any significant differences when comparing each pair of conditions. Therefore, **we cannot accept** H_{collC}. As a result, we **rejected** H_{coll} given PC2D+VR2D did not have the best communication effects, yet PC2D+VR3D motivated more discussion.

Group Awareness. Based on the collected data shown in Figure 3.5(G), we did not find a significant difference between the three conditions in terms of *social presence*. Thus, we **reject** H_{aware} that PC2D+VR3D required more group awareness.

Preferences Rankings. In general, participants rated PC2D+VR3D better than PC2D+VR2D, with marginal significance ($p = 0.06$, $\cdot$). Specifically, in Figure 3.5(H), PC2D+VR3D was considered significantly more user-friendly than PC2D+VR2D ($p = 0.02$, $*$). Additionally, it was marginally rated as more productive than PC2D+VR2D ($p = 0.06$, $\cdot$). Therefore, $H_{preference}$ **is partially accepted** . We were also interested in whether the perceived preference changed for different dimensionalities in the same environment and did not find statistical significance in the analysis.

Qualitative Feedback. In PC2D+VR2D, participants with the 2D PC commented they are *"comfortable"* and *"easy to interact with"* in this environment. Generally, participants dislike the flat layout and *"more physical movement"* in VR 2D, however, since both of the users are using the interface with the same dimensionality, VR P11 mentioned that *"I mean the one that I preferred the most was probably the linear layout. Um, why is because the way that we executed our strategy, you know, I didn't need curvature, right?"*.

For PC2D+VR3D, participants are able to use their best-fit environment with the devices, it makes the collaboration with *"better communication"*. For example, VR P7 commented that

"I mean the second one was easier just because I know he [PC P7] was having trouble with the third one. So it just like made communication easier." Moreover, compared to `PC2D+VR2D`, the VR users are happy by reducing their physical movement. VR P13 pointed out that *"overall the third one. Cause it was just like the easiest one to look at. ... the third one [`PC2D+VR2D`] was the most like circular, so I just kind of had to move my head."*

Lastly, for `PC3D+VR3D`, by introducing PC 3D, again, we made both users collaborate in the same dimensionality. However, participants dislike the interaction because it requires bimanual input, which provides a steep learning curve. For example, PC P15 stated that *"[...] because I used to play video games like a very long time ago, [...] it took me a while to navigate things with. So, but if someone, like someone else already knew how to play video games, it would be easier for them to navigate."*

3.5 Discussion

Individual effectiveness significantly influenced user preference: `PC2D+VR3D` was overall preferred. We consider 2D as the optimal environment for our tested task on a PC because participants frequently complained about using a mouse and keyboard to navigate a 3D environment. Meanwhile, leveraging the 3D display space was found to be more effective than having a flat layout by Liu et al. [53, 54]. Based on their findings, the 3D environment was more ideal for the collaborative task for VR users. Overall, our collected data found that participants favored conditions in which individual effectiveness was the priority. This effect was predominantly seen in the `PC2D+VR3D` condition, where both the PC user and the VR user had the interface be individually effective to both users. This was felt in multiple participant responses, in which `PC2D+VR3D` significantly ranked higher than `PC2D+VR2D`, such as in overall preference, productivity, and user-friendliness. For overall preference, when

adjusted according to user type, the VR users preferred the `PC2D+VR3D` condition more than the `PC2D+VR2D` condition, which aligns well with previous studies from Liu et al. [53, 54].

Collaborative effectiveness was positively correlated with the completion time: `PC2D+VR2D` was overall fastest. Despite the fact that `PC2D+VR3D` was the most preferred condition, `PC2D+VR2D` performed significantly faster with respect to task completion time than `PC2D+VR3D`. A possible reason why `PC2D+VR2D` significantly outperformed `PC2D+VR3D` is because of the added collaborative effectiveness in the VR user and the PC user sharing the same view. A reason why `PC3D+VR3D` did not have the same level of performance despite this added collaborative effectiveness could be caused by the added mental demand required to operate the system for the PC users, hindering the individual effectiveness considerably and therefore making the condition result in lower completion time. A possible design implication could mean that VR users could find it easier to adjust to a flat display than PC users would with a 3D display, meaning VR users are more capable of compromising individual effectiveness given the improvement provided to collaborative effectiveness.

Participants did not perceive an obvious difference when the layout dimensionality of their collaborator was changed. Many participants noted similarities between conditions. This was the intention of the study design, given that in `PC2D+VR2D` and `PC2D+VR3D`, the PC user was given the same display, and in `PC2D+VR3D` and `PC3D+VR3D` the VR user was given the same display. PC P9 wrote *"I don't remember any significant moments where things felt different in both the 2D ones."*, VR P4 wrote *"there wasn't much of a noticeable difference between the first and second experiments because I felt like really nothing changed"* and VR P7 wrote *"I couldn't really notice that my partner had a completely different experience."* Other participants did, however, notice this difference in collaboration, with PC P6 claiming *"When my partner had 2D he went too fast, and I had the most problem keeping up with him"* and PC P2 saying *"The 2nd and 3rd condition was essentially the same for*

me because PC was 2D, but I ranked the VR3D higher because we were a lot more efficient, however, this was a small minority given most participants did not notice a collaborative difference across conditions. Despite users not noticing the visual difference in the environment, collaborative effects were still felt as participants performed differently given the collaborative effort required by them or their partners. This is indicated by `PC2D+VR2D` having a lower completion time than `PC2D+VR3D`.

Having consistent dimensionality across PC and VR resulted in a "follow the leader" workflow, while different dimensionalities (`PC2D+VR3D`) promoted hypothesis discussion and led to more collaborative effects. *"Follow the leader"* is a workflow in asymmetric collaboration identified by Saffo et al.'s [77], where users with views designated to their task took on a leadership position. We observed that the PC user in the `PC2D+VR2D` user was more frequently the *leader* given that the view for the VR user was altered to correspond to the PC user, making the PC user the leader, with a similar effect happening with the VR user being a leader in the `PC3D+VR3D` condition. We found that this *follow the leader* may have occurred, given in the communication analysis, the condition with the most DH and SA in the pairwise comparison was `PC2D+VR3D`, the condition in which individual effectiveness was prioritized for both users. In `PC2D+VR3D`, since the participants are *equal* in terms of leadership, they both have equal opportunities to DH, as DH could be seen as a type of *leadership* communication, given the leader discussed hypotheses necessary to the completion of the trial and if both participants felt as if neither was *in charge*, both shared equally. Similarly, SA could be seen as a *follower* communication type, being inquisitive. Given that both participants may be seen as equal, both participants could feel inclined to ask questions to the other. The analysis does not find strong, concrete evidence of *follow the leader* attributes, but these two collaborative effects in communication could have been from *follow the leader*.

One conclusion that could be made clear as a result of `PC2D+VR3D` having the most DH and SA occur could be that the inconsistency in layout dimensionalities leads to more discussion. Given that users are in an environment where individual effectiveness is prioritized, both users could feel equally inclined to discuss amongst one another, thus leading to more DH and SA than the other conditions.

Navigating a 3D environment was cumbersome on a PC: `PC3D` was frequently complained about and `PC3D+VR3D` was more mentally demanding. Specifically for PC users, the `PC3D+VR3D` condition was found to have higher mental demand than the `PC2D+VR3D` users. Possible reasons for this discrepancy could be because of the extra degrees of freedom and extra controls required to interact with the 3D environment. Users moving with one hand and looking at the environment with their mouse, combined with general unfamiliarity with 3D environments on a PC screen, could contribute to the added mental demand perceived by the users. As a result, the effort required to manipulate the camera and understand the environment contributed to this effect of `PC3D+VR3D` users having more added mental demand. This claim was also facilitated by some of the users, with PC P16 saying about the `PC3D+VR3D` condition *"Using both hands and communicating with my partner at the same time was very challenging."* and PC P3 claiming *"I found the use of the keyboard [...] a little irritating and took time to move around. So I was a bit slower, so the partner was initially taking the lead. Once I got used to it, I could pace up with my partner and started to take the lead with the preferences said first from my end".*

Different layout dimensionality combinations resulted in similar spatial referencing behavior. Although we identified influences of layout dimensionality combinations for asymmetric collaboration, no statistical significance was found between spatial referencing types and amounts across conditions. This suggests users employed similar methods to discuss items spatially, regardless of dimensionality differences. Surprisingly, this contrasts our

original expectation that the distinct visual presentation and navigation affordances across conditions would confer different levels of spatial awareness. Specifically, the clear difference between 2D and 3D on PC should manifest in the data. We posit the designed visual awareness tool (Figure 3.2) reduced the required explicit spatial referencing (i.e. verbal communication), with participants instead checking the visual indicator implicitly. While we still consider the total spatial referencing action count affected by dimensionality combinations, future studies should collect data on implicit spatial referencing actions (e.g., looking at the visual indicators) to verify this.

Chapter 4

Object Translation in Room-scale, Mobile Virtual Reality

4.1 Acknowledgement

This work was completed equally from my fellow co-author, Hayoun Moon, who consented to share our work in the presentation of this thesis. We both contributed equally to the production of the work displayed in this chapter. More specifically, Hayoun Moon worked on the research methods, user study, and discussion equally alongside myself, while completing most of the analysis.

4.2 Introduction

To address **RQ-2**, this chapter will define the implementation of a mobile VR system that aims to understand different manipulation methods to understand how people interact inside the given environment, followed by two user studies to evaluate the differences in the manipulation methods and then a discussion that compiles the considerations from the results of the study.

As identified in Related Work, mobile VR serves as a good tool for an accessible, co-located

Figure 4.1: A sketch of an application of mobile device users interacting with a person inside of a virtual environment.

experience. Implementations of mobile VR allow for the integration of many users into a single virtual environment without the need to be immersed via a HWD. The implementation of this system could allow for mobile device users to look into a virtual environment that an immersed HWD VR user is seeing and interact within their environment, as seen in 4.1. The integration of these users can modify the experience of the immersed VR user while providing a sense of scale and understanding that VR offers to the non-immersed user. To integrate both mobile VR and immersed HWD VR users into the environment, we can allow for interactions using their independent domains to provide manipulation onto the world's objects. The literature involving immersive VR interaction is extensive, with many methods to support interaction of the environment. However, for mobile VR, the literature specifically analyzing in detail the specifics of the mobile device in a VR domain is limited.

As a result, we chose to focus on the evaluation of interaction within mobile VR as it is necessary to understand the ways in which mobile device users can integrate their experience into the VR system. Given mobile VR's applications are typically in larger scale environments than mobile AR, it is necessary to understand how mobile VR's interface functions in relation to space, given the difference in UI. As a result, object manipulation - more specifically, object translation - serves as an effective means to evaluate spatial interaction in 3D environments. Object translation is a common task in deployments of mobile VR and mobile AR systems. For entertainment and manufacturing, moving objects is quite common to complete tasks. in education, interacting and moving objects are necessary to apply kinesthetic experiences to learners.

An example of this can be seen in the Virtual Solar System [70], in which many mobile device users are able to access a learning experience inside of a virtual environment. As mentioned in the literature review, young children may not be suitable to wear immersive HWD but can still benefit from the sense of space and scale offered by a virtual environment. Users that are not immersed can still have the benefits offered from the environment, allowing collaboration among one another and other immersed users.

Object rotation and object scaling were not chosen as the interaction were similar to their smaller environment counterparts, given the lack of spatial information needed for task completion. As such, different manipulation techniques were created to highlight the differences in space offered by mobile VR and differences in UI.

A spherical ball was chosen as the movable object as rotation was not a factor in this experiment, with a transparent spherical ball showing the target location for the position of the translation. By default, the movable object would remain selected throughout the application.

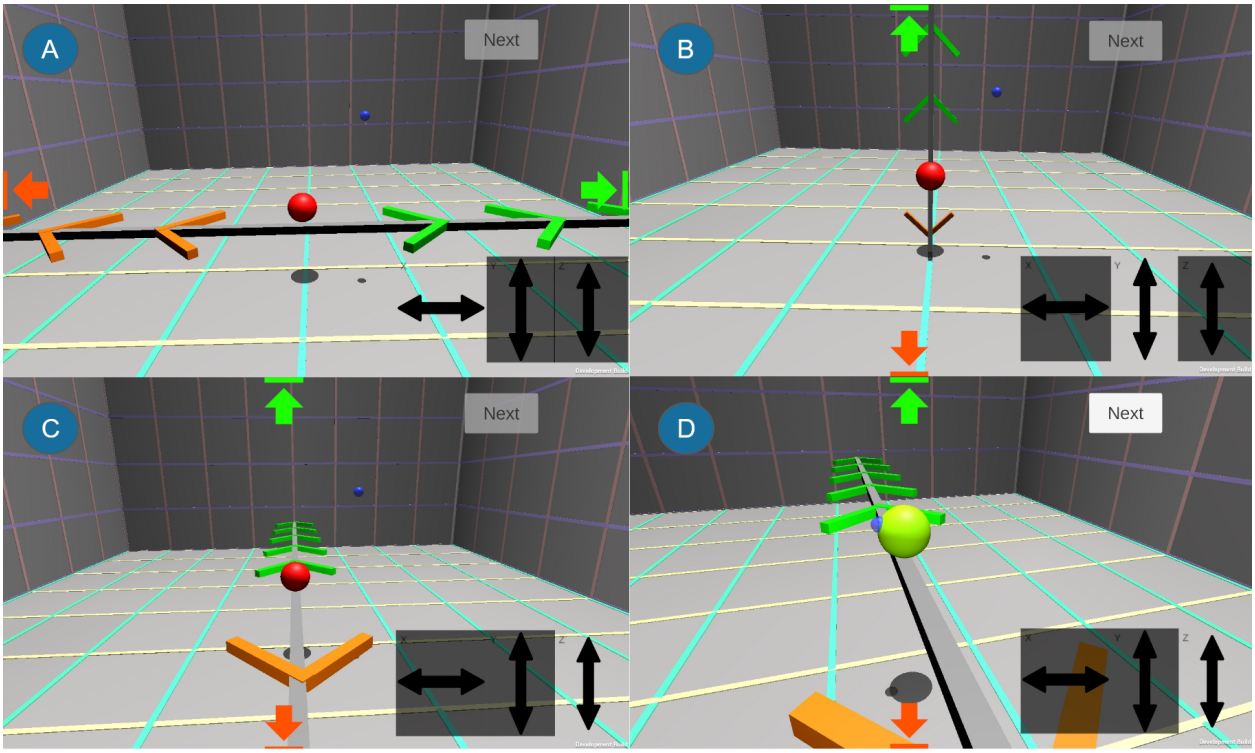

Figure 4.2: The UI for `3DSlide`. (a) Initial state of the environment, with a movable object (red) that needs to be translated to touch a target object (blue) with x-axis. (b) The state of the environment with the y-axis button selected. (c) The state of the environment with the z-axis button selected. (d) The movable object has been translated successfully and turned green with the z-axis button selected.

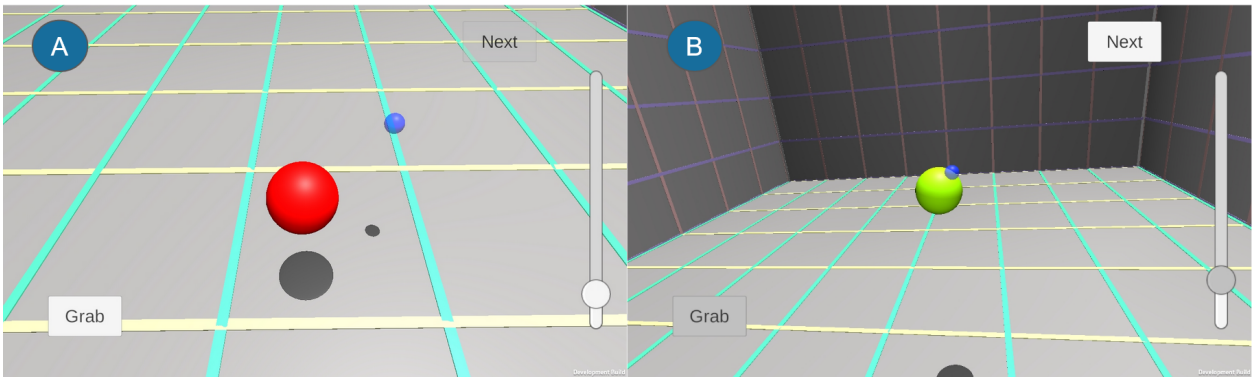

Figure 4.3: The UI for `VirtualGrasp`. (a) Initial state of the environment, with a movable object (red) that needs to be translated to touch a target object (blue). (b) The user has grabbed the movable object by pressing the "Grab" button, translated it using physical device movements, and changed its depth using the slider, so that the movable object is touching the target.

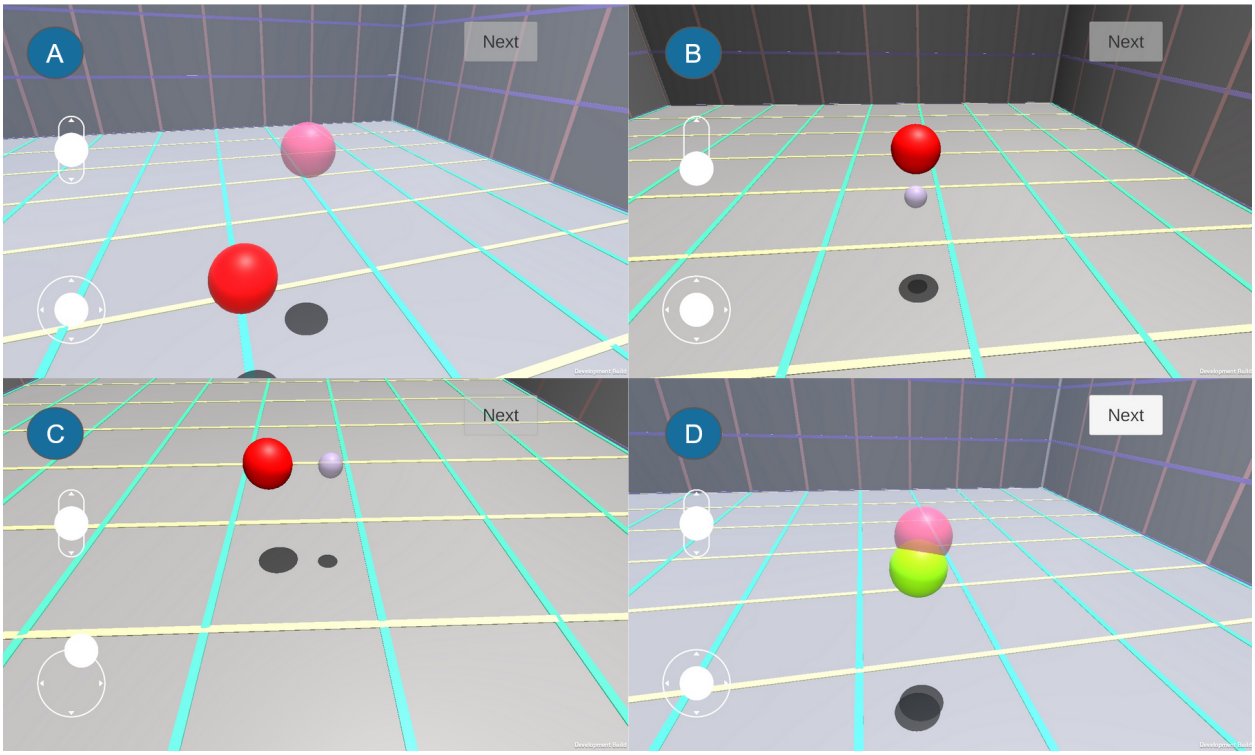

Figure 4.4: The UI for `Joystick`. (a) Initial state of the environment, with a movable object (red) that needs to be translated to touch a target object (blue), alongside two joysticks. (b) The state of the environment with the vertical joystick moving. (c) The state of the environment with ground-plane joystick moving (d) The movable object has been translated successfully with no joysticks being moved.

4.3 Design

4.3.1 Translation Techniques

We decided to look at three translation techniques that would encompass a wide variety of previously studied mobile AR interaction techniques and familiar techniques for input. As a result, we decided upon the methods described below.

`3DSlide` is a touch-based interaction method that is a modification of 3DTouch developed by Mossel et al. [63] and SlidAR developed by Polvi et al. [73]. Instead of restricting movement to one axis based on device orientation or an epipolar line, it restricts movement based on user input. Users would translate the object by selecting a given axis from a set of on-screen buttons and then sliding their finger across the screen in the 2D axis that represents the direction of the translation. This direction would be on the UI indicator for the axes buttons as seen in Figure 4.2. Because users have direct access to control the axis of translation, this method uses a global-coordinate system. To aid users in translation, the VR environment would display arrows in different colors that would match the direction of the UI. Given that the users would be required to translate an object in front of them, the global-coordinate axis would always align with the user's starting local coordinate axis.

`VirtualGrasp` is a device-based interaction method that is akin to HOMER-S developed by Mossel et al. [63], in which an object would be positioned by the users' device position. Since the movable object was constantly selected, the technique used a "grab" button that would grasp the movable object, regardless of range and can be see in Figure 4.3. The original design of HOMER-S restricted the distance of the movable object in positional translation to be that of the original selection of the movable object. Since the size of the environment was larger than the size of the room, it was apparent that a way to control the distance from

the movable object and the user would be necessary. The ”drag-and-drop” style of HOMER-S would require users to excessively move their device if their target location were to be outside of their physical movement range (e.g., moving backwards to increase the distance from the movable object to the user, and moving forward to move the movable object, and then repeating this until the movable object is at the correct position). Thus, to control object distance more easily, a UI slider was added to control distance.

`Joystick` is a touch-based interaction method that uses local-coordinates to manipulate the object. There are two joysticks as seen in Figure 4.4. One controls the ground plane, moving the object left, right, forward, and backward relative to the user's perspective, resulting in `Joystick` using local-coordinates. The other joystick controls the vertical plane, moving the object up and down. The vertical joystick can only move up or down as shown in Figure 4.4b. The virtual joysticks are similar to those found in mobile-phone applications.

This defined three translation techniques, each with varying restrictions on simultaneous DoF control. `3DSlide` was restricted to simultaneous control of 1 DoF, `Joystick` restricted to 1 or 2 DoF, and `VirtualGrasp` with 3 DoF.

4.3.2 Environment Design

The environment was made in Unity [88], using SteamVR [87] as the primary means of tracking. The VR environment is loaded on the mobile device and the device position would be synchronized through a Vive Tracker [18].

Similar to Olin et al. [69], the mobile device serves as a camera with its VE position synchronizing to the position of the device. The tracker's position is synchronized to the mobile device in Unity using Photon Unity Networking [71].

Depth cues were added to aid in object translation as is common practice in many VEs [21].

Lines were on the floor of the VE to indicate 1 meter of physical and VE space on the X (horizontal) and Z (depth) axis, as well as on the walls of the VE to indicate 1 meter of space on the Y (vertical) axis. Shadows were directly beneath the objects aid in positioning the object along the ground plane. The implementations of these depth cues can be seen in Figures 4.2, 4.3, and 4.4.

4.3.3 Exploratory Study

We performed an exploratory study to inform how mobile AR translation UI would differ from mobile VR given the room-scale environment. To analyze how larger environments could inform UI factors, we chose to randomize user position, movable object starting position, and goal position. With 5 participants, we gauged a few key design takeaways that would help with the creation of the task.

We determined the distance from the starting position of the movable object to the user was important to keep static as most common object manipulation tasks depend on the distance to move the movable object to a given target destination, rather than the distance from the user to the target.

We decided to study the variables that led towards the most disparities between small and large scale manipulation. Whether or not users were able to walk was found to be critical in our tests, given certain manipulation methods would appear more or less challenging when users were allowed to walk the environment. Allowing users not to walk would replicate scenarios where virtual objects could not be reached, such as a virtual object extending past the boundaries of the physical room. From this, we decided that the VE should be larger than the size of the physical room. We also decided that the distance the object would travel would be the most interesting variable. It was also apparent that some manipulation

methods may be better suited for precision tasks while others may be suited for speed tasks. As such, target size would be variable across trials.

4.4 User Study

4.4.1 Overview

In this study, we conducted user studies in two parts. In User Study 1, participants performed the translation task at the designated location, whereas in User Study 2, participants were free to walk around the room to perform the same task. The goals of User Study 1 were to theoretically examine how each element would affect performance and to understand translations when users choose not to walk or are unable to walk (e.g., out of reach for the user). User Study 2 aimed to investigate user experience in a more practical setting. Other than being stationary versus mobile, all procedures and tasks were identical for both user studies.

4.4.2 Experiment Design

For both user studies, we employed a $3 \times 3 \times 3$ within-subject design involving three independent variables: **translation technique**, **target size**, and **distance**. Each participant experienced all three translation techniques – `3DSlide`, `VirtualGrasp`, and `Joystick`. Target size was designed in three levels: large (0.25 m, equivalent to the movable object's radius), medium (0.125 m, half the radius); and small (0.075 m, one-third of the radius). Additionally, we varied the distance between the movable object and the target destination across three levels: 1 m, 2 m, and 4 m.

The set of target object locations was predefined in the combination of inclination angle, azimuth angle, and distance in accordance with those presented by Cha and Myung [15]: inclination angle represents the angle between the positive z-axis (pointing front) and the target location, azimuth angle represents the angle between the positive x-axis (pointing to the right) and the projected target location on the x-y plane, and distance represents the radial distance to the target. In this study, a total of 18 unique target locations were provided: 3 inclination angles ($\theta_1 = 30°$, 45°, and 60°) $\times$ 2 azimuth angles ($\theta_2 = 0°$, 45°) $\times$ 3 distances ($A = 1$ m, 2 m, 4 m).

4.4.3 Measures

In this study, our primary aim was to examine the impact of different translation techniques and task-related factors on performance and user behavior. In each trial, we measured the completion time for participants to translate the movable object from its initial position to the target location. Measures were recorded from the moment the participant first touched the tablet display to initiate the object's movement until the object reached its final position as determined by the participant.

To assess user behavior, we measured the percentage of time the participant spent fine tuning the object's position near the destination. We defined the fine tuning region as a sphere centered on the target location, with a radius equal to 20% of the total distance between the object's initial position and the target location. The time the movable object was present within this region was then divided by the completion time for the trial. To quantify user movement, we recorded the tablet's live position throughout the task, and divided its total travel distance by the completion time for the trial.

Upon completing each translation condition, participants evaluated its usability using HARUS [79],

which measures usability through two subcategories, manipulability and comprehensibility, with eight question items. Participants also rated their perceived workload using NASA-TLX [33]. At the conclusion of the entire study, we collected participants' preferences among the three translation techniques and solicited their comments explaining their preferences.

4.4.4 Apparatus

The participants used a Samsung Galaxy A7 Lite (8.7", 32 GB), weighing 366g and running a 1.8GHz octa-core MediaTek Helio P22T (MT8768T) processor [78]. The mobile device would run the application and the device's position would be located through a Vive tracker [18] that would be synchronized using Photon Unity Networking [71] through a computer running Unity v2018.3.27f1 [88] SteamVR [87]. The computer used to host the SteamVR system was a Dell Alienware Aurora R12 [2] with an Intel Core i7 11700F CPU running at 2.5 GHz and an NVIDIA GeForce RTX 3080 with 16 GB of memory. The system updated 90 frames per second.

4.4.5 Procedure

Each study was approximately 60-90 minutes in duration, with participants being compensated with an Amazon gift card worth $15, with an additional $5 per 30 minutes above 90 minutes in case participants took additional time. Prior to the study, the participants read a consent form approved by the Institutional Review Board, and the researcher obtained verbal consent from the participant. Participants filled out a demographic form that asked their age, prior experience with video games and mobile games, each on a scale from 1 to 7 (1 = *"no experience"*, 7 = *"very experienced"*). The participants were given a training task comprising 10 trials to understand the translation method. The translation method was

chosen according to a Complete Latin Square. Participants were informed that their goal was to translate the ball to make it touch the goal object. After successful completion of a translation, a chime could be heard, and a button displaying ”Next” was enabled, allowing the user to immediately begin the next trial. For every trial, the movable object, a ball with a radius of 0.25 m, appeared at the same position (as seen in Figure 4.3a).

For User Study 1, participants stood in a marked position facing the movable object and were told not to walk throughout the condition duration. For User Study 2, participants were asked to begin each trial in a marked position facing the movable object and were encouraged to walk.

Participants had a maximum time of 90 seconds to complete a given trial. If this maximum time was reached, an audio cue could be heard and the next trial would begin. After completion of the practice trials, the participants moved onto performing the recorded trials. In total, each condition consisted of 55 trials, the first one of which was discarded to account for the starting time of the application and did not effect the remaining 54 trials. After each trial, participants could take an optional break in case fatigue was felt. The 10 training trials and the 54 recorded trials were then repeated for the remaining translation methods. The trial order was random for all participants.

After completing a translation method, the user would fill out a questionnaire to assess the usability and perceived workload of the translation type.

After the completion of all three translation methods, the users would then be given a questionnaire discussing overall preferences and rationale behind rankings.

4.4.6 Participants

For User Study 1, we recruited 30 participants with ages ranging from 19 to 36, with an average of 26.33 and a standard deviation of 4.33. The participants' ratings on their experience with video games averaged 5.47, with a standard deviation of 1.94. For their experience with mobile games, the participants' ratings averaged 4.4 with a standard deviation of 1.59. Both ratings had a scale from 1 to 7.

For User Study 2, we recruited another 30 participants with ages ranging from 18 to 36, with an average of 25.67 and a standard deviation of 4.44. The participants' ratings on their experience with video games averaged 5.27 with a standard deviation of 2.00. For their experience with mobile games, the participants' ratings averaged 4.30 with a standard deviation of 1.58.

No one participated in both User Study 1 and User Study 2.

4.5 Results

4.5.1 Objective Measures

For each user study, we collected 4860 data points (3 translation techniques $\times$ 3 target sizes $\times$ 3 distances $\times$ 6 trials $\times$ 30 participants). We conducted repeated measures ANOVA to compare the effects of translation techniques, target sizes, and distances on completion time, fine-tuning phase, and user movement. When the sphericity assumption was violated in data, we applied Greenhouse-Geisser correction for $\epsilon < 0.75$ and Huynh-Feldt correction for $\epsilon \geq 0.75$. We used the Bonferroni correction with an adjusted alpha level of 0.0167. All statistical analyses were performed using SPSS Version 29.0.1.0.

Table 4.1: Results of repeated-measures ANOVA on objective measures. Significant results with $p < 0.05$ are bolded. (Technique: Translation Technique, Size: Target Size, Distance: Target Distance).

	Completion Time			Fine-tuning Phase			User Movement		
User Study 1	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Technique	13.87	**<0.001**	0.32	157.75	**<0.001**	0.85	18.25	**<0.001**	0.39
Size	54.30	**<0.001**	0.65	57.50	**<0.001**	0.67	4.78	**0.01**	0.14
Distance	125.43	**<0.001**	0.81	1.06	0.35	0.04	48.98	**<0.001**	0.63
Technique × Size	1.96	0.14	0.06	6.27	**<0.001**	0.18	8.62	**<0.001**	0.23
Technique × Distance	5.36	**0.01**	0.16	3.69	**0.02**	0.11	16.37	**<0.001**	0.36
Size × Distance	0.86	0.44	0.03	0.88	0.48	0.03	0.64	0.64	0.02
User Study 2	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2
Technique	6.96	**0.01**	0.19	119.23	**<0.001**	0.80	20.15	**<0.001**	0.41
Size	22.47	**<0.001**	0.44	57.37	**<0.001**	0.66	12.27	**<0.001**	0.30
Distance	195.70	**<0.001**	0.87	3.22	0.06	0.10	2.20	0.13	0.07
Technique × Size	2.81	**0.04**	0.09	0.69	0.60	0.02	7.71	**<0.001**	0.21
Technique × Distance	3.19	0.06	0.10	3 .09	**0.04**	0.10	1.84	0.16	0.06
Size × Distance	0.31	0.87	0.01	0.73	0.58	0.03	0.63	0.58	0.02

Table 4.2: Conditional mean and standard deviation of objective measures. (Technique: Translation Technique, Size: Target Size, Distance: Target Distance)

	User Study 1			User Study 2		
	Completion Time	Fine-Tuning Phase	User Movement	Completion Time	Fine-Tuning Phase	User Movement
Technique						
`3DSlide`	8.03 (6.37)	23.92% (12.39%)	0.056 (0.005)	9.35 (8.78)	23.32% (13.15%)	0.074 (0.007)
`VirtualGrasp`	6.94 (6.00)	44.54% (18.73%)	0.086 (0.005)	8.13 (7.49)	40.75% (17.36%)	0.153 (0.015)
`Joystick`	5.96 (3.56)	28.22% (14.39%)	0.067 (0.005)	6.89 (4.76)	26.66% (15.14%)	0.089 (0.007)
Size						
Small	7.58 (5.50)	34.94% (18.53%)	0.067 (0.004)	9.09 (8.26)	32.93% (18.34%)	0.100 (0.007)
Medium	7.07 (6.03)	32.09% (17.23%)	0.071 (0.004)	8.07 (7.51)	30.21% (16.89%)	0.107 (0.007)
Large	6.28 (4.88)	29.66% (17.14%)	0.071 (0.004)	7.16 (7.83)	27.60% (15.47%)	0.109 (0.007)
Distance						
1 m	5.24 (4.18)	32.43% (19.87%)	0.082 (0.005)	6.17 (6.32)	29.84% (18.54%)	0.110 (0.007)
2 m	6.17 (4.25)	31.75% (16.18%)	0.067 (0.004)	7.49 (6.87)	29.82% (15.90%)	0.106 (0.008)
4 m	9.52 (6.76)	32.50% (17.06%)	0.060 (0.003)	10.66 (7.83)	31.08% (16.65%)	0.100 (0.008)

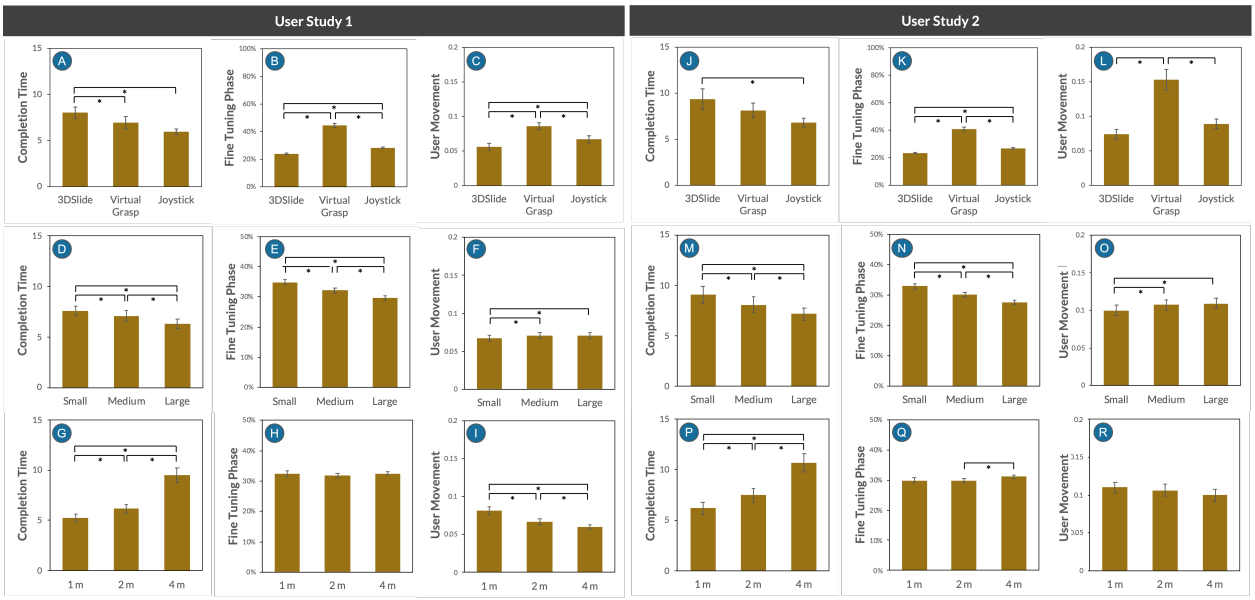

Figure 4.5: Objective measures summary showing the main effects of each independent variable. The first three columns are from User Study 1, with the first column representing completion time (a,d,g), the second column representing fine-tuning phase (b,e,h), and the third column representing user movement (c,f,i) with the first row showing each translation technique (a-c), the second row showing each target size (d-f), and the third row showing each target distance (g-i). The next three columns are from User Study 2 with the same layout. Error bars represent standard error. Asterisk ($*$) indicates a statistically significant difference between the two conditions with Bonferroni correction: $p < 0.05/3 = 0.0167$.

Completion Time

As shown in Table 4.1, all three studied variables had significant main effects on the completion time in both User Study 1 and 2. Among the translation techniques, `3DSlide` took a significantly longer time to complete tasks. In terms of size and distance, completion time increased when the target was smaller and located at longer distances, following the relationship defined by Fitts' Law. For User Study 1, an interaction effect between the translation technique and distance was detected (see Figure 4.7a). While the overall trend of the longer distance taking longer to complete was identical across all techniques, and `3DSlide` being the slowest were consistent across all distances, there were disparities in whether the differences were statistically significant. For example, for 1 m distance, while `3DSlide` was slower

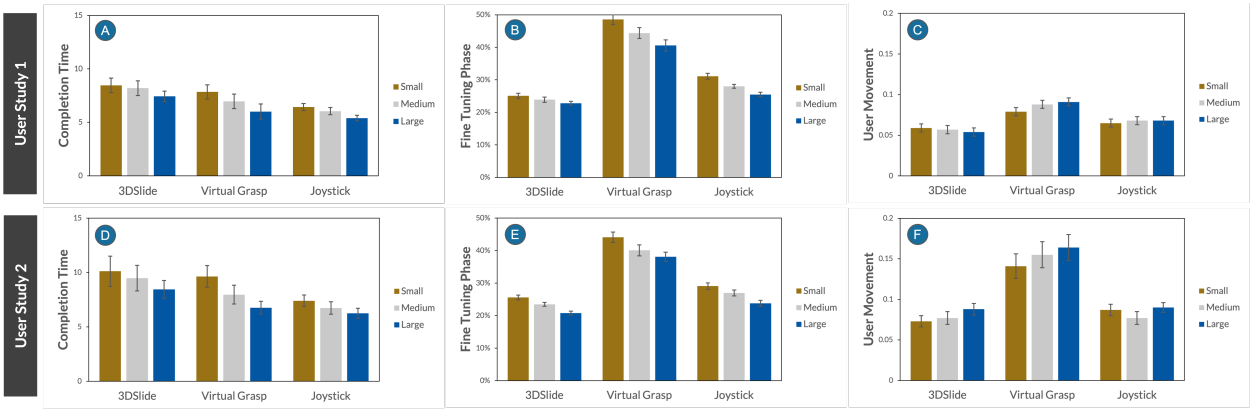

Figure 4.6: Objective measures for combinations of translation technique and target size. The top row indicates results from User Study 1: (a) Completion time, (b) Fine-tuning phase, and (c) User movement. The bottom row indicates results from User Study 2: (d) Completion time, (e) Fine-tuning phase, and (f) Use movement. Error bars represent a standard error.

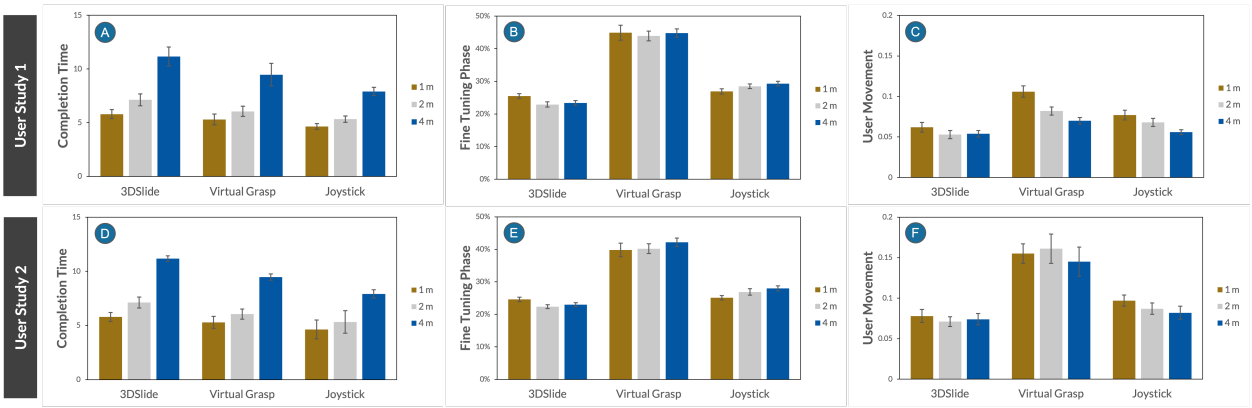

Figure 4.7: Objective measures for combinations of translation technique and distance. The top row indicates results from User Study 1: (a) Completion time, (b) Fine-tuning phase, and (c) User movement. The bottom row indicates results from User Study s: (d) Completion time, (e) Fine-tuning phase, and (f) User movement. Error bars represent a standard error.

than `VirtualGrasp`, and `VirtualGrasp` was slower than `Joystick`, the differences were not statistically significant; and the same was between `VirtualGrasp` and `Joystick` for 4 m distance. For User Study 2, when participants were allowed to walk around the room, an interaction effect between the translation technique and size was observed (see Figure 4.6d). Again, the overall trend of the smaller size taking longer time across all techniques, and

`3DSlide` being the slowest across all sizes were identical. However, when the target size was small, `VirtualGrasp` was not significantly faster than `3DSlide`; and when the target size was large, `VirtualGrasp` was not significantly slower than `Joystick`.

Fine-Tuning Phase

For both User Study 1 and 2, translation technique and target size showed significant main effects on the proportion of time spent on fine-tuning. Among the translation techniques, `VirtualGrasp` required a significantly greater proportion of time to fine-tune the object before it reached the target. While `3DSlide` and `Joystick` spent around 20-30% of the total time in the region of the last 20% of total distance, `VirtualGrasp` spent more than 40% in the same region. The final adjustment also required greater effort for targets with smaller sizes.

While there was no main effect observed for target distance, an interaction effect was identified between target distance and translation technique for both user studies (see Figure 4.7b and 4.7e). Regardless of whether the participants were stationary or movable, `3DSlide` had the most substantial fine-tuning phase when the target was at a 1 m distance; in contrast, `Joystick` had the smallest fine-tuning phase at this distance. However, for `VirtualGrasp`, no significant differences were found among the distances. The overall trend of `VirtualGrasp` requiring the largest fine-tuning phase was consistent across difference distances.

In User Study 1, an interaction effect was observed between translation technique and target size (see Figure 4.6b). The general trend of smaller target sizes requiring larger fine-tuning phase was consistent across translation techniques. Also, the finding that `VirtualGrasp` required the largest fine-tuning phase remained consistent regardless of target sizes. However, there were variations in whether these differences were statistically significant. Specifically,

for both `VirtualGrasp` and `Joystick`, it was statistically significant that smaller target sizes required a larger fine-tuning phase when compared to larger target sizes, with differences between all pairs being statistically meaningful. In case of `3DSlide`, the small-sized target required a significantly larger fine-tuning phase compared to the large-sized target. However, the differences between small-sized and medium-sized targets, as well as between medium-sized and large-sized targets, did not reach statistical significance.

User Movement

In User Study 1, significant main effects were observed for all variables: translation technique, target size, and distance. Among the translation techniques, `VirtualGrasp` resulted in the most active movement, followed by `Joystick`, and then `3DSlide`. Regarding target size and distance, participants moved less when dealing with small-sized targets compared to medium-sized and large-sized targets, as well as at longer distances compared to shorter distances.

User Study 2 produced similar result in trends, but with a higher overall activity level. Regardless of whether participants were allowed to move or not, `VirtualGrasp` consistently led to the most active movement during the task. The trend of small targets resulting in lower activity levels, as well as targets located at longer distances leading to reduced activity, remained consistent. However, in User Study 2, the variance in distances had no statistically significant impact on movement levels.

Interaction effects between translation technique and size were observed in both User Study 1 and 2 (see Figure 4.6c and 4.6f). In both studies, the amount of movement among different target sizes was not significantly different for `3DSlide` and `Joystick`. However, for `VirtualGrasp`, both user studies consistently showed lower activity levels when targets were

in smaller sizes.

In Study 1, an interaction effect between translation technique and distance was also identified (see Figure 4.7f). The overall trend was consistent with the main effects observed or translation technique and distance. Specifically, `VirtualGrasp` was consistently associated with the highest level of activity, while targets located at greater distances resulted in lower activity levels. However, there were variations in whether these differences reached statistical significance. While differences in most pairs of different distances were significant, there was no significant difference in activity levels between targets placed at 2 m and 4 m for `3DSlide`.

4.5.2 Subjective Measures

We conducted repeated measures ANOVA to compare the effects of translation techniques on usability and perceived workload for each user study. When the sphericity assumption was violated in data, we applied Greenhouse-Geisser correction for $\epsilon < 0.75$ and Huynh-Feldt correction for $\epsilon \geq 0.75$. The Chi-square goodness of fit test was used to confirm statistically significant variance in the distribution of preference and vote for the most easy to learn, intuitive, and fun. Comments collected during exit interviews were categorized using an Affinity Diagram [45].

Usability

As shown in Table 4.3, translation technique had a significant effect on both manipulability and comprehensibility in User Study 1 and 2. Manipulability assessed the physical comfort and ease of operation of the device and user interface. Regardless of whether participants were stationary or movable, `Joystick` was consistently rated as the most comfortable and

Table 4.3: Conditional mean, standard deviation, and results of repeated-measure analysis on subjective measures. Significant results with $p < 0.05$ are bolded.

User Study 1	3DSlide	Virtual Grasp	Joystick	F	p	η_p^2
Usability						
Manipulability	4.35 (0.25)	3.63 (0.24)	5.03 (0.23)	18.48	**<.001**	0.39
Comprehensibility	5.64 (0.19)	5.77 (0.19)	6.18 (0.11)	4.08	**0.02**	0.12
Workload						
Mental Demand	2.70 (0.31)	3.60 (0.29)	2.23 (0.23)	9.16	**<.001**	0.24
Physical Demand	3.57 (0.32)	4.33 (0.30)	2.97 (0.27)	15.26	**<.001**	0.35
Temporal Demand	3.37 (0.32)	3.57 (0.34)	3.37 (0.32)	0.30	0.68	0.01
Performance	5.90 (0.21)	5.50 (0.22)	6.23 (0.15)	6.19	**0.004**	0.18
Effort	3.33 (0.30)	4.33 (0.26)	2.70 (0.27)	13.52	**<.001**	0.32
Frustration	2.77 (0.36)	3.57 (0.32)	2.37 (0.27)	7.47	**0.003**	0.21
User Study 2	3DSlide	Virtual Grasp	Joystick	F	p	η_p^2
Usability						
Manipulability	4.36 (0.22)	4.14 (0.24)	5.08 (0.18)	7.37	**0.001**	0.20
Comprehensibility	5.58 (0.24)	5.83 (0.19)	6.23 (0.14)	5.02	**0.01**	0.15
Workload						
Mental Demand	2.87 (0.31)	3.20 (0.33)	2.6 (0.27)	1.87	0.16	0.06
Physical Demand	3.07 (0.31)	4.43 (0.30	3.03 (0.32)	8.79	**<.001**	0.23
Temporal Demand	3.30 (0.30))	3.43 (0.30)	3.40 (0.32)	0.13	0.87	0.01
Performance	6.03 (0.19)	5.77 (0.20)	6.333 (0.14)	3.51	**0.04**	0.11
Effort	3.50 (0.35)	3.80 (0.31)	3.03 (0.29)	3.56	**0.04**	0.11
Frustration	3.17 (0.32)	3.27 (0.34)	2.53 (0.32)	3.09	0.05	0.10

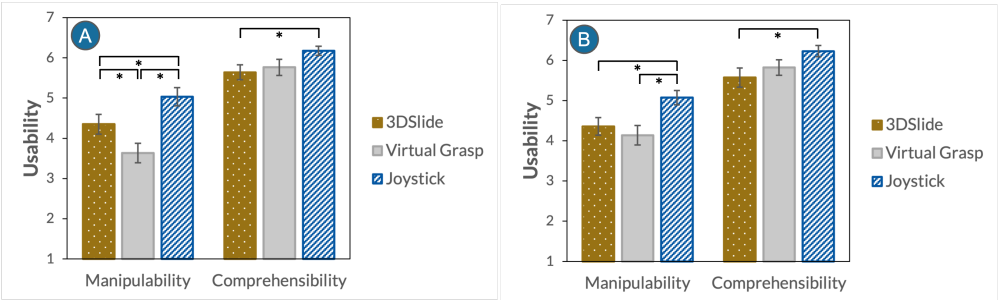

Figure 4.8: Mean usability score for each translation technique. (a) User Study 1. (b) User Study 2. Asterisk ($*$) indicates a statistically significant difference between the two conditions with Bonferroni correction: $p < 0.05/3 = 0.0167$

straightforward to use. In User Study 1, `3DSlide` received the second-highest score, followed by `VirtualGrasp`, with statistically significant differences between all pairs. In Study 2, although the average score for `VirtualGrasp` was lower than that of `3DSlide`, the difference did not reach statistical significance.

Comprehensiblity assessed factors such as response speed, ease of reading and understanding, and consistency of the interface. Overall, all three translation techniques received higher scores than neutral, indicating positive evaluations in terms of comprehensibility. Once again, `Joystick` received the highest score, while `3DSlide` received the lowest score. However, there were no significant differences between `Joystick` and `VirtualGrasp`, as well as between `3DSlide` and `VirtualGrasp` in terms of comprehensibility.

Workload

Figure 4.9 shows the results of participants' perceived workload, assessed using the NASA-TLX questionnaire, which consists of six subsets: mental demand, physical demand, temporal demand, performance, effort, and frustration. Across all subsets and user studies, `VirtualGrasp` consistently received the highest scores for demand and the lowest score for perceived performance. In User Study 1, `VirtualGrasp` was rated to be statistically more complex (mental demand), strenuous (physical demand), demanding (effort), and stressful (frustration) in comparison to one or both of the other translation techniques. Additionally, participants expressed lower satisfaction with the performance of `VirtualGrasp`. Temporal demand, which gauged the level of time pressure experienced by participants, did not exhibit any significant differences among translation techniques, with scores indicating neutral responses.

User Study 2 yielded a similar trend, with `VirtualGrasp` being evaluated as significantly

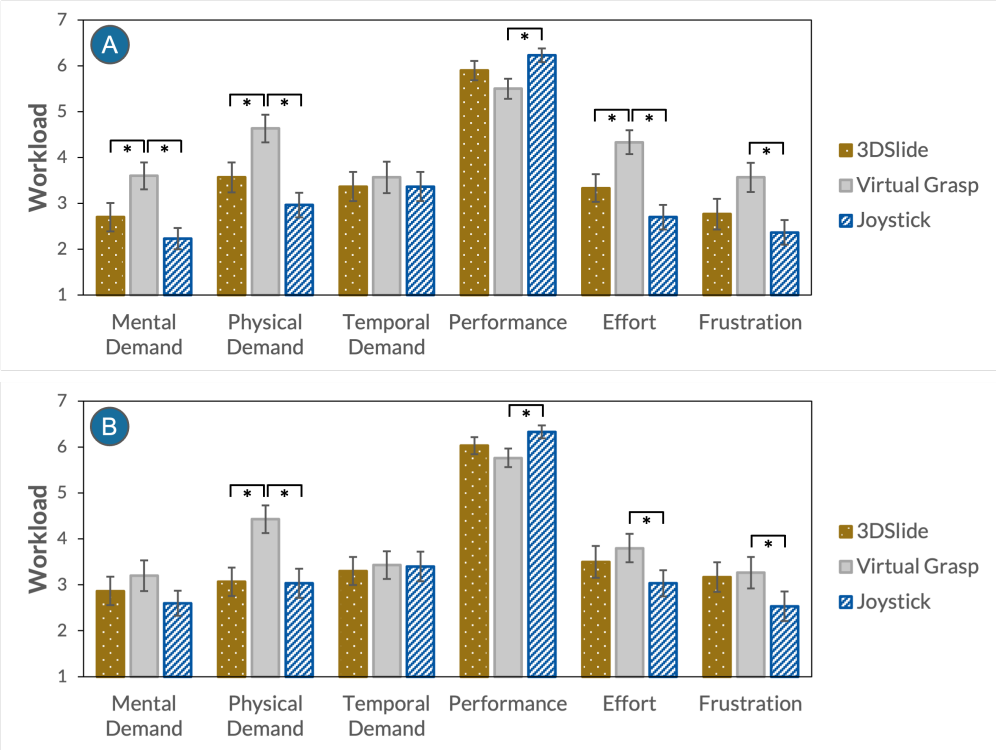

Figure 4.9: Mean workload score for each translation technique. (a) User Study 1. (b) User Study 2. Asterisk ($*$) indicates a statistically significant difference between the two conditions with Bonferroni correction: $p < 0.05/3 = 0.0167$

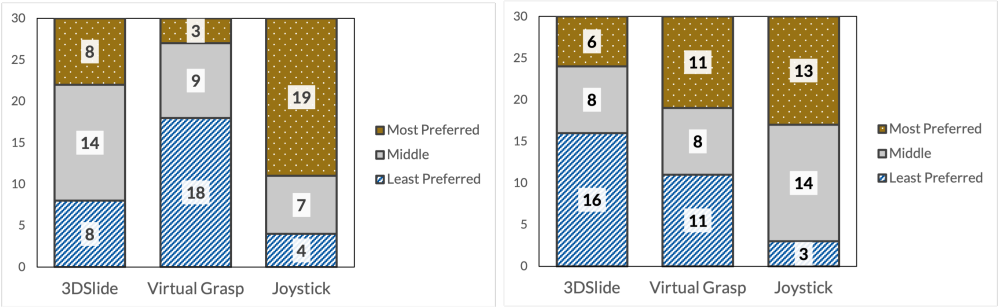

Figure 4.10: Preference rank distribution for each translation technique: from User Study 1 with fixed position (left) and from User Study 2 with movable condition (right).

more strenuous (physical demand), demanding (effort), and stressful (frustration) with lower satisfaction (performance) compared to one or both other translation techniques. One notable result was the mental demand, which assessed whether participants found the condition to be easy and simple or demanding and complex. In contrast to User Study 1, `VirtualGrasp` was no longer significantly more demanding in terms of complexity (mental demand).

Preference

Figure 4.10 shows the distribution of preference ranks for translation techniques. A Chi-square goodness of fit test showed statistically significant discrepancies in the number of votes for both User Study 1 ($\chi^2(2) = 10.94, p = 0.004$) and User Study 2 ($\chi^2(2) = 9.25, p = 0.026$). In User Study 1, `Joystick` received the highest number of votes to be the most preferred translation technique. While `Joystick` maintained its first place for getting the highest number of votes for being the most preferred technique, it is noteworthy that in User Study 2, `VirtualGrasp` received far more votes than it did in Study 1, when participants were restrained in a stationary location.

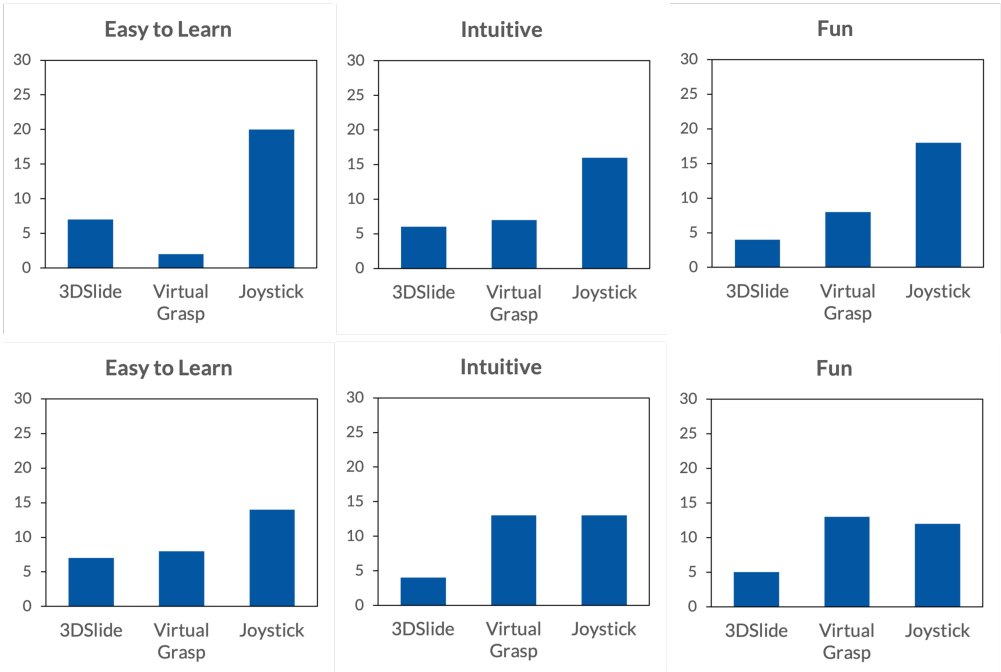

Figure 4.11: Distribution on votes for the most easy to learn, intuitive, and fun among the three translation techniques: from User Study 1 with fixed position (top row) and from User Study 2 with movable condition (bottom row).

Exit Interview

In addition to general preference, participants were asked to vote for the method they thought was the most intuitive, easy to use, and fun (see Figure 4.11). In User Study 1, all three components showed statistically significant variances in the distribution of the votes (Easy to learn: $\chi^2(3) = 25.75, p =< 0.001$; Intuitive: $\chi^2(3) = 11.75, p = 0.008$; Fun: $\chi^2(3) = 19.00, p =< 0.001$) with `Joystick` receiving the highest number of votes. Following `Joystick`, `VirtualGrasp` received more votes than `3DSlide` for being intuitive and fun, while `VirtualGrasp` was rarely voted to be the most easy to learn.

User Study 2 showed distinctive results from User Study 1. For easiness to learn, `Joystick` received the most votes, but failed to reach statistical significance to reject the null hypothesis of equal distribution across all techniques ($\chi^2(3) = 7.75, p = 0.05$). Also, `VirtualGrasp`

Table 4.4: Equations and R^2 values from Fitts' Law and its extended model.

Model	Equation	ID	3DSlide	VirtualGrasp	Joystick
Fitts'[24]	$MT = a + b \cdot ID$	$\log_2(\frac{2A}{W})$	0.54	0.60	0.56
Shannon's [57, 84]	$MT = a + b \cdot ID$	$\log_2(\frac{A}{W} + 1)$	0.54	0.60	0.56
Hoffman's [38]	$MT = a + b \cdot ID$	$\log_2(\frac{2A}{W+F})$	0.710	0.65	0.66
Murata and Iwase's [66]	$MT = a + b \cdot \sin\theta + c \cdot ID$	$\log_2(\frac{A}{W} + 1)$	0.58	0.69	0.68
Cha and Myung's [15]	$MT = a + b \cdot \theta_1 + c \cdot \sin\theta_2 + d \cdot ID$	$\log_2(\frac{2A}{W+F})$	0.75	0.76	0.83

received equal number of votes as `Joystick` for being the most intuitive, and one more vote than `Joystick` for being the most fun.

An affinity diagram was made to document the reasons participants justified for their decisions for preferring and disliking their chosen conditions. In this affinity diagram, a participant's response could be mapped onto multiple reasons, depending on the statements provided.

4.5.3 Fitts' Law

To apply Fitts' Law and its extended models to our data, we calculated the average completion time across participants from User Study 1 for each unique combination of variables: target size, distance, azimuth angle, and inclination angle. This process provided us with 54 data points for each translation technique, which were then used for model fitting. We employed five models [15, 24, 38, 57, 66, 84], as outlined in Table 4.4. For Fitts' and Shannon's, the models aimed to predict completion time (MT) based on the index of difficulty (ID) determined by target size (W) and distance (A). For Fitts', a total of 7 IDs were calculated, ranging from 3 to 6.73, and for Shannon's, a total of 7 IDs, ranging from 2.32 to 5.76. Hoffman's model incorporated an additional factor, finger pad size (F). In this study, we referred to the size of the movable object. A total of 9 IDs were calculated, ranging from 2 to 6.74. Murata and Iwase's model employed the same definition of ID as Shannon's but included the sine of azimuth angle as an additional factor. Lastly, Cha and Myung's model

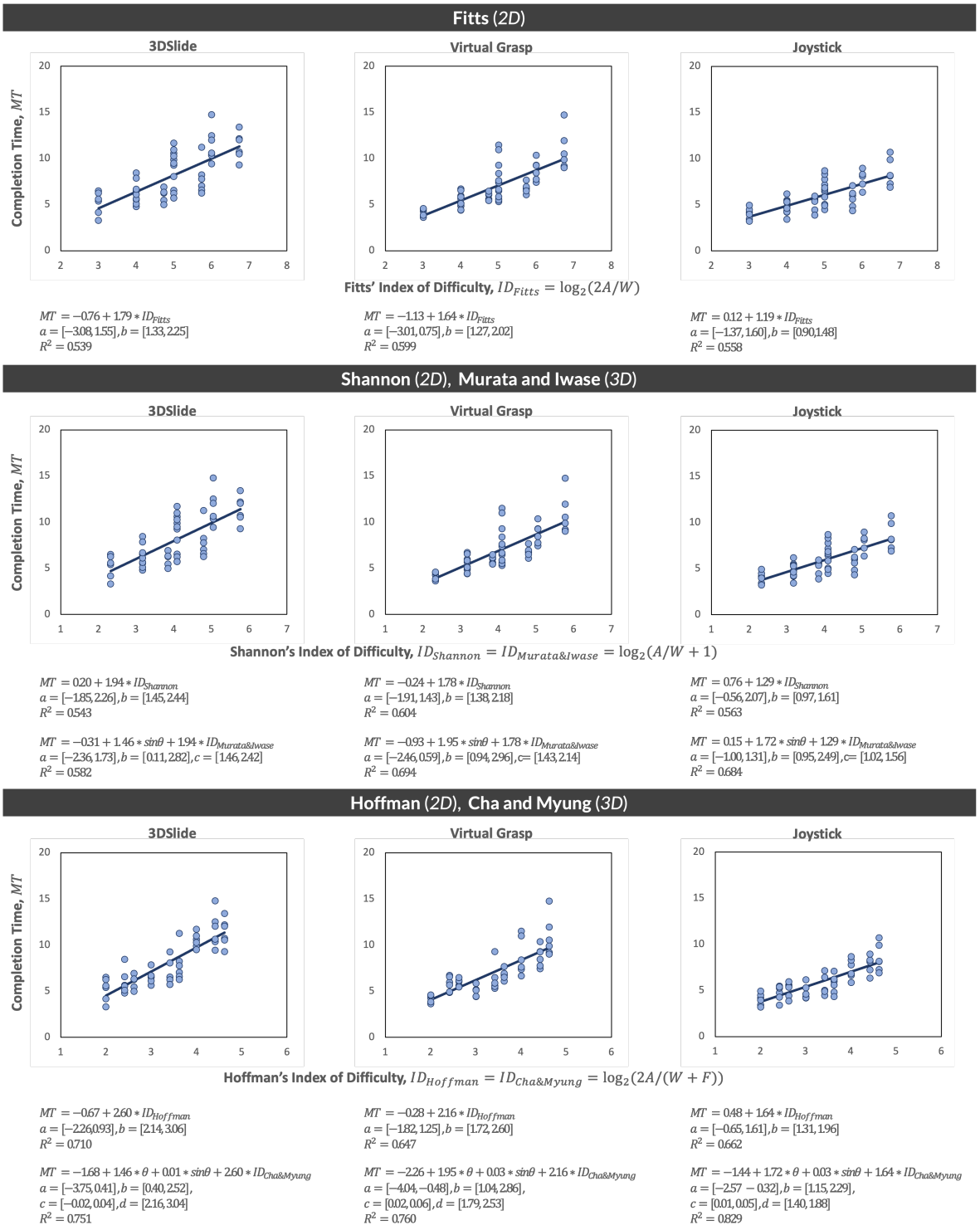

Figure 4.12: The linear regression equation, the coefficients' lower and upper bound confidence intervals of the constants (CI=95%), and R^2 for the fitted Fitts' Law and related models.

integrated both the sine of azimuth and the inclination angles, and employed the same ID as Hoffman's. Although the 2D models do not consider angles, we opted to use the same number of data points for these models instead of averaging across unique pairs of target size and distance. This decision was made to maintain consistency and ensure fair comparison across the models. Using the same number of data points for all models is important because the number of repetitions embedded in a single data point significantly impacts the explainability (R^2) of the model, as discussed by Triantafyllidis and Li [91].

Figure 4.12 presents the R^2 values along with the linear regression equations, including the coefficients' lower and upper bound confidence intervals (CI=95%), for each fitted model. As shown in Table 4.4, Cha and Myung's model demonstrated the best fit for all three translation techniques, while Fitts' and Shannon's models yielded the lowest fit. Specifically, for `VirtualGrasp` and `Joystick`, models that incorporated angle(s) proved to be more effective at explaining the data compared to the 2D models. Interestingly, for `3DSlide`, Hoffman's model emerged as the second-best performer in explaining the data, showcasing its unique characteristics in this context.

Our study's R^2 values ranged from the lowest at 0.54 to the highest at 0.83. These values may appear relatively low when compared to the typical values reported in the literature in this field. This discrepancy could be attributed to the limited number of repetitions embedded in each data point, as each participant completed a single repetition. These values are expected to improve with greater repetitions [91].

4.6 Discussion

We designed an experiment to compare three different translation techniques to evaluate object translations in room-scale environments. To do so, the translation techniques were

evaluated at a variety of target location distances and target sizes. The results show that the translation was completed the fastest with the `Joystick` condition, with `3DSlide` requiring the least amount of fine-tuning and `VirtualGrasp` resulting in the most amount of user movement. Participants overall preferred `Joystick` the most, however, `VirtualGrasp` was significantly more preferred in User Study 2 when movement was allowed as opposed to when it was not allowed in User Study 1.

Limiting simultaneous DoF control helps to control precision

Given the amount of time spent in the fine-tuning phase was smaller for `3DSlide` and `Joystick` as compared to `VirtualGrasp`, it is evident that objects are easier to control when limiting the DoF users can control simultaneously. This does, however, come at a trade-off, given that `3DSlide` was the most restrictive, and therefore, users had to spend the least amount of time trying to be accurate with the position of the movable ball, and had to spend more time trying to get the movable ball to the target location. `Joystick` offers a good compromise to for limiting simultaneous DoF control, and precision control. This was because of the design choice to limit movement to a singular plane or axis, depending on input. This evidence is supported by Benko and Feiner [8], given that their evaluated technique of `Balloon Selection`, performed best. `Balloon Selection` was a 3 DoF positioning technique that restricted its axes to either 2 DoF or 1 DoF translation, similar to the movement allowed from the `Joystick` method. Participants also agreed with this, as P-15 stated when discussing `Joystick`, *"it allows me to reach the destination diagonally in one plane. It feels faster than having to navigate the coordinates x, y and z one by one."* further facilitating this claim with `Joystick`.

Users also preferred `3DSlide` in comparison to `VirtualGrasp`, as `3DSlide` was reported by participants to be more precise because of the individual axis controls that were offered to users with statements such as P1-29 saying, *"[3DSlide], because manipulating each axis*

individually offered more preciseness in moving the object around." and P2-21 stating, *"I preferred [3DSlide] the most because it was the one that was easiest to control and offer the most precise control."* supporting this claim.

For `VirtualGrasp`, participants spent more than 40% of their movement time for both static and movable conditions in the fine-tuning phase of their manipulation. Evidence against `VirtualGrasp`'s axes control was prevalent, with P1-21 claimed they disliked it the most *"because all three axis responded simultaneously, I was putting most physical efforts to maintain the stable position."* and P2-19 claiming *"[Virtual Grasp]: I felt very unintuitive, moving in all 3 axes at the same time was difficult."*.

Device-based methods can be physically fatiguing for mobile VR, especially when movement is restricted

For `VirtualGrasp`, the main complaint gathered from both user studies was that the tablet would have to be held up for a significant amount of time, causing fatigue to the participants. `VirtualGrasp` received a significantly lowest score in manipulability in Study 1, where participants were asked to deliberately not walk. In this condition, participants were required to tilt the tablet and hold steady to accurately pinpoint the target location from a fixed distance, as opposed to adjusting perspective and distance toward the target, allowing for more leniency in their accuracy. This fatigue is quite common in device-based methods [8, 28, 96] as users have to hold their arms steady to "aim" at the target correctly. In `VirtualGrasp`, given that the device's position and the movable object's position would be linked, any shifts in perspective when holding the tablet will move the object according to distance in relation to the azimuth and inclination angle of the tablet. Participants shared a similar sentiment, with P1-17 stating *"I didn't like the [Virtual Grasp] ... I had to be very steady while grabbing the object otherwise, if the tablet moved even slightly, I would lose the correct position; I felt it needed too much physical effort compared to the other two"* and P2-2 *"it was difficult*

and demanding to hold the tablet". A possible explanation for the disparity between HWD VR and mobile VR in device-based movements is that the camera and the hand are coupled to one another in mobile VR, yet independent in HWD VR. This contrasts interactions in HWD VR, as in HWD VR, device-based methods are oftentimes the easiest to control and intuitive methods to control translation in VR, with Homer and other controller grasping metaphors being quite common [11, 42]. This makes device-based interaction in HWD VR much easier to select and manipulate rather than mobile AR/VR.

People overall preferred less physical movement, with `Joystick` ranking least in mental demand, physical demand, effort, and frustration, with the best-perceived performance based on NASA-TLX results. Participants also claimed to have liked `Joystick` because of the lack of physical movements required to complete a translation with P1-1 saying *"I prefer the [Joystick] method as it does not require me to physically move..."* and P2-20 saying *"it was the easiest to quickly get the answer with little to no movement"*, whereas `VirtualGrasp` required the most mental demand, physical demand, effort, and frustration, with the worst perceived performance.

According to preference results, more participants preferred `VirtualGrasp` in User Study 2, when movement was allowed, and overall, the positive impressions increased in User Study 2. Evidence can be seen in User Study 1 as `VirtualGrasp` had a significantly higher mental demand and frustration compared to the other two techniques, whereas this was not the case in User Study 2. More participants also described `VirtualGrasp` as intuitive in Study 2 than in Study 1. The results of Mossel et al. indicated that users preferred HOMER-S for object translation tasks as opposed to 3DTouch in 3D translation and 2D translation tasks, given that it is similar to the real-world metaphor of translation [63]. As `VirtualGrasp` was derived from HOMER-S, and that the original evaluation of HOMER-S allowed for participant movement [63], device-based methods are preferable in environments in which

movement is allowed, given that participants are able to use this real-world metaphor of translation. The results from our study align with the room-scale mobile AR evaluation from Hellmuth et al. [36], given that their device-based manipulation technique performed the slowest and was the most inaccurate.

An explanation of this can be derived from Jacob et al. [41]. The mechanism for `VirtualGrasp` does not align with the perceptual structure of the task as it is difficult to maintain device orientation to position the movable object. However, in User Study 2, the physical structure of the task allowed users to move, making the translation's perceptual structure more apparent.

These observations from `VirtualGrasp` appear to defy Hinckley et al. [37], as integration of 3 DoF control scheme should align for a 3 DoF translation task. However, given that at a larger distance, changes in device orientation would significantly effect the movement, the separation of the task's perceptual structure would result in the translation not being completed accurately, resulting in a higher completion time. The theory of Hinckley et al. [37] however can be seen in `Joystick` and `3DSlide`, as the higher simultaneous control in DoF from `Joystick` made the translation more apparent, given the lower separability of the task.

When presented with strong depth cues, users prefer not to move

Results across User Study 1 and 2 remained fairly similar for the `3DSlide` and `Joystick` conditions, as participants typically chose not to move, even when they could move for translation tasks in Study 2. Workload and preference were similar across Study 1 and Study 2 for `3DSlide` and `Joystick`. The participants in User Study 2 frequently cited the usage of the depth cues as their excuse not to move, as they felt the need not to move as they understood the position of the virtual objects well enough and that a shift in perspective

would not have aided significantly in the understanding in the position of the objects. These sentiments are shared by the participants, with P1-9 stating for `3DSlide` *"It was very easy to slide the ball to the other one's shadow through use of the X and Z swipe, then using Y to locate the ball"* and P2-7 stating *"[Joystick]: It felt more intuitive, going towards the shadow with joystick was easier"*.

When applying to mobile AR techniques, depth cues from shadows are important to allow for aligning the user's perceptions on distance of virtual objects [95]. As a result, in larger, room-scale environments, users are likely to rely on depth cues when performing translations in both mobile AR and VR.

Users prefer familiar and intuitive interfaces

`Joystick` was the most preferred method when it came to participants' rankings when compared to `3DSlide` across both user studies. In User Study 1, `Joystick`'s intuitiveness was more pronounced in contrast to the device-based `VirtualGrasp` method, making Joystick more preferred over `VirtualGrasp`. The participants' preference rankings reflected their performance, given that `Joystick` performed the fastest when considering completion time. A possible reason for this explanation is that the `Joystick` implementation closely follows ones that are on most mobile device applications and on gaming consoles. Familiarity could play a significant factor when considering interacting and using a UI in a different setting. This reason is compounded by some of the participant statements, with P1-7 stating *"The Joystick felt the most familiar"* and when discussing which method they preferred the most, P1-10 wrote *"Joystick. It's similar other precedent games."* and P1-30 saying Joystick was *"Intuitive"*.

Applying Fitts' Law on Mobile Virtual Reality

Overall, we observed that the impact of task variables on completion time followed the

fundamental principles of Fitts' law for all three translation methods, with greater completion time when target sizes were smaller sizes and when the distance to the target was greater.

At the outset of this study, we posed the question of whether the translation task in mobile VR would be better explained in 2D or 3D settings, given the 2D input system and 3D task environment. After fitting our data to the original Fitts' law and its extended models, we found that the model incorporating both azimuth and inclination angles, within a 3D spherical coordinate system, provided the most effective explanation of the relationship between the studied variables and completion time for all three techniques.

While `VirtualGrasp` and `Joystick` benefited from 3D models that outperformed models based solely on target size and distance, `3DSlide` demonstrated its second-best fit with Hoffman's model, which does not consider angles. This suggests that the impact of angles on predictability was relatively lower for `3DSlide`. This the unique characteristics of the `3DSlide` method, where object translation was limited to 1 DoF at a time. This characteristic resembles the integration of multiple 1D translation tasks. In summary, our findings emphasize the importance of understanding both the environment and the characteristics of the translation technique to gain insight into the underlying dynamics of the task.

Chapter 5

Conclusion

In this chapter, the results of the two systems and their corresponding user studies will be summarized with respect to answering their corresponding research questions.

5.1 Layout Dimensionality

To answer **RQ-1** a user study was performed to analyze key differences of layout dimensionalities across collaborators of different levels of immersion.

To this effect, an empirical study was performed that utilizes both quantitative and qualitative analysis to examine the impact of different designs of asymmetric collaborative systems between PC and VR on the collaborative decision-making process. The study focuses on a classic data-driven collaborative task, specifically the hotel search task, and employs a small-multiple design to present the hotel information. The small-multiple layout varies in dimensionalities across different conditions, namely `PC2D+VR2D`, `PC2D+VR3D`, and `PC3D+VR3D`. To facilitate task completion, semantic zooming and awareness support was incorporated in both environments. Findings indicate that optimizing the individual environment enhances satisfaction and improves engagement. On the other hand, interfaces that prioritize reducing collaborative cost lead to faster completion times. However, we observed that navigating a 3D environment on a PC was ineffective for our tested task, although no similar trend was observed in VR. We believe that our results contribute to an empirical understanding of the

collaborative experience in asymmetric collaboration and can serve as inspiration for future designs.

5.2 Manipulation in Mobile VR

To answer **RQ-2** two user studies were performed to analyze how people interacted with mobile devices inside of a VR environment.

Three translation methods were investigated to understand object translation in room-scale, mobile VR environments. Through two user studies, it was determined that restrictions on simultaneous DoF control were useful to increase the accuracy of a translation, however, too much restriction would result in slower manipulations. This trade off of speed also came at the cost of accuracy, as too high simultaneous DoF control was difficult to control and resulted in lowering completion time. Similarly, too low simultaneous DoF control took too much time to change and control each axis, despite being highly precise. The implications of this study allow for designers to make insights when creating room-scale mobile VR experiences. These implications also help the designers of mobile AR systems in larger scale environments as mobile computing technology grows to allow for better tracking and interaction capabilities given increased environment size.

5.3 Limitations and Future Work

For layout dimensionality, different combinations of dimensionalities were tested for PC+VR asymmetric collaborative decision-making. In order to gain a nuanced understanding of PC+VR collaboration, we adapted a well-established hotel search task [43]. Given the nature of this task, we employed a small-multiple design, where all windows appeared identical ex-

cept for displaying distinct hotel information. In more complex data-driven decision-making scenarios, collaborators may require access to diverse information presented in different visual formats, such as a dashboard. Although we believe that many of our findings could be applied to such scenarios, it is important to note that due to differences in navigation behaviors, we might observe contrasting collaborative actions, especially concerning spatial referencing. Hence, further investigation through future studies is necessary.

Additionally, it is worth mentioning that although the small-multiple layouts were 3D in certain conditions, the visual content within each individual window remained inherently 2D. As a result, our future direction involves examining task scenarios that involve 3D content or a combination of 2D and 3D content. This will allow us to explore the impact of spatial depth and visual immersion on collaborative decision-making in PC+VR environments more comprehensively.

We ensured that our participants had the necessary interactions to successfully accomplish the study task. Nevertheless, we acknowledge the potential benefits of incorporating additional interactions, such as the ability to move and resize windows. By introducing these interactions, we would face new technical challenges, including real-time layout synchronization between 2D and 3D environments. However, such additions would also enhance collaborative behaviors, for instance, allowing for the interactive formation of clusters [19, 55]. Furthermore, as more interactions are introduced, awareness cues will play an even more crucial role, given the increased activity of both users and visual content. Therefore, improving awareness techniques to effectively support more interactive experiences is imperative.

Lastly, our study revealed that employing both 2D and 3D environments within VR presents its own set of advantages and disadvantages. One potential solution is to offer VR users the option to switch between 2D and 3D environments or even implement automatic switching. For instance, using 3D for individual work and transitioning to 2D for collaborative discus-

sions could be beneficial. This approach would allow users to leverage the strengths of both environments based on the specific task or stage of the collaborative process.

For object translation, the study evaluated different translation techniques in room-scale mobile VR and how their design could effect usability in this unique environment. Across each translation technique, many factors were changed that could influence the result of translation. Future work could more directly analyze the broad implications made from the effects of each translation type to precisely determine which design factor influenced the translation.

The environment that was created for this user study could have affected the performance of the translation methods. Participants often noted that the usage of the depth cues made the touch-based interfaces easier to use. Future work could look into different environment designs to understand which environments are better designed for a given translation Future work can also examine other manipulation techniques, such as scaling and rotation.

Future work could analyze other translation techniques to understand how manipulations function in mobile VR. One of which, Gesture-based interactions, were not examined due to current technological limitations, as it is currently unable to integrate the pass-through of objects such as hands. Allowing a user to see their own hands would make gesture input feasible through this technology and is the key avenue that went unexplored in this study that would normally be transferable from AR to VR. This however is not technically feasible currently due to the load required to both use AR and VR features onto a single device paired with pass-through recognition and gesture detection. Future work could integrate this technology and evaluate the feasibility of gestures in a mobile VR environment.

Although beyond the scope of the project, the study could be replicated in an AR environment to evaluate mobile input and compare VR to AR manipulation to assess how well one

can provide a testing ground for the other. This will aid in understanding when manipulations are preferable across mobile AR and mobile VR environments.

5.4 Closing Remarks

This thesis presented design implications and understandings on two fundamental scenarios of cross-device collaboration - layout dimensionalities across differently immersed users and mobile device interaction into VR. With the integration of AR/VR technology, come a host of cross-device compatibility and iteration through collaboration work that is fundamental towards the adoption of immersive tools. One of which is the differences in space that are offered by a 3D environment as opposed to a 2D environment. Leveraging collaborative effectiveness, designers can give leniency towards the flexibility of the 3D space from an immersive environment to compromise in individual effectiveness for task completion. The other is the integration of outside observers into a local instance of VR. Offering a compromise of accuracy and DoF control provides speed in task completion, offering an intuitive middle ground between restricting movement for control and enabling movement for speed. Takeaways from this work allow designers to integrate users of different levels of immersiveness into their AR/VR experiences.

Bibliography

[1] Owlchemy, 2021. URL https://owlchemylabs.com/blog/owlchemy-mobile-spectator-ar-spectator-camera.

[2] Alienware. Alienware aurora r12 gaming desktop | dell usa, 2020. URL https://www.dell.com/en-us/shop/desktop-computers/alienware-aurora-r12-gaming-desktop/spd/alienware-aurora-r12-desktop.

[3] Sriram Karthik Badam, Fereshteh Amini, Niklas Elmqvist, and Pourang Irani. Supporting visual exploration for multiple users in large display environments. In *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 1–10. IEEE, 2016.

[4] Huidong Bai, Lei Gao, Jihad El-Sana, and Mark Billinghurst. Markerless 3d gesture-based interaction for handheld augmented reality interfaces. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*, pages 1–1, 2013.

[5] Till Ballendat, Nicolai Marquardt, and Saul Greenberg. Proxemic interaction: designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces*, pages 121–130, 2010.

[6] Daniel Bambušek, Zdeněk Materna, Michal Kapinus, Vítězslav Beran, and Pavel Smrž. How do i get there? overcoming reachability limitations of constrained industrial environments in augmented reality applications. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 115–122. IEEE, 2023.

[7] Victoria Bellotti and Sara Bly. Walking away from the desktop computer: distributed

collaboration and mobility in a product design team. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 209–218, 1996.

[8] Hrvoje Benko and Steven Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. In *2007 IEEE symposium on 3D user interfaces*. IEEE, 2007.

[9] Dorian Gorgan Bianca-Cerasela-Zelia Blaga. Dar: Implementation of a drone augmented reality video game.

[10] Frank Biocca, Chad Harms, and Jennifer Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. *4th annual International Workshop on Presence*, Jan 2001.

[11] Doug A Bowman and Larry F Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 35–ff, 1997.

[12] Doug A Bowman, Sabine Coquillart, Bernd Froehlich, Michitaka Hirose, Yoshifumi Kitamura, Kiyoshi Kiyokawa, and Wolfgang Stuerzlinger. 3d user interfaces: New directions and perspectives. *IEEE computer graphics and applications*, 28(6):20–36, 2008.

[13] Doug A Bowman, Cheryl Stinson, Eric D Ragan, Siroberto Scerbo, Tobias Höllerer, Cha Lee, Ryan P McMahan, and Regis Kopper. Evaluating effectiveness in virtual environments with mr simulation. In *Interservice/Industry Training, Simulation, and Education Conference*, volume 4, page 44, 2012.

[14] Frederik Brudy, Christian Holz, Roman Rädle, Chi-Jui Wu, Steven Houben, Clemens Nylandsted Klokmose, and Nicolai Marquardt. Cross-device taxonomy: Sur-

vey, opportunities and challenges of interactions spanning across multiple devices. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–28, 2019.

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

[16] Tom Chandler, Maxime Cordeil, Tobias Czauderna, Tim Dwyer, Jaroslaw Glowacki, Cagatay Goncu, Matthias Klapperstueck, Karsten Klein, Kim Marriott, Falk Schreiber, et al. Immersive analytics. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8. IEEE, 2015.

[17] Logan D Clark, Aakash B Bhagat, and Sara L Riggs. Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies*, 139:102413, 2020.

[18] HTC Corporation. Vivetm | discover virtual reality beyond imagination, 2011. URL https://www.vive.com/.

[19] Kylie Davidson, Lee Lisle, Kirsten Whitley, Doug A Bowman, and Chris North. Exploring the evolution of sensemaking strategies in immersive space to think. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[20] Tobias Drey, Patrick Albus, Simon der Kinderen, Maximilian Milo, Thilo Segschneider, Linda Chanzab, Michael Rietzler, Tina Seufert, and Enrico Rukzio. Towards collaborative learning in virtual reality: A comparison of co-located symmetric and asymmetric pair-learning. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2022.

[21] Fatima El Jamiy and Ronald 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.

[22] Barrett Ens, Benjamin Bach, Maxime Cordeil, Ulrich Engelke, Marcos Serrano, Wesley Willett, Arnaud Prouzeau, Christoph Anthes, Wolfgang Büschel, Cody Dunne, et al. Grand challenges in immersive analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2021.

[23] Barrett M Ens, Rory Finnegan, and Pourang P Irani. The personal cockpit: a spatial interface for effective task switching on head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3171–3180, 2014.

[24] Paul M Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6):381, 1954.

[25] Adrien Fonnet and Yannick Prie. Survey of immersive analytics. *IEEE transactions on visualization and computer graphics*, 27(3):2101–2122, 2019.

[26] Jann Philipp Freiwald, Sünje Gollek, and Frank Steinicke. Vr invite: A project-independent smartphone app for vr observation and interactivity. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part I 18*, pages 352–370. Springer, 2021.

[27] Bernhard Fröhler, Christoph Anthes, Fabian Pointecker, Judith Friedl, Daniel Schwajda, Andreas Riegler, Shailesh Tripathi, Clemens Holzmann, Manuel Brunner, Herbert Jodlbauer, et al. A survey on cross-virtuality analytics. In *Computer Graphics Forum*, volume 41, pages 465–494. Wiley Online Library, 2022.

[28] Eg Su Goh, Mohd Shahrizal Sunar, and Ajune Wanis Ismail. 3d object manipulation

techniques in handheld mobile augmented reality interface: A review. *IEEE Access*, 7: 40581–40601, 2019.

[29] Jerônimo Gustavo Grandi, Henrique Galvan Debarba, and Anderson Maciel. Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 127–135. IEEE, 2019.

[30] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4021–4033, 2017.

[31] Jiechang Guo, Yigang Wang, Peng Du, and Lingyun Yu. A novel multi-touch approach for 3d object free manipulation. In *Next Generation Computer Animation Techniques: Third International Workshop, AniNex 2017, Bournemouth, UK, June 22-23, 2017, Revised Selected Papers 3*, pages 159–172. Springer, 2017.

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

[33] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[34] Rex Hartson and Pardha Pyla. *The UX Book: Process and Guidelines for Ensuring a Quality User Experience*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1st edition, 2012. ISBN 0123852412.

[35] Syed Ali Hassan, Tariq Rahim, and Soo Young Shin. Childar: an augmented reality-based interactive game for assisting children in their education. *Universal Access in the Information Society*, 21(2):545–556, 2022.

[36] Carl-Philipp Hellmuth, Miroslav Bachinski, and Jörg Müller. Interaction techniques for 3d-positioning objects in mobile augmented reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pages 604–612, 2021.

[37] Ken Hinckley, Joe Tullio, Randy Pausch, Dennis Proffitt, and Neal Kassell. Usability analysis of 3d rotation techniques. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, pages 1–10, 1997.

[38] Errol R Hoffmann. Effective target tolerance in an inverted fitts task. *Ergonomics*, 38 (4):828–836, 1995.

[39] Sungwon In, Tica Lin, Chris North, Hanspeter Pfister, and Yalong Yang. This is the table i want! interactive data transformation on desktop and in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 2023.

[40] Petra Isenberg and Danyel Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. In *Computer Graphics Forum*, volume 28, pages 1031–1038. Wiley Online Library, 2009.

[41] Robert JK Jacob, Linda E Sibert, Daniel C McFarlane, and M Preston Mullen Jr. Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(1):3–26, 1994.

[42] Jason Jerald. *The VR book: Human-centered design for virtual reality.* Morgan & Claypool, 2015.

[43] Hans-Christian Jetter, Jens Gerken, Michael Zöllner, Harald Reiterer, and Natasa Milic-Frayling. Materializing the query with facet-streams: a hybrid surface for collaborative search on tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3013–3022, 2011.

[44] Polyxeni Kaimara, Andreas Oikonomou, and Ioannis Deliyannis. Could virtual reality applications pose real risks to children and adolescents? a systematic review of ethical issues and concerns. *Virtual Reality*, 26(2):697–735, 2022.

[45] Jiro Kawakita. *The Original KJ Method*. Kawakita Research Institute, 1991.

[46] Chutisant Kerdvibulvech. A review of augmented reality-based human-computer interaction applications of gesture-based interaction. In *HCI International 2019–Late Breaking Papers: 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pages 233–242. Springer, 2019.

[47] KyungTae Kim, Waqas Javed, Cary Williams, Niklas Elmqvist, and Pourang Irani. Hugin: A framework for awareness and coordination in mixed-presence collaborative information visualization. In *ACM International Conference on Interactive Tabletops and Surfaces*, pages 231–240, 2010.

[48] Matthias Kraus, Johannes Fuchs, Björn Sommer, Karsten Klein, Ulrich Engelke, Daniel Keim, and Falk Schreiber. Immersive analytics with abstract 3d visualizations: A survey. In *Computer Graphics Forum*, volume 41, pages 201–229. Wiley Online Library, 2022.

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

[50] Benjamin Lee, Xiaoyun Hu, Maxime Cordeil, Arnaud Prouzeau, Bernhard Jenny, and

Tim Dwyer. Shared surfaces and spaces: Collaborative data visualisation in a co-located immersive environment. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1171–1181, 2020.

[51] Lee Lisle, Xiaoyu Chen, JK Edward Gitre, Chris North, and Doug A Bowman. Evaluating the benefits of the immersive space to think. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 331–337. IEEE, 2020.

[52] Lee Lisle, Kylie Davidson, Edward JK Gitre, Chris North, and Doug A Bowman. Sensemaking strategies with immersive space to think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 529–537. IEEE, 2021.

[53] Jiazhou Liu, Arnaud Prouzeau, Barrett Ens, and Tim Dwyer. Design and evaluation of interactive small multiples data visualisation in immersive spaces. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 588–597. IEEE, 2020.

[54] Jiazhou Liu, Arnaud Prouzeau, Barrett Ens, and Tim Dwyer. Effects of display layout on spatial memory for immersive environments. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):468–488, 2022.

[55] Weizhou Luo, Anke Lehmann, Hjalmar Widengren, and Raimund Dachselt. Where should we put it? layout and placement strategies of documents in augmented reality for collaborative sensemaking. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2022.

[56] Haohua Lyu, Cyrus Vachha, Qianyi Chen, Odysseus Pyrinis, Avery Liou, Balasaravanan Thoravi Kumaravel, and Bjoern Hartmann. Webtransceivr: Asymmetrical communica-

tion between multiple vr and non-vr users online. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7, 2022.

[57] I Scott MacKenzie. A note on the information-theoretic basis for fitts' law. *Journal of motor behavior*, 21(3):323–330, 1989.

[58] Narges Mahyar and Melanie Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE transactions on visualization and computer graphics*, 20(12):1633–1642, 2014.

[59] Anthony Martinet, Géry Casiez, and Laurent Grisoni. Integrality and separability of multitouch interaction techniques in 3d manipulation tasks. *IEEE transactions on visualization and computer graphics*, 18(3):369–380, 2011.

[60] Arwa Michelle Mboya. The oculus go wasn't designed for black hair, Nov 2020. URL https://debugger.medium.com/the-oculus-go-a-hard-ware-problem-for-black-women-225d9b48d098.

[61] Meta. Horizon workrooms virtual office & meetings | meta for work. https://forwork.meta.com/horizon-workrooms/. Accessed: 2023-06-10.

[62] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–292. Spie, 1995.

[63] Annette Mossel, Benjamin Venditti, and Hannes Kaufmann. 3dtouch and homer-s: intuitive manipulation techniques for one-handed handheld augmented reality. In *Proceedings of the virtual reality international conference: laval virtual*, pages 1–10, 2013.

[64] Jens Müller, Roman Rädle, and Harald Reiterer. Virtual objects as spatial cues in collaborative mixed reality environments: How they shape communication behavior

and user task load. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1245–1249, 2016.

[65] Jens Müller, Roman Rädle, and Harald Reiterer. Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 6481–6486, 2017.

[66] Atsuo Murata and Hirokazu Iwase. Extending fitts' law to a three-dimensional pointing task. *Human movement science*, 20(6):791–805, 2001.

[67] Michael Nebeling, Katy Lewis, Yu-Cheng Chang, Lihan Zhu, Michelle Chung, Piaoyang Wang, and Janet Nebeling. Xrdirector: A role-based collaborative immersive authoring system. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.

[68] Muhammad Nur Affendy Nor'a, Fazliaty Edora Fadzli, Ajune Wanis Ismail, Zuraifah Syazrah Othman Vicubelab, Mohamad Yahya Fekri Aladin, and Wan Ahmad Asyraf Wan Hanif. Fingertips interaction method in handheld augmented reality for 3d manipulation. In *2020 IEEE 5th international conference on computing communication and automation (ICCCA)*, pages 161–166. IEEE, 2020.

[69] Patrick Aggergaard Olin, Ahmad Mohammad Issa, Tiare Feuchtner, and Kaj Grønbæk. Designing for heterogeneous cross-device collaboration and social interaction in virtual reality. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction*, pages 112–127, 2020.

[70] Chase Parker. Virtual solar system transforms young learners into astronauts. https://news.vt.edu/articles/2023/09/univlib-virtual-solar-system.html. Accessed: 2023-11-29.

[71] Photon. Photon unity 3d networking framework sdks and game backend | photon engine, 2011. URL https://www.photonengine.com/pun.

[72] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, 6:5, 2019.

[73] Jarkko Polvi, Takafumi Taketomi, Goshiro Yamamoto, Arindam Dey, Christian Sandor, and Hirokazu Kato. Slidar: A 3d positioning method for slam-based handheld augmented reality. *Computers & Graphics*, 55:33–43, 2016.

[74] Luc Renambot, Thomas Marrinan, Jillian Aurisano, Arthur Nishimoto, Victor Mateevitsi, Krishna Bharadwaj, Lance Long, Andy Johnson, Maxine Brown, and Jason Leigh. Sage2: A collaboration portal for scalable resolution displays. *Future Generation Computer Systems*, 54:296–305, 2016.

[75] Jonathan C Roberts, Panagiotis D Ritsos, Sriram Karthik Badam, Dominique Brodbeck, Jessie Kennedy, and Niklas Elmqvist. Visualization beyond the desktop–the next big thing. *IEEE Computer Graphics and Applications*, 34(6):26–34, 2014.

[76] Tom Rodden. A survey of cscw systems. *Interacting with computers*, 3(3):319–353, 1991.

[77] David Saffo, Andrea Batch, Cody Dunne, and Niklas Elmqvist. Through their eyes and in their shoes: Providing group awareness during collaboration across virtual reality and desktop platforms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2023.

[78] Samsung. Galaxy tab a7 lite, 2021. URL https://www.samsung.com/levant/tablets/galaxy-tab-a/galaxy-tab-a7-lite-gray-32gb-sm-t220nzaameb/.

[79] Marc Ericson C Santos, Takafumi Taketomi, Christian Sandor, Jarkko Polvi, Goshiro Yamamoto, and Hirokazu Kato. A usability scale for handheld augmented reality. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pages 167–176, 2014.

[80] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Tobias Czauderna, and Bernhard Jenny. Maps around me: 3d multiview layouts in immersive spaces. *Proceedings of the ACM on Human-Computer Interaction*, 4(ISS):1–20, 2020.

[81] Alexander Schäfer, Gerd Reis, and Didier Stricker. A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems. *ACM Computing Surveys*, 55 (6):1–27, 2022.

[82] Mickael Sereno, Xiyao Wang, Lonni Besançon, Michael J Mcguffin, and Tobias Isenberg. Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2530–2549, 2020.

[83] Marcos Serrano, Kadek Ananta Satriadi, Yalong Yang, Barrett Ens, Arnaud Prouzeau, and Stefanie Zollmann. Immersive analytics spaces and surfaces. In *Companion Proceedings of the 2022 Conference on Interactive Surfaces and Spaces*, pages 68–71, 2022.

[84] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

[85] Lauren Shupp, Robert Ball, Beth Yost, John Booker, and Chris North. Evaluation of viewport size and curvature of large, high-resolution displays. In *Graphics Interface*, pages 123–130, 2006.

[86] Spatial. Spatial - your world awaits. https://www.spatial.io/. Accessed: 2023-06-10.

[87] Steam. Steamvr on steam, 2016. URL https://store.steampowered.com/app/250820/SteamVR/.

[88] Unity Technologies. Unity - unity, 2019. URL https://unity.com/.

[89] Balasaravanan Thoravi Kumaravel, Cuong Nguyen, Stephen DiVerdi, and Bjoern Hartmann. Transceivr: Bridging asymmetrical communication between vr users and external collaborators. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pages 182–195, 2020.

[90] Wai Tong, Meng Xia, Kam Kwai Wong, Doug A Bowman, Ting-Chuen Pong, Huamin Qu, and Yalong Yang. Towards an understanding of distributed asymmetric collaborative visualization on problem-solving. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 387–397. IEEE, 2023.

[91] Eleftherios Triantafyllidis and Zhibin Li. The challenges in modeling human performance in 3d space with fitts' law. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2021.

[92] Edward R Tufte. The visual display of quantitative information. *The Journal for Healthcare Quality (JHQ)*, 7(3):15, 1985.

[93] Virbela. Virbela: A virtual world for work, education, & events. https://www.virbela.com/. Accessed: 2023-06-10.

[94] Mark A Whiting, Chris North, Alex Endert, Jean Scholtz, Jereme Haack, Carrie Varley, and Jim Thomas. Vast contest dataset use in education. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pages 115–122. IEEE, 2009.

[95] Jason Wither and Tobias Hollerer. Pictorial depth cues for outdoor augmented reality.

In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, pages 92–99. IEEE, 2005.

[96] Cik Suhaimi Yusof, Huidong Bai, Mark Billinghurst, and Mohd Shahrizal Sunar. A review of 3d gesture interaction for handheld augmented reality. *Jurnal Teknologi*, 78 (2-2):15–20, 2016.

[97] Qian Zhou, George Fitzmaurice, and Fraser Anderson. In-depth mouse: Integrating desktop mouse into virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2022.

Appendices

Appendix A

IRB Forms for Layout Dimensionality Study

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Take Part in a Research Study

Title of research study: Design and Collaboration of Immersive and Non-Immersive Users Across a Shared Workspace Environment (IRB#23-119)

Principal Investigator: Dr. Yalong Yang, yalongyang@vt.edu

Other study contact(s):

- Daniel Enriquez - denriquez@vt.edu

Key Information: The following is a short summary of this study to help you decide whether to be a part of this study. More detailed information is listed later in this form.

Virtual Reality (VR) has great capabilities that allow people to work and understand data through new perspectives. A major roadblock to being in VR is excluding those that collaborate with other people outside of VR or interact with different environments that can not transition into a non-immersive environment seamlessly and synchronously.

As a result, this work aims to resolve and understand the mismatch between working conditions where users can either be fully immersed in VR or working on a desktop/PC environment and where users can expect to collaborate with others in an environment different from their own.

The procedure for the research study will require participants to attend only one session. The session will last for no longer than approximately 90 minutes. During the session, participants can expect to spend up to 15 minutes familiarizing themself with either the VR environment or the desktop environment, afterwards. Afterwards, for approximately 45 minutes, they will be given a task to analyze a given data representation and then collaborate and discuss with the other participant to determine insights on the data. During this one participant will be using VR equipment and the other will be using a PC environment. Questionnaires will be administered before the training session and after the discussion, lasting a total of approximately 15 minutes. Participants will then be asked open ended questions about their experience with the application, lasting approximately 10 minutes.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at Daniel Enriquez (denriquez@vt.edu).

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team

Consent to Take Part in a Research Study

- You want to talk to someone besides the research team to provide feedback about this research

How many people will be studied?

We plan to include up to twenty people in this research study.

We plan to include about twenty people at this location out of ten people in the entire study nationally.

What happens if I say yes, I want to be in this research?

The procedure for the research study will require participants to attend only one session. The session will last for no longer than approximately 75 minutes.

Study participants will be interacting with one or more of application developers who will be administering the tasks: Daniel Enriquez. The research will be conducted on-campus at Virginia Tech towards February/March of 2023.

Depending on the participant, you will either use a VR Vive headset with controllers or a PC.

What happens if I say yes, but I change my mind later?

You can leave the research at any time, for any reason, and it will not be held against you. If you decide to leave the research, contact the investigator so that the investigator can dispose of your data.

Is there any way being in this study could be bad for me? (Detailed Risks)

Risk: Motion Sickness

Probability: Low

Description: Motion sickness in virtual reality is a result of the brain observing movement while the user's body remains static. The risk of this occurring in the VR application is low as participants will remain in a static/stationary position within the virtual environment and in reality. Several Known physical factors associated with the use of a VR headset include nausea, dizziness, eye strain, headache, and loss of balance. Participants who indicate or expect any physical discomfort related to the use of a VR headset will be excluded from this study.

What happens to the information collected for the research?

We will make every effort to limit the use and disclosure of your personal information, including research study and medical records, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

Your information or samples that are collected as part of this research will not be used or distributed for future research studies, even if all of your identifiers are removed.

The results of this research study may be presented in summary form at conferences, in presentations, and academic papers.

Consent to Take Part in a Research Study

Can I be removed from the research without my OK?

The person in charge of the research study or the sponsor can remove you from the research study without your approval. Possible reasons for removal include disruptive behavior.

What else do I need to know?

Any expenses accrued for seeking or receiving medical or mental health treatment will be your responsibility and not that of the research project, research team, or Virginia Tech.

We will not offer to share your individual test results with you. You may accept or decline these results.

Signature Block for Capable Adult

Your signature documents your permission to take part in this research. We will provide you with a signed copy of this form for your records.

Signature of subject	Date
Printed name of subject	
Signature of person obtaining consent	Date
Printed name of person obtaining consent	

Participant ID *

Your answer

What is your gender *

- ◯ Male
- ◯ Female
- ◯ Prefer not to say
- ◯ Other: _____

Please indicate your age *

Your answer

On a scale of 1 to 5, how much experience do you have with Virtual Reality (VR)? *

	1	2	3	4	5	
No VR Experience	◯	◯	◯	◯	◯	Plentiful VR Experience

Please briefly describe your experience with VR. *

Your answer

On a scale of 1 to 5, how much experience do you have with collaborative shared workspaces (Miro boards, SAGE displays, Google Jamboard, etc.)? *

	1	2	3	4	5	
No Experience	○	○	○	○	○	Plentiful Experience

Please briefly describe your experience with collaborative shared workspaces *

Your answer

Participant ID *

Your answer

Please indicate how you felt while using the application for each of the items, on the following scale:

1 = not at all, 2 = slightly, 3 = moderately, 4 = fairly, 5 = extremely

My actions depended on the other's actions. *

	1	2	3	4	5	
Not at all	○	○	○	○	○	Extremely

The other's actions were dependent on my actions. *

	1	2	3	4	5	
Not at all	○	○	○	○	○	Extremely

The other paid close attention to me. *

	1	2	3	4	5	
Not at all	○	○	○	○	○	Extremely

I paid close attention to the other. *

	1	2	3	4	5	
Not at all	○	○	○	○	○	Extremely

What the other did affected what I did. *

	1	2	3	4	5	
Not at all	○	○	○	○	○	Extremely

What I did affected what the other did. *

	1	2	3	4	5	
Not at all	○	○	○	○	○	Extremely

Please indicate how you felt while using the application for each of the items, on the following scale:

1 = Very low, 7 = Very High

How mentally demanding was the task? *

	1	2	3	4	5	6	7	
Very Low	○	○	○	○	○	○	○	Very High

How physically demanding was the task? *

	1	2	3	4	5	6	7	
Very Low	○	○	○	○	○	○	○	Very High

How hurried or rushed was the pace of the task? *

	1	2	3	4	5	6	7	
Very Low	○	○	○	○	○	○	○	Very High

How successful were you in accomplishing what you were asked to do? *

	1	2	3	4	5	6	7	
Very Low	○	○	○	○	○	○	○	Very High

How hard did you have to work to accomplish your level of performance? *

	1	2	3	4	5	6	7	
Very Low	○	○	○	○	○	○	○	Very High

How insecure, discouraged, irritated, stressed, and annoyed were you? *

	1	2	3	4	5	6	7	
Very Low	○	○	○	○	○	○	○	Very High

Is there anything you'd like to share about using this application in this condition? *

Your answer

Rank the conditions overall in terms of which you you preferred the most *

	Most Preferred	Middle	Least Preffered
PC2D-VR3D	○	○	○
PC3D-VR3D	○	○	○
PC2D-VR2d	○	○	○

Is there anything else you would like to share that wasn't previously addressed? *

Your answer

Participant ID *

Your answer

Rank the conditions in terms of which one felt more user friendly to you *

	Most preferred	Middle	Least Preferred
PC2D-VR3D	○	○	○
PC3D-VR3D	○	○	○
PC2D-VR2D	○	○	○

Rank the conditions in terms of which one felt more productive to you *

	Most Preferred	Middle	Least Preferred
PC2D-VR3D	○	○	○
PC3D-VR3D	○	○	○
PC2D-VR2D	○	○	○

Rank the conditions in terms of which one felt easier to communicate to your partner *

	Most Preferred	Middle	Least Preferred
PC2D-VR3D	○	○	○
PC3D-VR3D	○	○	○
PC2D-VR2D	○	○	○

Rank the conditions overall in terms of which you you preferred the most *

	Most Preferred	Middle	Least Preffered
PC2D-VR3D	○	○	○
PC3D-VR3D	○	○	○
PC2D-VR2d	○	○	○

Is there anything else you would like to share that wasn't previously addressed? *

Your answer

Appendix B

IRB Forms for Object Manipulation Study

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Take Part in a Research Study

Title of research study: ***23-178 Large Scale Object Manipulation in XR***

Principal Investigator: *Dr. Sang Won Lee, sangwonlee@vt.edu, 540-231-4857*

Other study contact(s): *Dr. Myounghoon Jeon, Daniel Enriquez, Hayoun Moon {myounghoonjeon@vt.edu, denriquez@vt.edu, moonhy@vt.edu}*

Key Information: The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form. The objective of this study is to explore virtual object manipulation methods using hand-held tablets. By collecting both objective and subjective data, the study will identify how manipulation methods differ in user's performance (accuracy, time), movement behavior, usability, workload, and preference. You will experience nine sessions of three one-minute-long tasks to move an object to a target location in a 3D virtual space. The entire study will take approximately 60-90 minutes. You will be compensated of $10/hour Amazon gift card for your time and effort.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at *sangwonlee@vt.edu (Dr. Sang Won Lee).*

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

How many people will be studied?

We plan to include about 40 people in this research study.

What happens if I say yes, I want to be in this research?

After consenting to the study, you will be asked to fill out a pre-study survey on your basic demographic and previous game experiences. You will then be trained to use three virtual object manipulation methods using tablets. There will be nine sessions of three one-minute-long tasks. For all tasks, you will walk around a room-sized environment and move an object to a target location in a 3D virtual space. You will perform the task using a tablet display, and there will be no headset involved. After each session, you will evaluate its usability and workload. The entire session including a training, nine sessions, and evaluation questionnaire swill take no longer than 90 minutes. The experiment will take place at the Blacksburg campus of Virginia Tech.

Consent to Take Part in a Research Study

What happens if I say yes, but I change my mind later?

You can leave the research at any time, for any reason, and it will not be held against you.

Is there any way being in this study could be bad for me? (Detailed Risks)

There is no critical expected by participating in this study. We anticipate no discomfort or danger commonly associated with VR technology as it does not involve the use of a headset.

What happens to the information collected for the research?

We will make every effort to limit the use and disclosure of your personal information, including research study and medical records, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech. If identifiers are removed from your private information or samples that are collected during this research, that information or those samples could be used for future research studies or distributed to another investigator for future research studies without your additional informed consent. The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation.

Can I be removed from the research without my OK?

The person in charge of the research study or the sponsor can remove you from the research study without your approval. Possible reasons for removal include if an individual does not show up or becomes belligerent and need to be asked to leave.

What else do I need to know?

If you agree to take part in this research study, you will receive $10/hour Amazon gift card for your time and effort.

Signature Block for Capable Adult

Your signature documents your permission to take part in this research. We will provide you with a signed copy of this form for your records.

Signature of subject	Date
Printed name of subject	
Signature of person obtaining consent	Date
Printed name of person obtaining consent	

Participant ID *

Your answer

How old are you? *

Your answer

Rate your experience with **video games** *

	1	2	3	4	5	6	7	
No Experience	○	○	○	○	○	○	○	Very Experienced

If so, which **consoles or platforms** have you used most frequently? *

Your answer

Describe your experience with **video games** *

Your answer

Rate your experience with **mobile games** *

	1	2	3	4	5	6	7	
No Experience	○	○	○	○	○	○	○	Very Experienced

Briefly describe your experience with **mobile games** *

Your answer

Participant ID *

Your answer

I think that interacting with this application requires a lot of body muscle effort. *

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I felt that using the application was comfortable for my arms and hand. *

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I found the device difficult to hold while operating the application. *

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I think the operation of this application is simple and uncomplicated. *

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I felt that the information display was responding fast enough. *

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I thought that the information displayed on screen was confusing. *

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I thought the words and symbols on screen were easy to read. *

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

I thought that the information displayed on screen was consistent. *

	1	2	3	4	5	6	7	
Strongly Disagree	○	○	○	○	○	○	○	Strongly Agree

How much mental and perceptual activity was required? Was the method easy or demanding, simple or complex? *

	1	2	3	4	5	6	7	
Easy and Simple	○	○	○	○	○	○	○	Demanding and Complex

How much physical activity was required? Was the method easy or demanding, slack or strenuous? *

	1	2	3	4	5	6	7	
Easy and Slack	○	○	○	○	○	○	○	Demanding and Strenuous

How much time pressure did you feel due to the pace at which the method or method elements occurred? Was the pace slow or rapid? *

	1	2	3	4	5	6	7	
Slow	○	○	○	○	○	○	○	Rapid

How successful were you in performing the method? How satisfied were you with your performance? *

	1	2	3	4	5	6	7	
Not Successful	○	○	○	○	○	○	○	Very Successful

How hard did you have to work (mentally and physically) to accomplish your level of performance? *

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very much so

How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the method steps? *

	1	2	3	4	5	6	7	
Not at all	○	○	○	○	○	○	○	Very Irritated, Stress, and Annoyed

Participant ID *

Your answer

Rank the Methods in terms of which you prefer *

	Most Preferred	Middle	Least Preferred
3DTouch	☐	☐	☐
HomerS	☐	☐	☐
DPad	☐	☐	☐

Which method did you prefer the most and why? *

Your answer

Which method did you prefer the least and why? *

Your answer

In what order did the following factors influence your decision: easy to use, speed, and accuracy? *

Your answer

Participant ID *

Your answer

Which method did you **learn to use** the quickest, and why? *

Your answer

Which method did you find the most **intuitive to use**, and why? *

Your answer

Which method felt the most **fun to use**, and why? *

Your answer

Anything else you would like to share? *

Your answer