FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Remote Collaboration over 3D Content in DeskVR

Nuno Miguel da Silva Alves

WORKING VERSION

Mestrado em Engenharia Informática e Computação

Supervisor: Daniel Filipe Martins Tavares Mendes

Second Supervisor: Rui Pedro Amaral Rodrigues

June 29, 2024

Remote Collaboration over 3D Content in DeskVR

Nuno Miguel da Silva Alves

Mestrado em Engenharia Informática e Computação

June 29, 2024

Abstract

The transition of real-world collaborative and teamwork tasks to the digital world has proven challenging, often resulting in the loss of the inherent social aspects of real-world collaboration. Research has addressed this gap by exploring social translucence – an approach to designing systems for social processes by implementing visibility, awareness, and accountability.

Virtual Reality (VR) is an excellent vector for incorporating the principles of social translucence. Due to its high sense of presence and immersion, VR inherently possesses enhanced social capabilities. A subset of VR of particular interest in this domain is DeskVR. DeskVR allows users to be fully immersed in a virtual environment while remaining seated at their desks. It offers a practical solution for extended working hours by reducing physical fatigue associated with standing and mid-air gestures.

Nevertheless, close interaction collaboration, such as the shared manipulation of an object, is complex due to the nature of the system. One fundamental challenge is the lack of awareness regarding other users' intentions. This absence contributes to the unpredictability of concurrent object manipulation, exacerbated by the lack of a physical link between users and the object. Furthermore, devising interaction techniques for DeskVR is difficult due to limited physical mobility and space.

This work proposes a method for implementing social awareness in the context of shared object manipulation in DeskVR. The proposed method involves designing a specialized multi-user interaction technique suited to the constraints of DeskVR. By designing with social translucence in mind, this technique helps users understand each other, fostering harmonious communication and collaboration. The research will also comprise an empirical user study to validate the effectiveness of this approach, evaluating the solution's usability and effectiveness.

Keywords: virtual environments, virtual reality, shared object manipulation, collaboration, social translucence, awareness, DeskVR

ACM Classification:

- Human-centered computing → Human computer interaction (HCI) → Interaction paradigms → Virtual reality
- Human-centered computing → Human computer interaction (HCI) → Interaction paradigms → Collaborative interaction

"It's a meaningless slab of iron you can't even lift,
for killin' dragons and monsters that ain't even real."

Godo, Berserk "He Who Hunts Dragons" by Kentaro Miura

Contents

List of Figures

List of Tables

Abbreviations

SA	Situation Awareness
CSCW	Computer-Supported Cooperative Work
VR	Virtual Reality
XR	Extended Reality
VE	Virtual Environment
CVE	Collaborative Virtual Environment
DOF	Degrees of Freedom
CAVE	Cave Automatic Virtual Environment
HMD	Head Mounted Display
WIM	World-in-Miniature
RPC	Remote Procedure Call

Chapter 1

Introduction

Throughout human history, collaboration has been an integral part of our pursuit of knowledge. Effective communication and a deep comprehension of each other have propelled us to advance our understanding and technological progress. In the digital age, it is only natural to harness the power of technology to extend collaboration to the digital realm, connecting people around the world. This is particularly pertinent in the aftermath of the COVID-19 pandemic, which has underscored the importance of digital tools in facilitating collaboration.

Virtual Reality (VR) emerges as a compelling platform for collaboration, offering a heightened sense of presence and immersion that inherently enhances social capabilities. A noteworthy subset of VR is DeskVR, immersing users in a virtual environment using a stereoscopic head-mounted display (HMD), all while comfortably seated at an office desk. DeskVR presents unique opportunities for interactions with the virtual environment, focusing on enhancing comfort, reducing physical fatigue, and improving accessibility to VR.

1.1 Context and Motivation

The transition of real-world collaborative tasks to the digital realm demands careful consideration and a thorough understanding of how we communicate. Often, certain tasks become more challenging in the digital space, leading users to prefer and be more efficient in performing those tasks in the real world. During this transition, social information tends to become opaque, limiting users' understanding of each other [16].

VR is an excellent medium for collaboration since users can coexist in a virtual space, visually observing each other's actions and communicating through voice. Still, designing for awareness in VR presents some challenges, particularly with interfaces and information presentation. For example, reading text can be difficult in VR [45, 27]. Nonetheless, advancements in resolution and HMD technology, such as eye gaze tracking, are addressing this issue.

Using VR while standing and relying on mid-air interaction for extended periods can induce fatigue and may be less accessible for individuals with mobility impairments. DeskVR addresses this concern, allowing users to interact with the environment while seated. The use of an office

desk offers additional benefits, serving as a surface for touch-based approaches, providing passive haptic feedback, and serving as a virtual representation for presenting social information or operational instructions [69, 59]. However, constraining users to a seated position limits their movement, requiring special attention while designing for DeskVR.

For instance, one of the most common interactions in VR is object manipulation, which often relies on techniques designed for users who are standing and engaging in physically demanding mid-air movements. Given that DeskVR users are confined to a desk, their mobility and range for mid-air movements are limited. Therefore, it is important to design object interaction techniques considering range, physical demand, and ergonomics for seated users [2].

Collaboration can be used to help object manipulation. For example, handling large objects in VR is difficult as they may obscure the user's view during positioning tasks, often requiring users to place and inspect the object from alternative angles [7]. A collaboration involving multiple users can be advantageous in these situations, as diverse viewpoints contribute to a better understanding of the environment, assisting the primary user in successfully manipulating the object [44]. The challenge lies in seamlessly integrating efficient collaboration and social awareness within the limitations of DeskVR.

1.2 Objectives

The goal of this work is to design and implement a collaborative framework for VR that seamlessly integrates efficient collaboration and awareness within DeskVR. This will involve a comprehensive examination of the current state-of-the-art techniques in collaboration, computer-supported cooperative work (CSCW), and DeskVR. The objective was to discern the difficulties and challenges inherent in these areas, hoping to identify a promising avenue for exploring collaboration. Subsequently, this framework will undergo user evaluation to analyze its effectiveness as a tool for communication and cooperation.

1.3 Document Structure

Chapter 2, *Related Work*, explores the current state-of-the-art relevant to this research. It starts with an investigation into computer-supported cooperative work, then examines techniques employed to maintain consistency in shared virtual environments, and subsequently delves into DeskVR interaction techniques. The chapter ends with a discussion of the examined work, relating the concepts between various topics and offering insights into the design of the proposed solution.

Chapter 3, *Design of the Proposed Method*, presents a high-level overview of the design of the proposed solution, offering detailed specifications and descriptions for each component necessary for the solution's implementation.

Chapter 4, *Implementation of a Prototype*, showcases the concrete implementation following the design standards established in Chapter 3. It details the technologies and hardware used, as well as delving into more technical aspects.

Chapter 5, *Evaluation*, outlines the testing methodology and the subsequent evaluation of the obtained results.

Finally, Chapter 6, *Conclusions*, brings this work to a close by revisiting key points, suggesting potential future research directions, and providing concluding remarks.

Chapter 2

Literature Review

This chapter explores the literature on virtual reality and collaboration. Section 2.1 delves into computer-supported cooperative work and describes social translucence, concepts of workspace awareness, and group spatial interactions. Section 2.2 explores methods to ensure consistency in shared virtual environments across multiple users. Section 2.3 explores various interaction techniques used in DeskVR research. Lastly, section 2.4 assesses the techniques' applicability for a DeskVR environment, providing insights that inform the design of the proposed solution.

2.1 Computer-Supported Cooperative Work

The transition of real-world collaborative tasks to the digital world has rendered social information invisible, leading to challenges in effective communication and collaboration [16]. This section explores approaches to designing digital systems to enhance collaboration and information visibility by addressing these challenges.

2.1.1 Social Translucence

Erickson and Kellogg [16] define *socially translucent systems* as those that facilitate coherent behavior in human-to-human communication by making participants and their activities visible to one another. They believe that social information provides the basis for inferences, planning, and coordination of activities. They describe three properties of these systems: *visibility* of socially significant information, *awareness* of what is happening, and *accountability* for users' actions. Interestingly, they denote that while accountability and awareness are typically correlated in the physical world, they may not necessarily coincide in the digital realm.

The rationale for using the term *translucent*, rather than *transparent*, is that in the real world, social information is not entirely transparent due to various constraints. For example, physical distances between conversing groups prevent each group from hearing the other's conversation. Erickson and Kellog argue that there is power in these constraints, shaping our words and actions based on the audience present or its size. In digital environments, it is hard to understand the constraints or whether they are even shared, such as muting someone on a chat room without

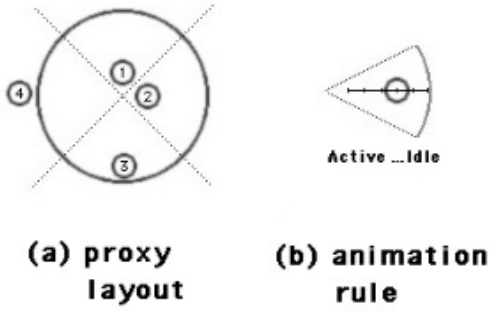


Figure 2.1: The illustration on the left shows the layout of the proxy in Babble [16], where dots 1, 2, and 3 are part of the conversation, and 4 is in another. The illustration on the right shows how the dots animate and drift further away depending on their activity level.

their knowledge or instances of shadow-banning. They say that the shared awareness of these constraints is critical in structuring our interactions, so they highlight the importance of supporting this in the digital realm.

Moreover, in the same article [16], they discuss how social translucence may be implemented in practice. First, they explore ways to make individuals' activity visible. For this purpose, they describe three designs: a realistic approach – projecting real-world social information to the virtual world through means such as video conferencing; a mimetic approach – creating a representation of social information in the digital world as closely as possible, such as avatars; and an abstract approach – portraying social information in the digital domain through abstract representations.

As a result of this discussion, they [16] chose the abstract method for further analysis. This decision stems from the apparent characteristics of these representations using text and simple graphics: they are easy to create, persist over time, and can be easily found using search and visualization engines.

Consequently, Erickson and Kellog describe a prototype called Babble that uses abstract representations to portray social information. Babble illustrates social communication using two approaches: a persistent textual representation that displays comments, participants' names, and timestamps; and a synchronous visual representation titled *social proxy*, which shows conversations as a circle, where participants are dots inside that circle, and their closeness to the center represents their activity level, as shown in Figure 2.1. With this, they show the efficacy and the importance of abstract representations in portraying social information.

2.1.2 Workspace Awareness

Endsley [15] defines *awareness* as "knowing what is going on." Several traits have emerged in previous studies on awareness [1, 40, 15, 22]: Awareness is the knowledge about the state of the environment in both space and time. Since environments change over time, maintaining and updating this knowledge is part of awareness. Individuals can achieve this by interacting with and

exploring the environment. Finally, awareness is not the primary goal of a task-oriented system but rather a secondary goal, the primary objective being to complete the task.

Situation Awareness (SA) is a concept that arose in research about more dynamic and complex environments with high information load, variable workload, and risk [19]. Adams et al. [1] define SA as "the up-to-the-minute cognizance required to operate or maintain a system." Endsley [15] describes a three-stage definition of SA that focuses more on the process. The first stage is the perception of relevant elements in the environment. The second stage is the comprehension of the current situation, or the ability to relate perceptual information retrieved in the first stage with past knowledge to understand their meaning. The third and last stage is the forecast of the status of those elements, which is invaluable for decision-making.

Gutwin and Greenberg [22] define the concept of workspace awareness as "the up-to-the-moment understanding of another person's interaction with the shared workspace" [22]. It is a specialized kind of SA due to the shared workspace context – it is about being aware not only of the environment but also of other individuals and their interactions with that environment. Similarly to Erickson and Kellog [16], they believe the difficulty of maintaining awareness in collaborative tasks in digital systems lies in the technological constraints of the medium, the lack of information due to the limited perception of others, and the misrepresentation of that information.

To aid the understanding of concepts and purposes of designing workspace awareness in groupware systems, Gutwin and Greenberg [22] developed a three-part conceptual framework. The first part focuses on what information is used and what is necessary for workspace awareness. The second part is about how that information is gathered. Lastly, the third part concerns the application of workspace awareness in collaboration.

Previous research [14, 58, 35, 22] explore a set of workspace information that people keep track of during collaborative work, which is the elements that answer the questions "who, what, where, when, and how." Gutwin and Greenberg [22] present specific elements and questions answered by those elements according to what they believe is the core of workspace awareness, seen summarized on tables 2.1 and 2.2.

Earlier findings [53, 40, 12, 26, 22] propose three primary sources and mechanisms for gathering workspace information: through people's bodies and consequential communication, through workspace artifacts and feedthrough, and through conversation, gestures, and intentional communication. People's bodies provide extensive details about their current work, making watching and hearing other people a mechanism for obtaining workspace information [22, 53]. This mechanism for gathering information is called *consequential communication* [53].

The second source of information is the artifacts produced in the workspace [12, 20]. The information that can be collected from these artifacts can be obtained either through visible characteristics [22] or through the sound that is made when manipulating them [20]. The mechanism used for gathering this information is called *feedthrough* [12], meaning that when manipulating an artifact, information is received by not only the person doing the action as a form of feedback but also by others who are watching.

Table 2.1: Elements of workspace awareness of the present [22]

Category	Element	Questions
Who	Presence	Is anyone in the workspace?
	Identity	Who is participating? Who is that?
	Authorship	Who is doing that?
What	Action	What are they doing?
	Intention	What goal is that action part of?
	Artifact	What object are they working on?
Where	Location	Where are they working?
	Gaze	Where are they looking?
	View	Where can they see?
	Reach	Where can they reach?

Table 2.2: Elements of workspace awareness of the past [22]

Category	Element	Questions
How	Action history	How did that operation happen?
	Artifact history	How did this artifact come to be in this state?
When	Event history	When did that event happen?
Who	Presence history	Who was here, and when?
Where	Location history	Where has a person been?
What	Action history	What has a person been doing?

The third source of workspace information is conversation and gesture through intentional communication [9, 25, 6]. People can gather information from verbal communication in different ways: through explicit conversational exchange of awareness elements [22], by picking up relevant information from other peoples' conversations [22], or through *verbal shadowing* or "outlouds" [25] which are comments made by individuals while working that are addressed to no one in particular.

Gutwin and Greenberg [22] suggest five types of activities that benefit from workspace awareness. The identified types include: the management of coupling – the degree to which people are working together [52] or the amount of work a person can do before requiring the help of another [22]; simplification of communication – allowing using deictic references, demonstrations, and visual evidence; coordination of actions; anticipation; and assistance.

2.1.3 Spatial Group Interaction

Unlike conventional CSCW settings, real-time collaboration in virtual environments introduces novel interaction possibilities, particularly when multiple users simultaneously engage with one or more objects [8]. Therefore, it becomes essential to acknowledge the importance of enhancing awareness in this specific context.

A spatial approach to collaborative virtual environments consists of spaces or rooms allowing individuals to navigate and engage with one another and the objects, or artifacts, within these designated areas [4]. With the goal of enhancing collaboration and scalability within large virtual environments, Benford and Falhén [4] propose the spatial model of interaction.

The spatial interaction model introduces the concepts of medium, aura, awareness, focus, nimbus, and adapters. The medium is the means through which interactions with objects occur, for example, through text, audio, or visuals. The aura represents an area in which objects can interact within a given medium. When two auras collide, the interaction between the objects in the medium becomes possible. Objects can have multiple auras, such as their size, shape, and color.

Awareness is the basis on which objects control their potential interactions, as determined by their auras. This awareness is not necessarily symmetrical. For example, between object A and object B, object A can be more aware of B than B is aware of A. The awareness of an object towards another is calculated based on the combination of its focus and nimbus. Focus and nimbus are, like auras, subspaces of an object and describe its attention or presence, respectively. Benford and Falhén [4] describe this as "the more an object is within your focus, the more aware you are of it and the more an object is within your nimbus, the more aware it is of you".

Furthermore, Benford and Falhén [4] denote that a person does not need to be aware of their aura, focus, and nimbus, as these may be manipulated using natural interactions. Notably, they indicate three main ways of managing these subspaces. The first is through implicit interaction derived from, for example, the movement, orientation, or eye gaze of individuals. The second is through explicit adjustment of parameters, for instance, with the help of a user interface. The third is through adapters which modify a user's aura, focus, and nimbus. An example of an adapter would be a tool that a person can pick up or a table in which a user can sit, altering the user's aura, focus, and nimbus for that context.

Domingues et al. [13] develop a workflow-based approach to collaborative 3D interaction based on the spatial model of interactions. They describe the workflow concept, which manages tasks and ensures coordination in collaborative environments. The workflow comprises two components: the shared component, encompassing the data and state information of all users and sources within the environment; and the motor component, which uses the data from the shared component to execute actions on particular sources through assistance functions. Here, a *source* is an object that users can perceive, *particular sources* are objects that change during 3D interaction through assistance functions, serving as support tools, and *assistance functions* help manage coordination in 3D interaction.

Next, they specify five particular sources, dedicated to interaction in the virtual environment.

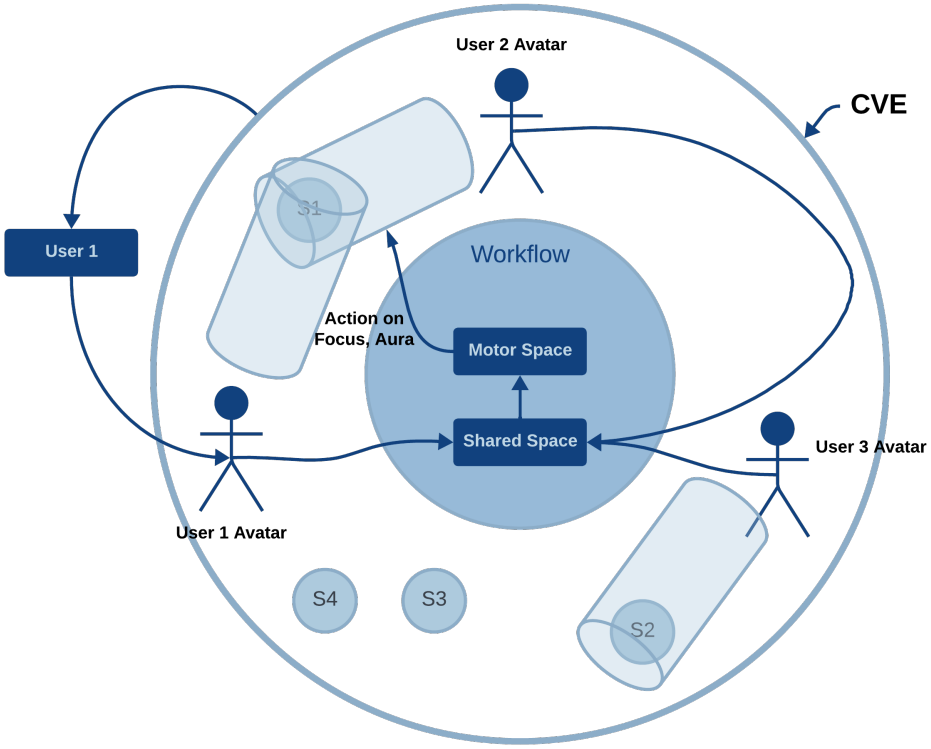


Figure 2.2: Portrayal adapted from the work of Domingues et al. [13]. This represents their workflow approach applied to a CVE, where cylinders represent user foci. In this example, the nimbus of the object S1 is the set of users 1 and 2.

3DIFocus coincides with objects the user can interact with, meaning in their field of view. *3DINimbus* symbolizes all users that might interact with this source. *3DIAura* is a zone that enables specific interactions if users are situated within its area. *3DIAssistant* assists users during specific actions through a virtual guide. Lastly, *3DIAvatar* is the digital counterpart of a user. These concepts are illustrated in figure 2.2

Using this model, Domingues et al. [13] describe several assistance functions, in particular, an assistance function for helping the coordination of multi-user interaction with an object. They model multi-user interaction by inserting joints between the user avatars and the object, allowing for manipulation of that object using single-user interaction techniques. The assistance function then helps manage this interaction by creating visual guides for the users by changing the color of the avatars engaged in multi-user interaction, or by displaying the object's direction, for instance.

2.2 Concurrency Control

Margery et al. [34] developed a classification to assess the cooperation levels in multi-user systems. In systems with level 1 collaboration, users can perceive and communicate with each other in the virtual world through avatars. Level 2 systems enable users to manipulate objects within the

scene. Level 3 systems allow users to manipulate the same object simultaneously. Level 3 can be further subdivided based on the degree of independence in users' actions, such as independently changing different aspects of an object or combining inputs codependently. This last collaborative level, particularly the codependent mode, is considered the most natural and immersive form of collaboration, offering enhanced efficiency.

Broll [8] had previously defined something similar to levels 2 and 3, cooperative and collaborative interaction, respectively. Recognizing which interactions are concurrent and which are not is essential yet challenging. Without a notion of simultaneity, interactions become a series of individual moves by each user rather than a coordinated combination. In distributed environments, typical in collaborative virtual reality, where each user has a copy of the world, determining concurrency is complicated, especially considering communication delays.

Concurrency control prevents conflicts of concurrent updates arising from adverse network conditions. In its absence, users manipulating shared objects in conflicting directions may experience difficulty anticipating outcomes and encounter divergence of state [46], introducing confusion among participants, offering disparate views of what should be a shared world and, in turn, violating the essential consistency requirement [62]. Therefore, the study of concurrency control emerges as an essential research topic for developing systems with level 3 collaboration.

The following subsections describe three distinct approaches for implementing concurrency control. Object ownership methods use locking mechanisms to block users from interacting with an object or its properties unless they possess a corresponding lock or ownership. Attribute separation methods assign specific attributes or degrees of freedom of an object to individual users, making each user responsible for that attribute. Finally, distributed average techniques calculate the average of participants' actions, integrating and updating them distributively.

2.2.1 Object Ownership Techniques

Greenberg and Marwood [21] describe how traditional concurrency control strategies can be used in real-time groupware systems. In particular, they demonstrate how the method of privileged access through locking can be applied in these systems, describing two different approaches. A non-optimistic locking policy blocks users from interacting with an object until a lock has been granted, signifying ownership over the object. An optimistic locking policy allows users to manipulate an object before they know if a lock has been granted to them. If the lock has been denied, the object returns to its original state, and a repair must be done. Optimistic locking schemes can be further divided into two approaches. A fully-optimistic approach allows users to manipulate multiple objects with tentative locks, while a semi-optimistic approach only allows users to manipulate one object at a time.

BrickNet [57] is a toolkit for network-based virtual environments. It uses a pessimistic locking mechanism for object manipulation where objects can only be updated by the current owner. Users have to request ownership over an object to the server, and the server mediates object updates. DIVE [23] is a multi-user distributed virtual environment system. DIVE also employs a pessimistic locking mechanism, similar to BrickNet, where users require an object-based token

to interact with the object. Spline [65] is a software platform for developing distributed virtual environments that implements the Interactive Sharing Transfer Protocol, or ISTP [64]. Objects in the ISTP protocol, like DIVE, are only subject to changes from its owner process. This ownership can be transferred between processes. CAVERNsoft [30] is a collaborative software architecture that was used by the CAVE Research Network, which also uses the pessimistic locking approach.

The PaRADE collaborative environment [49, 46] takes a different approach to pessimistic locking by employing a predictive strategy. Developed to mitigate adverse network effects, PaRADE aims to enhance both responsiveness and consistency through advanced time management. In PaRADE, all events are timestamped with both wall-clock time and causal time, and efforts are made to predict events to reduce the impact of network delays. Combining these times is used for sufficient causal ordering, serving as a middle ground for Lamport's causal ordering [28] to enhance concurrency and responsiveness. This approach restricts changes to an object to originate from a single user, leading to the adoption of a predictive object transfer paradigm that anticipates entity ownership in advance.

The predictive object ownership transfer relies on application knowledge and heuristics to anticipate a user's intention to interact with an object, enabling the transfer in advance of the ownership to minimize the impact of network delays. This advance transfer is important for optimizing responsiveness and in scenarios where ownership may have been erroneously transferred, allowing for retransmission to the correct recipient. Yang and Lee [67] identified scalability challenges in the widespread dissemination of token requests to all users in the network. In response, they introduced the concept of an entity radius, wherein messages are selectively delivered only to individuals within this radius, forming an "entity-centric multicast group."

Pessimistic approaches for locking mechanisms help preserve data integrity in databases, for example. However, in settings such as collaborative virtual environments, not only does obtaining a lock necessitate user waiting, but ensuring sequential updates becomes less scalable with a growing number of users. Sung et al. introduce CIAO [62], a large-scale 3D layout system that uses a semi-optimistic policy for managing shared object manipulation to enhance responsiveness. Unlike traditional systems, users in CIAO can manipulate objects without waiting for a lock. To address potential confusion arising from multiple users interacting with an object, the CIAO interface incorporates awareness features, using translucent clones to indicate ongoing manipulations yet to be validated. The authors also address the challenges of concurrently interacting with hierarchically related objects.

2.2.2 Attribute Separation Techniques

The approaches discussed in the previous section are limited to a collaboration level of 2, as there is no stage where multiple users synchronize their inputs on a shared object. Additionally, those concurrency control mechanisms often introduce issues commonly called "surprise" [31]. Such surprises manifest as conflicts leading to the reversal of changes or unexpected alterations in the anticipated order of events due to concurrency control conflicts.

This section introduces strategies that elevate collaboration to level 3, enabling independent modifications of various attributes of an object simultaneously. The concept involves assigning ownership of specific attributes to individual users, allowing multiple users to collaboratively manipulate an object without needing sequential ownership.

In their work, Lee et al. [29] refer to the manipulation of an attribute of a shared object as a "task." However, they explain the importance of avoiding "surprises" or, in this context, "task-surprises." These surprises may arise when interruptions occur during user tasks by other tasks or when task dependencies result in unexpected states when executed concurrently.

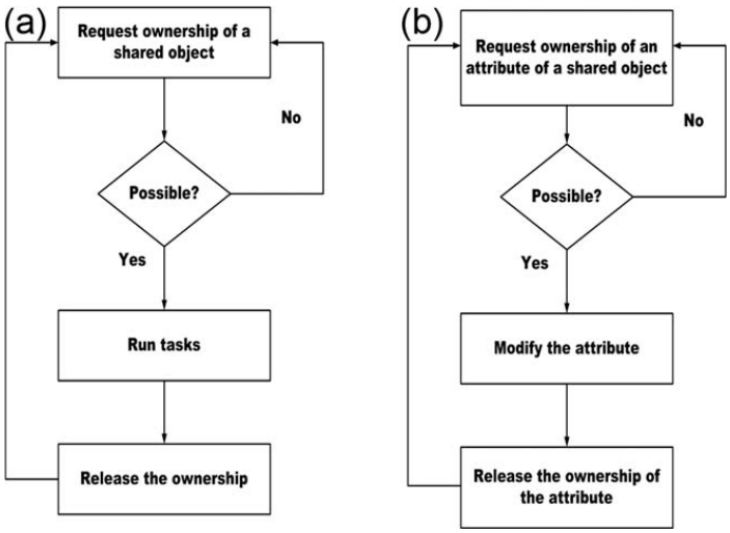


Figure 2.3: Illustration depicting the difference between ownership transfer and attribute separation [29].

Advancements in internet infrastructure around the turn of the century, particularly with the establishment of Internet2 [68] – a robust network implemented in the United States in 1996 for academic purposes – sparked new interest in exploring collaborative tasks in remote environments. Mortensen et al. [38] aimed to assess the feasibility of joint tasks when two users, separated by thousands of miles, interacted in a shared virtual environment within DIVE. The specific task involved both users lifting a stretcher together and moving it along a predefined path into a building. To achieve this, two handles were added to the stretcher for the users to manipulate, with the stretcher aligning itself based on the position and orientation of these handles. Their findings revealed that while data transmission speeds and throughput were sufficient, the lack of software support, particularly concerning lost data packets impacting consistency, posed a significant challenge.

Roberts et al. [47] investigated two distinct methods of shared object manipulation and how these methods were influenced by the display devices their users employed: an immersive walk-in display with tracked head and hands, and a desktop application. The two methods explored were sharing the same attribute and sharing through different attributes. To conduct this study,

they devised an experiment where users collaborated on constructing a gazebo in DIVE. Some sub-tasks required users to collaborate using the same attribute, such as jointly carrying a beam, while other tasks involved simultaneous manipulation of the object for different purposes, such as one person inserting a screw while another held the beam in place. This study also concluded that technological advancements had made such collaborative tasks feasible, but the ownership mechanism of DIVE failed to provide effective collaboration [48]. As such, it became evident that new systems needed to be developed with level 3 collaboration in mind.

Pinho et al. [43, 44] proposed a method for collaborative object interaction by separating degrees of freedom (DoF) and assigning them to different users. A degree of freedom in this context refers to a degree of control over the movement of an object, such as horizontal translation or rotation around an axis. As previously mentioned in section 2.3, the separation of DoF can be beneficial when precision is an important factor.

The goal of this approach was to use single-user interaction techniques, such as HOMER [7, 39], in collaborative manipulation, with the reasoning that they were well-researched, commonly understood, and perceived as more "magical" than natural, which was prevalent in most collaborative manipulation techniques then. The objective was to extend users' capabilities for multi-user object interaction. Additionally, the authors focused on awareness by implementing features like using a pointer to indicate the selected object, changing the object's color based on its state, and representing the DoF graphically, as shown in Figure 2.4. While the results were positive, two challenges were identified: testing the applicability with more than two users, and the selection of DoF for each user was predetermined in a configuration, limiting flexibility.

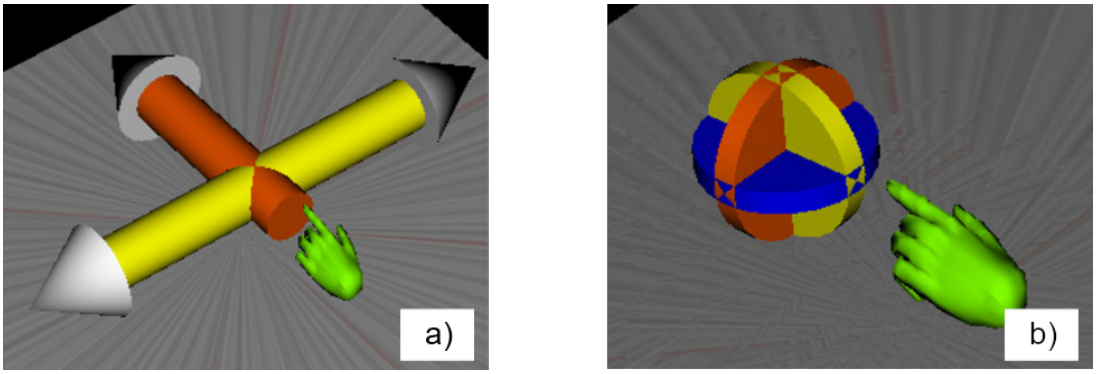


Figure 2.4: Illustration of translation (a) and rotation (b) pointers in [44]

Lee et al. [29] proposed a concurrency control mechanism based on fine-grained tasks. In this approach, if a user who does not own an object requests a task to manipulate a shared object, the task is permitted if it avoids conflicts with the owner's task and does not lead to task surprises. If the task is not allowed, non-owners can perform their tasks with duplicates of the object on a personal workspace, similar to the ghost images in CIAO [62]. The outcomes of these modifications in personal workspaces can be stored, allowing users to discuss and determine the final state of the object. Figure 2.5 demonstrates the flow of the concurrency control process.

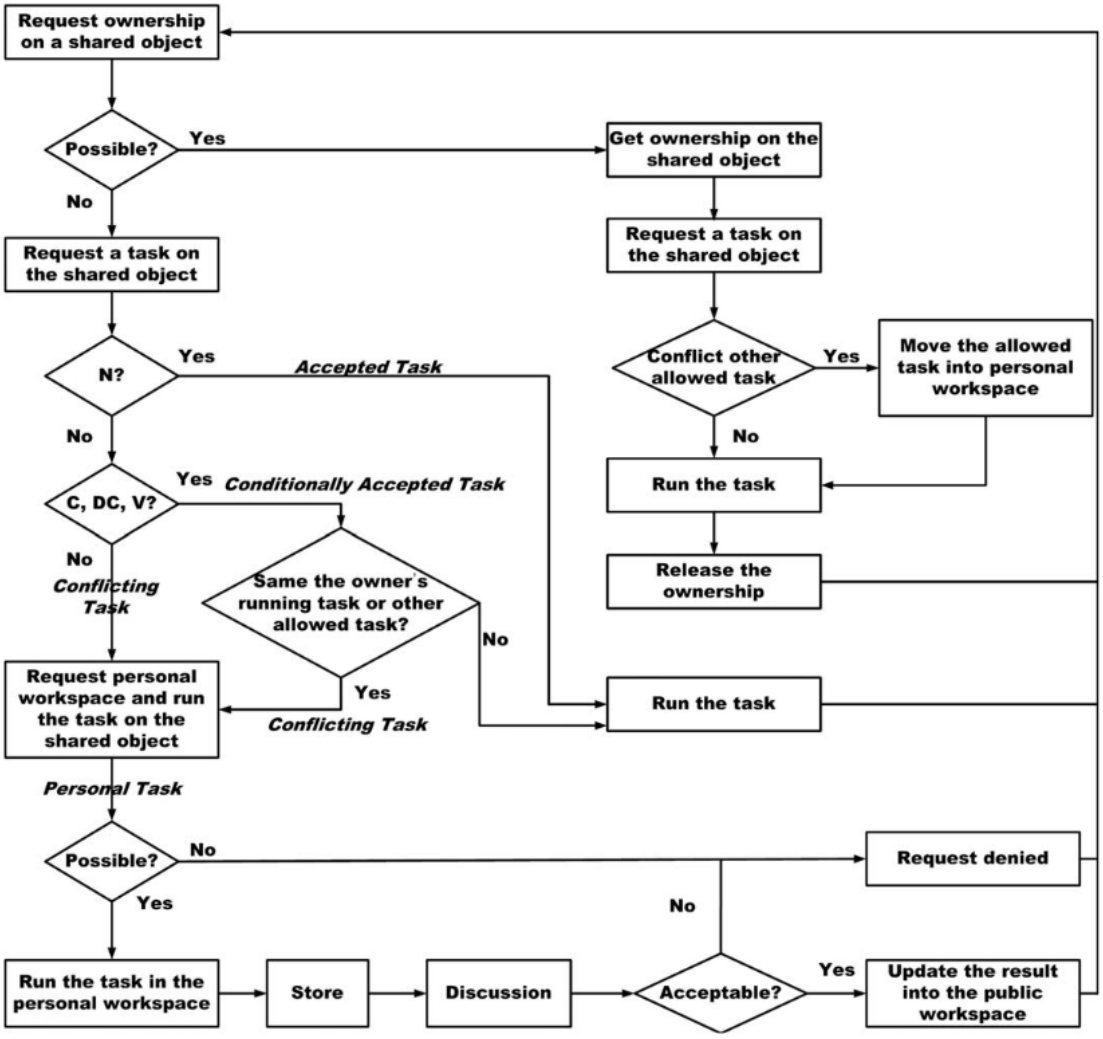


Figure 2.5: Illustration of the fine-grained task concurrency control system by Lee et al. [29].

The allowance types for each task classification can be seen on Table 2.3. *Conflicting tasks* are exclusive to the object's owner. *Conditionally accepted tasks* can be executed only if no one else is performing the task, while *accepted tasks* can always be executed. While their implementation demonstrated slower efficiency compared to other attribute-sharing mechanisms, Lee et al.'s approach reduced the occurrence of task surprises.

2.2.3 Distributed Average Techniques

This section explores techniques associated with codependent level 3 collaboration. Integrating level 3 collaboration into CVEs is challenging, especially in regards to the complexity of combining and integrating actions from multiple users, the absence of feedback inherent in physical interactions with an object, and network communication issues, especially concerning latency, which significantly influences the ability to detect simultaneous actions [50, 8].

Table 2.3: Allowance type for each task classification by Lee et al. [29].

Classification of Tasks	Allowance Type
Translation	Conflicting task
Rotation	Conflicting task
Scaling	Conflicting task
Fission	Conflicting task
Merge	Conflicting task
New	Accepted task
Remove	Conflicting task
Copy	Conditionally accepted task
Data calculation	Conditionally accepted task
Visual property	Conditionally accepted task

Ruddle et al. [50] investigated whether an asymmetric or symmetric approach was better suited for level 3 collaboration in the piano mover's task. This task involves two users maneuvering a large virtual object through confined spaces. Symmetric manipulation requires coordinated actions in direction and intensity, while asymmetric manipulation allows users to interact with the object differently, considering the average of the interactions. To address latency concerns, experiments were conducted on a single host computer. Interestingly, the results did not reveal a significant difference in the time participants took to complete the task, but instead, that performance depended on the task. However, the study highlighted the challenge of coordinating participants' movements, emphasizing the need for feedback, whether haptic or otherwise.

Friston et al. [18] present a concurrency control framework designed to enable the integration of level 3 collaboration in distributed virtual environments. This approach views the collaborative environment as a distributed data-fusion problem and incorporates elements such as prediction, distributed averaging, continuous authority, and constraint duplication. The methodology is rooted in consensus-based networking, where simulations share state updates with other nodes, and local solvers integrate these updates to establish a distributed average consensus of the system's state, ensuring consistency over time. Unlike bilateral operation or force-reflection systems, where clients submit inputs to a central authority, this approach calculates actions distributively, resulting in improved scalability. The authors conducted experiments that demonstrated the effectiveness of this technique in preserving causality, stacking objects, enabling level 3 collaboration, and providing support for haptic feedback.

2.3 DeskVR Interaction

Before the widespread adoption of stereoscopic displays, various effective approaches emerged for utilizing a multi-touch tabletop surface to interact with 3D objects with minimal fatigue [5, 61, 10]. Balloon Selection [5] employs the metaphor of manipulating a balloon attached to a string. This

approach divides a 3DOF positioning task into a 2DOF positioning task with one finger on the touch surface and a 1DOF string-pulling task with the other finger to control the height of the balloon. Pulling the fingers apart decreases the balloon's height and bringing them together increases its height. Balloon selection does not differentiate between right and left hands; instead, it prioritizes the order in which fingers are placed on the surface, designating the primary finger (anchor), secondary finger (stretching finger), and tertiary finger (selection finger). This design allows both right-handed and left-handed users to use the technique effortlessly. Additionally, the stretching finger may not always be held, allowing the balloon's height to remain fixed when removed. This feature, referred to as "String Height Clutching" by the authors, enables the extension of the balloon's height infinitely. This technique has shown low error rates, likely due to the user's hands being supported by the tabletop surface, resulting in significantly reduced hand tremor and arm fatigue.

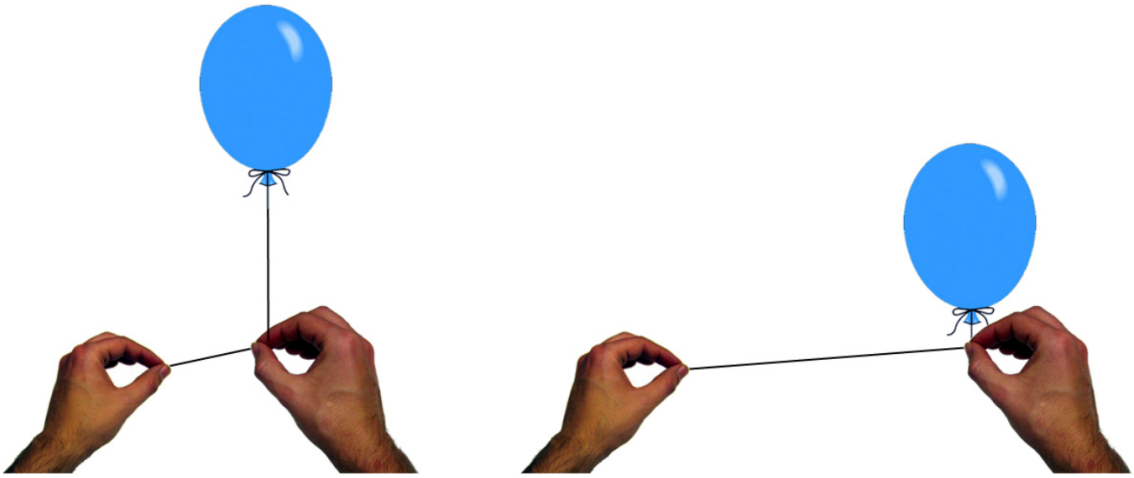


Figure 2.6: Illustration of the Balloon Selection metaphor [5].

Corkscrew Selection [5], similar to Balloon Selection, also breaks down the 3DOF task into a 2DOF positioning task and a 1DOF height task. However, it differs in the method used to adjust the height of the selection point, achieved through a rotational motion around a selection widget.

Triangle Cursor [61], on the other hand, creates an isosceles triangle with its two base vertices positioned at the touch points of two fingers, while the altitude's base point is located at the midpoint. Manipulating the triangle's position involves moving the fingers on the surface, while adjusting its height is controlled by scaling the triangle simlarly to a typical pinch gesture based on the distance between the two fingers. The height's upper limit is constrained by the diagonal of the touch surface. Additionally, users can initiate yaw rotation around an axis perpendicular to the table's surface by rotating the fingers around the midpoint.

In their evaluation comparing Triangle Cursor to Balloon Selection, the authors noted a slight preference in terms of speed and error minimization for Triangle Cursor. However, it's worth mentioning that the tasks necessitated the use of rotational techniques, which required an extension of Balloon Selection by using the rotation of the secondary finger around the primary finger as

rotational input.

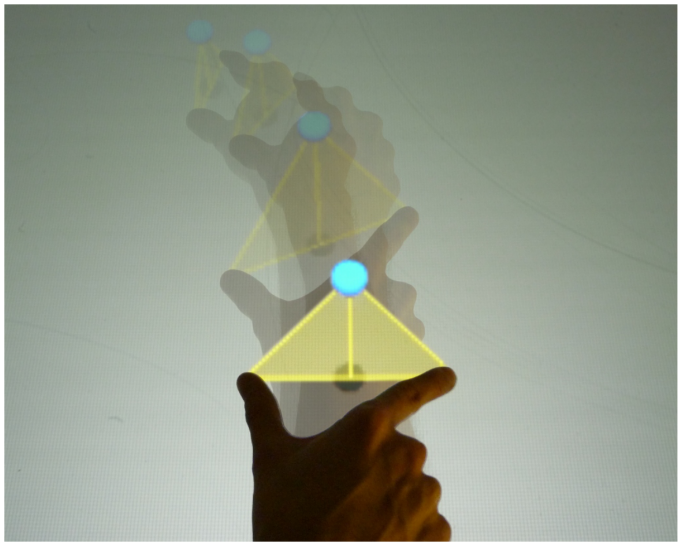


Figure 2.7: Illustration of the Triangle Cursor technique [61].

Zielasko et al. [71] introduced the concept of DeskVR, a fully immersive VR experience accessed entirely from an office desk while the user remains seated. Initially conceived to integrate with the workflow of data analysts, the application of this concept extends to broader applications, aiming to alleviate physical strain associated with standing, extend work periods, enhance accessibility, and boost overall productivity.

Zielasko et al. [70] surveyed to evaluate the advantages and disadvantages of standing versus sitting and the degree of embodiment in virtual reality experiences. The findings indicated that sitting generally scored higher in reducing cybersickness, enhancing comfort, ensuring safety, and improving accessibility. On the other hand, standing was preferred for perceived self-motion, locomotion preciseness, and engagement.

Given the constraints of DeskVR, environmental interactions must be re-evaluated, especially regarding selection, manipulation, and travel techniques. Because users remain seated in DeskVR, this enables interactions not common in typical VR scenarios, such as touch-based interactions on a desk. Zielasko et al. [69] explored this concept through the evaluation of four menu interaction arrangements in DeskVR: "Desk" aligns menus with the virtual desk, "Air" aligns menus with the user's task, and "DeskPlus" and "AirPlus" combine the previous scenarios with a physical desk to provide passive haptics. These configurations can be seen in Figures 2.8 and 2.9.

The results revealed diverse individual preferences for menu configurations. Some favored desk-aligned menus for tangibility, while others found it odd to touch a menu expected to be virtual. Additionally, menus aligned with the virtual desk required more head movement, making them less efficient. However, variants with a physical desk were favored for reducing physical strain.

Sousa et al. [59] introduced VRRRRoom, a virtual reality radiology reading room with a desktop surface for interacting with medical images. Medical images are displayed in 3D above a

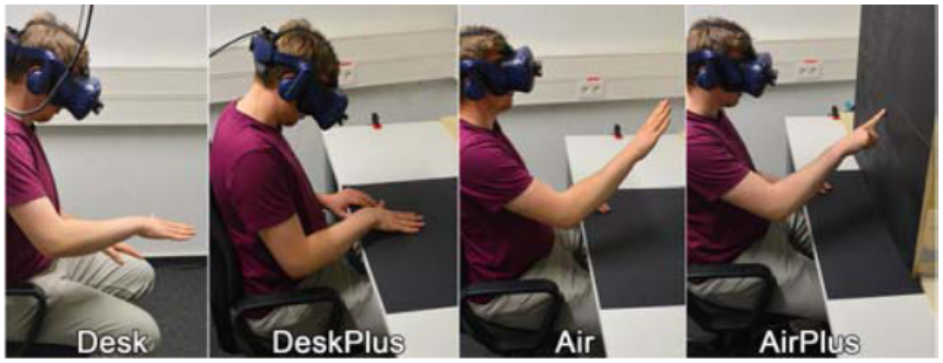


Figure 2.8: The four different scenarios studied in [69]. The "Desk" scenario aligns the menu with a virtual desk, while the "Air" scenario aligns the menu with the task. The "DeskPlus" scenario aligns the virtual desk with a physical desk, while "AirPlus" aligns the menu with the taks and a vertical board.

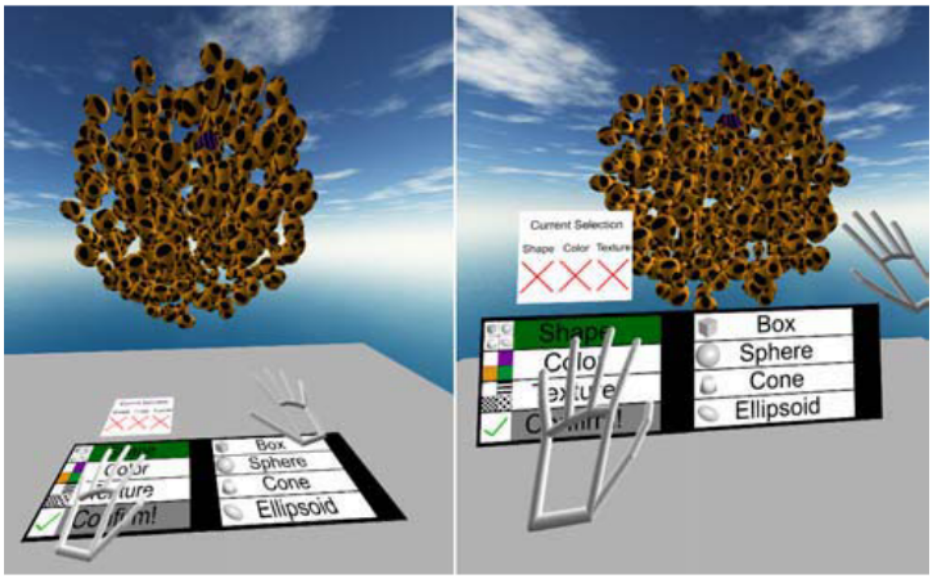


Figure 2.9: Experimental settings in [69] for desk-aligned menus on the left and task-aligned menus on the right.

virtual desk and can be manipulated using indirect touch controls. Users can change volume slices and adjust brightness with their left hand, while rotating the image and changing the scale can be done with their right hand, as depicted in Figure 2.10. The gesture-based interaction minimizes the need for users to move their head to view controls, as illustrated in Figure 2.11.

The visualization of controls on a virtual desk, exemplified in VRRRRoom [59] and illustrated in Figure 2.11, serves as a valuable tool for instructing users about available gestures and conveying information such as the current volume slice. Zielasko et al. [72] conducted an experiment to assess the impact of introducing this virtual desk into the virtual environment, analyzing its effects on performance, cybersickness, and presence. Their findings indicated no significant dif-

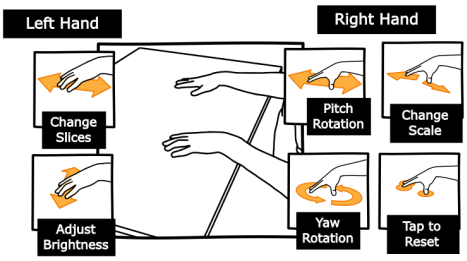


Figure 2.10: Gesture dictionary in [59]

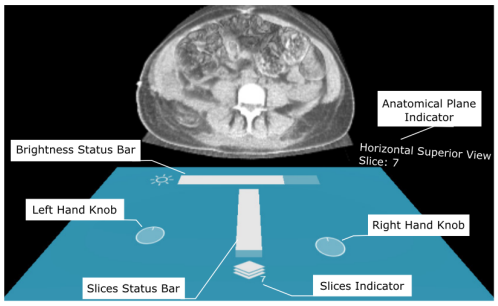


Figure 2.11: Virtual desk, control indicators, and medical images rendering in [59]

ferences in the evaluated aspects, suggesting the potential for expanding the desk functionality to incorporate menus or controls.

Regarding the selection and manipulation of 3D objects, many techniques are viable in both standing and sitting positions. However, techniques requiring extra controllers may be less suitable, as finding them after being set down while immersed in an HMD can be difficult [71]. Almeida et al. [2] developed a DeskVR-specific object manipulation technique called SIT6. This indirect touch-based technique utilizes a gesture dictionary with separated degrees of freedom: three for translation and three for rotation, as seen in Figure 2.12. Indirect touch means that users do not directly touch the object through a screen display but interact with it indirectly, as the former would not be practical in VR using an HMD [36]. While SIT6 may not be as fast as other state-of-the-art mid-air techniques, it was found to be as effective and less physically demanding.

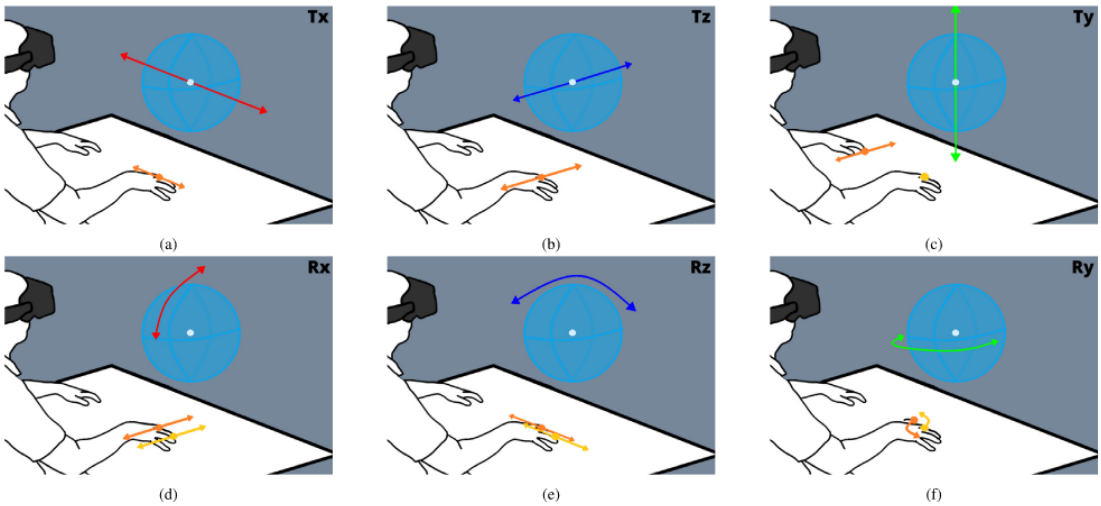
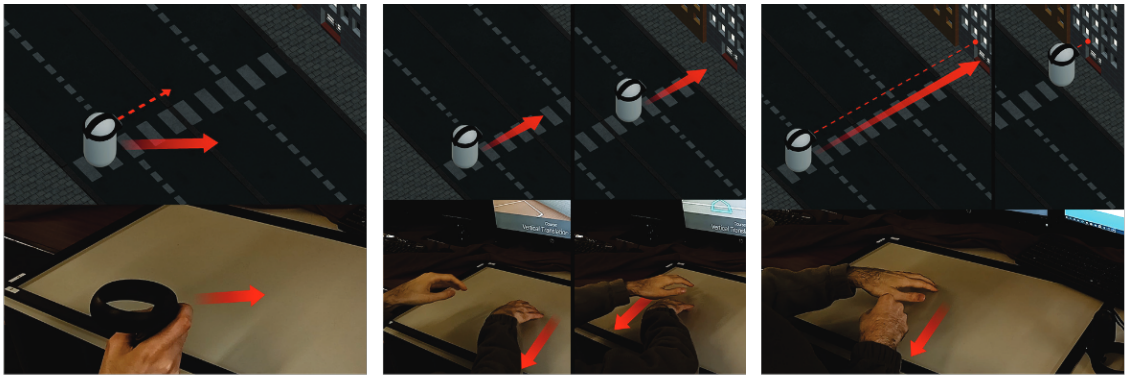


Figure 2.12: The gesture dictionary of SIT6 [2]. Gestures (a), (b), and (c) represent translation movements, while gestures (d), (e), and (f) represent rotation movements.

Designing traveling techniques in DeskVR is challenging because users are seated, rendering real walking impractical despite its potential presence enhancement. While presence is crucial for reducing cybersickness [71], an effective travel technique in DeskVR should aim to enhance

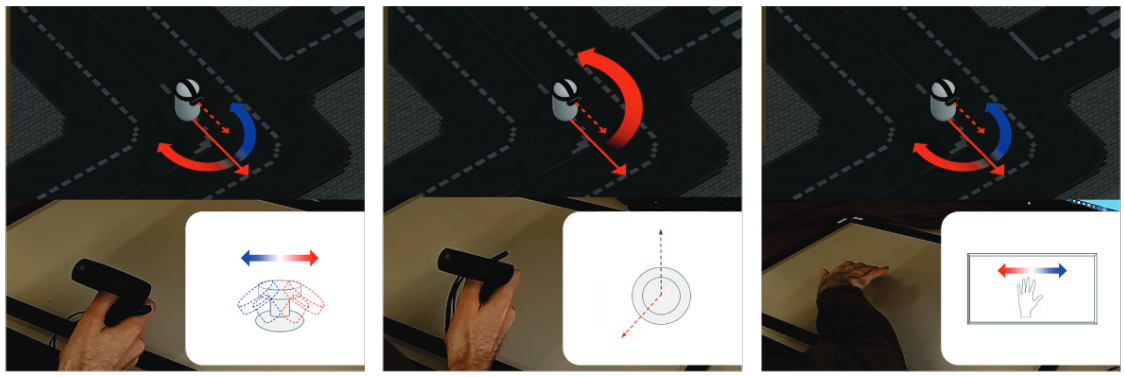
presence without necessitating tiring motions or additional body tracking, such as leg movements. Amaro et al. [3] devised three travel and four orientation techniques, seen in Figure 2.13. One travel technique utilized a VR controller named Continuous Directional Movement. In contrast, the others employed a touch surface and gestures, known as Dog Paddle and Drag'n Go.



(a) (b) (c)

Figure 2.13: Travel techniques designed for DeskVR by Amaro et al. [3]: Continuous Directional Movement (a), Dog Paddle (b), Drag'n Go (c).

For orientation, Amaro et al. [3] introduced two techniques utilizing VR controllers, named Continuous Directional Rotation and Choose & Click, one technique utilizing a tactile surface, named Tactile Surface Dragging, and one technique relying on the orientation of the user's head, named Gaze Convergence. These techniques are illustrated in Figure 2.14. Among the movement techniques, Continuous Directional Movement outperformed its counterparts in performance and comfort, although it appeared to induce more nausea than Dog Paddle. The discomfort in Dog Paddle could stem from its repetitive, exaggerated motions, indicating a potential avenue for exploring more straightforward and less straining interactions. Concerning orientation, Tactile Surface Dragging exhibited superior overall performance, showing fewer cybersickness symptoms than Continuous Directional Rotation, which exhibited the most.



(a) (b) (c)

Figure 2.14: Orientation techniques designed for DeskVR by Amaro et al. [3]: Continuous Directional Rotation (a), Choose & Click (b), Tactile Surface Dragging (c).

2.4 Discussion

The key takeaway in developing interactions for DeskVR is prioritizing user comfort and leveraging the potential benefits of having a desk in front of the user while being mindful of the associated constraints. Numerous studies have investigated the effectiveness of touch-based interactions in both non-steroscopic contexts, such as Balloon Selection [5], Corkscrew Selection [10], and Triangle Cursor [61], and in the context of DeskVR, exploring various scenarios such as menu navigation [69], medical image data examination for radiologists [59], travel techniques [3], and object interaction [2]. The consensus from these studies suggests that, although a touch-based approach may not always be the most time-efficient, it significantly reduces physical strain, enhancing overall comfort and prolonged usage time for individuals. This aspect is essential for DeskVR, where the primary objective is to alleviate physical strain, improve accessibility, and boost productivity [71, 70]. For this reason, touch-based approaches seem preferable over mid-air interactions. Furthermore, Zielasko et al. [72] demonstrated that presenting a virtual table as a surrogate for the physical table has minimal impact on task performance, cybersickness, and presence, making it a viable option for menus and other interactions without compromising the user experience.

The absence of physical feedback during simultaneous interactions by multiple users creates unpredictability in real-time multi-user collaboration [51]. While haptic devices like the Phantom Omni [55] offer a potential solution by allowing users to manipulate objects and perceive forces applied by others [50], their integration into DeskVR, potentially involving substitutional reality – replacing real physical objects such as haptic devices with virtual counterparts [56], is beyond the scope of this work. Furthermore, the limited mobility of users in DeskVR constrains simultaneous multi-user object interaction, particularly when considering physics-based interaction techniques such as concurrency control using consensus-based networking [18].

Considering these factors, level 3 collaboration is unsuitable, and thus, there is no need for distributed average concurrency control techniques. Additionally, degree-of-freedom separation techniques appear restrictive in configuration and interaction while potentially confusing and difficult to use. As such, the fine-grained approach of Lee et al. [29] is appealing, allowing users to discuss different object interactions and arrangements while sharing their perspectives. Relevant work has applied this concept in VR, such as the implementation of object previews in the work by Pereira et al. [42]. Extending this concept to DeskVR is a promising avenue for exploration.

Moreover, providing social visibility, awareness, and accountability in collaborative environments is pivotal in shaping how we communicate [16, 22]. The chosen approach should provide these elements, enabling users to communicate more effectively. For instance, it should indicate who is interacting with an object, its owner, and who is engaged in the conversation, seamlessly integrating these aspects within the DeskVR environment. Additionally, it should leverage the spatial attribute of VR, incorporating a spatial model of interactions [4, 13]. Users within the proximity of an object's aura could trigger interactions, such as audio communication, enabling them to focus on the task at hand with fewer distractions. This aligns with the concept of translucence as discussed by Erickson and Kellogg [16].

Chapter 3

Replico

After evaluating the literature, there appears to be a promising avenue for exploring a method for DeskVR users to engage in discussions and communicate about various objects and areas of interest within a shared virtual environment. This approach could enhance communication effectiveness, especially considering the challenges associated with verbally referencing objects and describing spatial locations [41]. Moreover, given the unique constraints of DeskVR, a research gap exists in applying such methods to this specific environment. Thus, this work introduces Replico, a collaborative touch-based DeskVR approach that utilizes the world-in-miniature (WIM) [60] metaphor to facilitate reasoning about 3D models.

3.1 Overview

After recognizing the potential of a collaborative DeskVR approach to enhance communication and reasoning about 3D models, it became important to outline the basic needs such an approach should meet. These requirements stem from the desire to address common challenges faced by users in virtual environments, aligning with DeskVR's goals of minimizing physical effort to enable longer periods of productive work, reducing mistakes to prevent frustration, and getting more done in less time. Additionally, the approach should be easy to understand to facilitate seamless collaboration with others and allow users to communicate effectively about objects and areas of interest even when they are out of sight.

Furthermore, users must know where others are, who they are, and what they are doing in the virtual space to work well together. Users should be able to easily determine the location and activities of their counterparts so they can coordinate and talk effectively. Importantly, all interactions and tasks should be achievable while seated so they can keep working without getting too tired.

To meet these needs, the chosen approach utilizes the world-in-miniature (WIM) metaphor [60]. The WIM is a miniature replica of the virtual environment that can be easily manipulated within arm's reach and viewed from multiple angles, making it a good fit for DeskVR. Changes

made in the miniature model are reflected in the full-scale model. Additionally, the WIM is effective for displaying social information, such as users' locations and where they are looking, as well as indicating who is working. Each user has a personal view of the WIM, while the to-scale model is shared.

The approach allows users to create points of interest to facilitate communication about objects and zones of interest. These points are uniquely identified by a number and an appearance corresponding to their owner. For instance, one user's points of interest might be represented by green striped spheres, while purple checkerboard cubes represent another user's points. These points of interest are visible in both the to-scale 3D model and its WIM counterpart and remain visible even if occluded.

Users are attached to virtual tables that correspond to their real-life counterparts. They can either remain at separate tables or join another user's table to share the same point of view of the to-scale model. Users can also teleport around the 3D model, taking their table with them if they are alone or creating a new one if they are at someone else's table.

To reduce physical effort, the approach uses touch-based gestures for interactions, such as moving the replica and creating points of interest. Literature suggests that touch-based interactions help reduce physical strain and increase comfort [5, 2]. Specifically, the approach incorporates the Balloon Selection metaphor [5] for creating points of interest. Balloon Selection was chosen mainly due to its "String Height Clutching" feature, which allows users to increase the height of the selection cursor infinitely. This is in contrast to the Triangle Cursor [61], and it is less straining than Corkscrew Selection [10], which requires rotational movement.

3.2 Actions

This section details the specific actions users can perform within Replico using touch-based gestures. These actions encompass various interactions and transformations. Similar to Balloon Selection [5], these gestures are handedness-agnostic, meaning they can be performed with either the left or right hand. The following subsections outline the key actions and how to perform them.

3.2.1 Replica Transformations

Transformations to the replica closely resemble the common gestures associated with touch screens on mobile phones for zooming, panning, and rotating images. Translation in the XZ plane is done by placing one or more fingers on the touch table and moving them. Yaw rotation around the Y-axis is achieved by rotating the fingers around the center of their positions, which also serves as the center of rotation. Scaling the replica in all dimensions is accomplished using a pinch gesture, with the scaling base centered on the midpoint of the fingers and the WIM's base y position.

Translation along the Y-axis is achieved by first placing a finger from the primary hand on the table, followed by two fingers from the secondary hand. The fingers of the secondary hand control translation on the XY plane, while the fingers of the primary hand can perform all the previously mentioned transformations.

3.2.2 Balloon Selection

Replico uses the Balloon Selection metaphor [5] to create, delete, and acknowledge points of interest, as well as teleportation, and join another user's table. To initiate the balloon selection gesture, the user places a finger from their primary hand on the table, followed by a finger from their secondary hand. This action brings up a balloon on the user's table relative to the WIM, with a copy visible on the to-scale 3D model. The primary finger controls the XZ position of the balloon while moving the fingers apart lowers the balloon, and bringing them together raises it. Users can remove the secondary finger without changing the balloon's height, allowing for "String Height Clutching."

To perform a *selection*, the user briefly adds a second finger from their secondary hand to the touch table. This *selection* action can create points of interest at the balloon's position, delete points of interest if the balloon intersects with one of the user's points, acknowledge points of interest if the balloon intersects another user's unacknowledged points, and join other users' tables if it intersects another user's table.

For *teleportation*, the user similarly adds a second finger from their secondary hand to the touch table but holds it in place until an arrow appears on the balloon, indicating the teleportation's end orientation. To rotate the balloon, the user performs a gesture similar to the replica's rotation gesture by rotating their fingers around their midpoint. Either hand can be removed to reposition, but removing both hands cancels the teleportation. To confirm the teleportation, the user taps again with the second finger of the secondary hand.

3.3 Awareness

Following Table 2.2, Replico incorporates most elements of workspace awareness. Users have distinct appearances that uniquely identify them, which is also reflected in their points of interest, albeit with some distinguishing features. User tables are represented in the WIM by outlined miniatures, positioned and oriented as they are in the to-scale model, and are visible behind objects. These miniature tables display who is at each table through miniature representations of the users.

Points of interest created by other users are initially in an *unacknowledged* state, marked by a vertical line capped with a symbol that resembles the owner's appearance and includes the balloon's identification number. This system, combined with the acknowledgment gesture described in 3.2.2, allows users to distinguish recently created points of interest from previously created ones.

These representations blend abstract and mimetic representations of social information as described in [16]. They are mimetic in representing real-world entities like tables and user avatars, yet abstract as they are symbolic and easy to create and manipulate.

Thus, Replico allows users to ascertain the presence of others in the workspace, their locations, and their viewing directions by observing the tables in the replica. Users can identify others by

their appearance, determine collaborators by noting who shares the same table, and discern authorship through the appearance of points of interest. Points of interest, in turn, indicate which objects users are working on. Additionally, users can track recent activities by checking unacknowledged points.

3.4 Summary

This chapter introduces Replico, a collaborative DeskVR approach designed to enhance communication and interaction in the analysis of 3D models. Using the world-in-miniature (WIM) metaphor, Replico addresses common challenges in virtual collaboration, such as spatial referencing and awareness of other users' activities.

First, it explains the requirements such an approach should consider, such as minimizing physical effort, reducing mistakes, ensuring efficiency, effective communication about objects and areas of interest, ease of understanding, providing awareness of other users, and enabling all interactions to be done while seated. To meet these requirements, Replico allows users to manipulate a miniature replica of the virtual environment (WIM). This approach ensures that changes made in the miniature are reflected in the full-scale model. Replico also enables the creation of uniquely identified points of interest, aiding the communication about objects and zones of interest within the virtual space. Users are attached to virtual tables corresponding to their real-life counterparts, allowing them to join others' tables or teleport around the 3D model.

Section 3.2 explains the different touch-based gestures users can perform within Replico: translation, rotation, and scaling of the WIM; and creation, deletion, and acknowledgment of points of interest, as well as teleportation and joining users' tables, using balloon selection. Section 3.3 explains how Replico incorporates workspace awareness: users and their points of interest share a unique appearance, making them identifiable; user tables are represented in the replica; and points of interest can be acknowledged to identify recent points of interest quickly.

Chapter 4

Implementation of a Prototype

This chapter describes the implementation of the prototype created to test the viability of Replico. It begins by explaining the architecture, hardware, software, and tools used to develop the prototype. The chapter then details how hand detection is performed, the various states within the system, how tables are tracked, how finger input is interpreted, the methods for displaying feedback to users, and the networking implementation details.

4.1 Architecture

To develop the prototype, two HTC Vive Pro 2 headsets and two multi-touch surfaces – a 32-inch infrared frame and a 47-inch capacitive Displax Skin Ultra[1] touchscreen – were used. The Unity[2] game engine was chosen for its robust VR support, personal previous experience, extensive community resources, and excellent compatibility with external C# libraries, which helped reduce development risks.

Unity's OpenXR plugin[3] controls VR hardware communication, handling input and rendering with minimal effort, and managing all API calls automatically. OpenXR[4] is an API standard developed by Khronos for XR applications, including VR, and is widely adopted across many XR devices with conformant runtimes. Due to its ubiquity and recency, it was chosen over alternatives such as using SteamVR directly.

Since Unity's OpenXR plugin interfaces with Unity's Input System, the Unity Enhanced Touch API[5] was used instead of the standard Unity Touch API to maintain a consistent input management system. The Enhanced Touch API provides automatic finger tracking and keeps a history of touch interactions.

[1] https://www.displax.com/skin-ultra
[2] https://unity.com/
[3] https://docs.unity3d.com/Packages/com.unity.xr.openxr@1.11/manual/index.html
[4] https://registry.khronos.org/OpenXR/specs/1.1/html/xrspec.html
[5] https://docs.unity3d.com/Packages/com.unity.inputsystem@1.0/api/UnityEngine.InputSystem.EnhancedTouch.html

The first-party Unity Netcode for GameObjects[6] library was used for networking. This library offers a straightforward abstraction of networking logic and is easy to use and set up for client-server or distributed authority topologies. Its simplicity is well-suited for a small number of clients, making it ideal for this prototype.

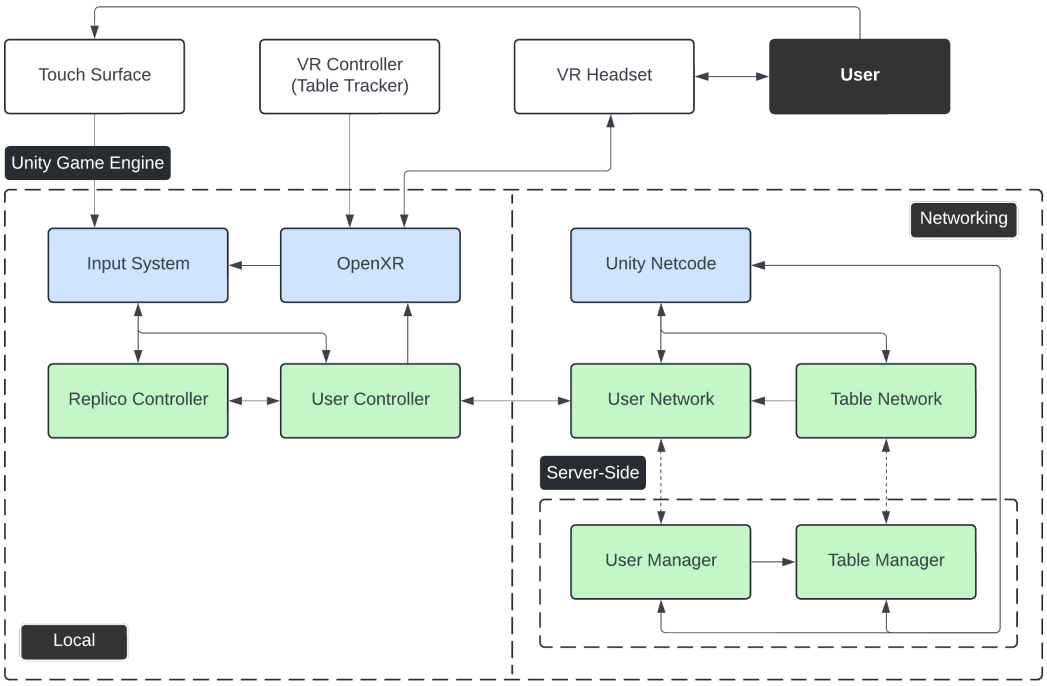


Figure 4.1: System architecture. Modules in blue represent Unity libraries, while the prototype implements modules in green.

The system architecture diagram in Figure 4.1 demonstrates the integration of various software and hardware components, describing both local and networked elements. OpenXR manages communication with the VR headset and the table tracker, rendering images to the headset, updating the virtual camera's position, and interfacing inputs to Unity's Input System. Unity's Enhanced Touch API processes touches sensed by the touch surface. The Replico Controller and User Controller manage interaction logic and user actions, communicating with the Input System and OpenXR. Networking is handled by Unity Netcode, which synchronizes user and table data across the network via the User Network and Table Network, using a client-server topology. The server-side User Manager and Table Manager manage data for users and tables. The User Manager tracks users in the system and communicates with the User Network objects through remote procedure calls (RPCs) and network variables. The Table Manager handles table creation, deletion, and management logic, communicating with Table Network objects using RPCs and network variables.

[6] https://docs-multiplayer.unity3d.com/netcode/current/about/

4.2 State Machine

The interpretation of touch surface input varies depending on the gestures employed. To address this, a state machine was created. A state machine consists of defined states with distinct transitions, where each state processes the input differently. The implementation was done through the state design pattern, wherein each step is represented as a class that alters the behavior of the Replico controller. Figure 4.2 provides a diagram illustrating the implemented state machine.

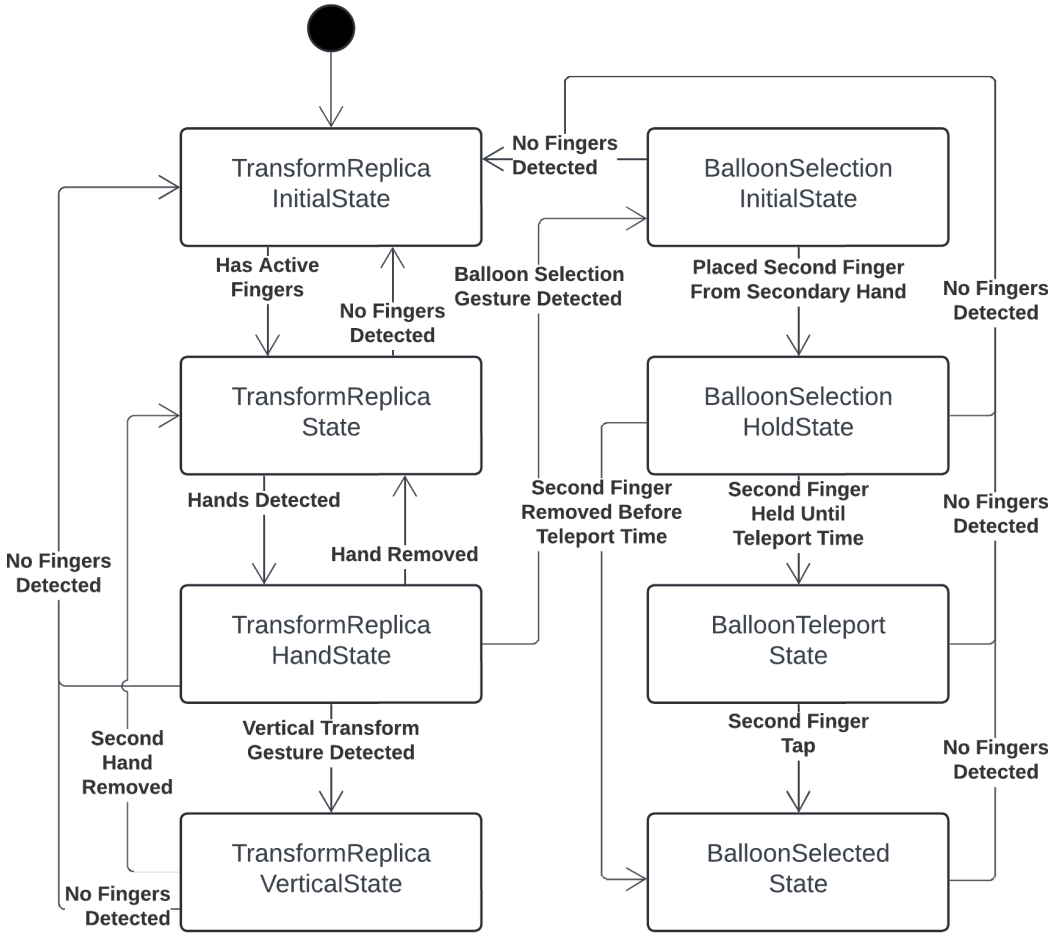


Figure 4.2: State machine diagram.

The `TransformReplicaInitialState` is the starting state where no fingers are detected, and therefore no controls are active. In this state, finger touches are checked every frame using the Enhanced Touch API, which updates every frame because the Input System update mode is set to Dynamic. When at least one finger is detected, the system transitions to `TransformReplicaState`. All states revert to this initial state whenever no fingers are detected.

The `TransformReplicaState` is entered when fingers are detected, but no hands are recognized yet. In this state, the user can move the replica using the gestures described in Section 3.2.1 and

implemented in Section 4.3. This state performs hand detetction using the method described in Section 4.4.1, every frame. When both hands are detected, it transitions to `TransformReplicaHandState`.

In `TransformReplicaHandState`, users can move the replica as in `TransformReplicaState`. This state checks what gesture the user is doing in every frame using the method described in Section 4.4.2. It transitions to `TransformReplicaVerticalState` if the vertical transform gesture is detected, or to `BalloonSelectionInitialState` if the balloon selection gesture is detected. If any hand is removed, it transitions back to `TransformReplicaState`.

In `TransformReplicaVerticalState`, the user can move the replica as in `TransformReplicaState` using their primary hand, while the secondary hand performs translation on the XY plane. To allow users to temporarily remove the secondary hand to reposition their fingers without terminating the gesture, the secondary hand may be removed for up to 0.55 seconds before the controller transitions back to `TransformReplicaState`.

In `BalloonSelectionInitialState`, the user performs the balloon selection gesture as described in Section 3.2.2. The primary finger moves the balloon on the XZ plane, while moving the fingers together raises the balloon and moving them apart lowers it. The balloon's height is saved between gestures, so the user doesn't have to raise it each time they start balloon selection. It transitions to `BalloonSelectionHoldState` when a second finger is added to the secondary hand. Instead of transitioning instantly to the initial state when no fingers are detected, there is a grace period of 0.15 seconds. This grace period accounts for potential hardware tracking failures that could prematurely stop the balloon selection gesture, potentially causing frustration.

The `BalloonSelectionHoldState` monitors how long the user holds down the second finger from the secondary hand. If the finger is held for 0.4 seconds or the primary hand moves, the state transitions to `BalloonTeleportState`. Removing the finger before 0.4 seconds results in different actions: joining a table if the balloon intersects one, acknowledging or deleting a point of interest if it intersects one, or creating a new point of interest if it intersects none, thereby transitioning to `BalloonSelectedState`.

In `BalloonTeleportState`, the user can rotate their balloon using the same gesture as rotating the replica, following the calculations described in 4.3. Confirmation of teleportation occurs if the user taps with the second finger of their second hand (removing and placing it again) or if two touches are detected within the hand detection distance outlined in 4.4.1. Upon confirmation, the controller transitions to `BalloonSelectedState`.

The purpose of `BalloonSelectedState` is to act as a buffer between balloon selection interactions and replica transformations. Users can only perform replica transformations after removing all finger touches, at which point the controller transitions back to `TransformReplicaInitialState`.

4.3 Replica Transformations

This section discusses the transformations applied to the replica in states `TransformReplicaState`, `TransformReplicaHandState`, and `TransformReplicaVerticalState`.

Translation is achieved by calculating the distance $\vec{D} = C_i - C_{i-1}$, where C_i and C_{i-1} are the centers of the active touches on the touch surface from the current frame and the previous frame, respectively, as shown in Equation 4.1.

$$\begin{aligned} \mathbf{F}_i &= \{(x_{i_k}, y_{i_k}) : \ k = 1, \ \ldots, \ n\} \\ \mathbf{minF}_i &= (\min_k x_{i_k}, \ \min_k y_{i_k}) \\ \mathbf{maxF}_i &= (\max_k x_{i_k}, \ \max_k y_{i_k}) \\ C_i &= \frac{\mathbf{minF}_i + \mathbf{maxF}_i}{2} \end{aligned} \tag{4.1}$$

Here, $\mathbf{F}_i$ represents the positions of the n fingers in the i-th frame, $\mathbf{minF}_i$ and $\mathbf{maxF}_i$ are the minimum and maximum x and y coordinates from the set of finger positions. Essentially, C_i is the geometric centroid of the smallest rectangle that can enclose all the touch points. The distance $\vec{D}$ is then multiplied by a factor t, which can be adjusted for each 3D model, resulting in $\vec{T} = t \cdot \vec{D}$. This yields a 2D vector that is added to the replica's position in the XZ plane.

Scaling is achieved by first calculating $s = \left(\bar{d}_i / \bar{d_{i-1}}\right)^c$, where $\bar{d}_i$ and $\bar{d_{i-1}}$ are the average distances of the active touches to the center of those touches from the current frame and the previous frame, respectively, and c is a constant used to modulate the scaling effect. The calculation for the average distance is shown in Equation 4.2.

$$\bar{d}_i = \frac{\sum_{k=1}^{n} \| C_i - F_{i_k} \|}{n} \tag{4.2}$$

Finally, this scaling is applied to the replica around a base point, based on the center of the fingers. To do this, the pivot point in world coordinates, pivot_w, is first converted to local coordinates, pivot_l, relative to the replica. The replica's local scale is then multiplied by s. After scaling, the position of the pivot in world coordinates, pivot_k, calculated by converting pivot_l back to world coordinates, will not be equal to pivot_w. To correct this, the displacement $\Delta\vec{\text{pivot}} = \text{pivot}_w - \text{pivot}_k$ is calculated and added to the replica's world position.

Rotation is achieved by calculating the fingers' average rotation $\bar{\theta}$. The calculation for this is shown in Equation 4.3, where $\vec{\text{dir}}_{i_k}$ is the vector direction from the center of the fingers to the k-th finger in the i-th frame, $|\theta_k|$ is the angle between the vector direction of the previous frame and the current frame for the k-th finger. The angle θ_k is determined by adjusting $|\theta_k|$ based on the cross product's z-component to account for direction, as the arccosine function's range only goes from 0 to π. The replica is then rotated around the Y axis that passes through $(C_{i_x}, 0, C_{i_y})$ with the angle $\bar{\theta}$, using Unity's `RotateAround`.

$$
\begin{aligned}
\vec{\text{dir}}_{i_k} &= \mathbf{F}_{i_k} - C_i \\
|\theta_k| &= \cos^{-1}\left(\frac{\vec{\text{dir}}_{i-1_k} \cdot \vec{\text{dir}}_{i_k}}{\|\vec{\text{dir}}_{i-1_k}\|\|\vec{\text{dir}}_{i_k}\|}\right) \\
\theta_k &= \begin{cases} |\theta_k| & \text{if} \quad (\vec{\text{dir}}_{i-1_k} \times \vec{\text{dir}}_{i_k})_z < 0 \\ -|\theta_k| & \text{if} \quad (\vec{\text{dir}}_{i-1_k} \times \vec{\text{dir}}_{i_k})_z \geq 0 \end{cases} \\
\bar{\theta} &= \frac{\sum_{k=1}^{n} \theta_k}{n}
\end{aligned}
\tag{4.3}
$$

In the `TransformReplicaVerticalState`, the primary hand can perform XZ translation, rotation, and scaling, while the secondary hand can only perform translation on the XY plane. Only the fingers from the primary hand are considered for transformations with the primary hand. The secondary hand's fingers can only perform translation as previously described, but instead of $\vec{T}$ being applied to the replica's XZ position, it is applied to the XY position.

The transformations are not directly applied to the replica; instead, they are applied to a target object. The replica then follows this target object using Unity's `SmoothDamp` for position and scale, and `SmoothDampAngle` for each Euler rotation angle. This helps to reduce jitter caused by low-frequency or low-accuracy touch input updates.

4.4 Gesture Detection

This section explains how the different gestures – transformation, vertical transformation, and balloon selection – are distinguished. Section 4.4.1 explains the method for detecting and distinguishing the user's hands, and Section 4.4.2 describes how the vertical transform and balloon selection gestures are differentiated.

4.4.1 Hand Detection

Detection and distinction of hands are important for recognizing Replico's touch-based gestures. The prototype took a simplistic approach to hand detection, using input solely from the touch surface and Unity's Enhanced Touch API. The method involves detecting two clusters based on finger proximity. For this purpose, this prototype uses a naive K-Means clustering algorithm [33, 32, 63] with a k value of 2. The algorithm was implemented using the ML.NET library[7], a machine learning library for .NET. For compatibility with Unity, the .NET Standard 2.1 version was used. The distance function used is the Euclidean distance between the finger positions on the touch surface in pixels, and the Yinyang initialization algorithm [11] is applied.

Other clustering algorithms, such as DBSCAN [17], were not used because they do not allow the specification of a fixed number of clusters (k) or because they perform better with a larger

[7] https://github.com/dotnet/machinelearning

number of clusters. K-means is perfectly adequate in this case with a maximum of 10 points (one for each finger) and only 2 clusters.

The K-means algorithm returns two clusters of fingers when more than one finger is placed on the touch surface. However, this can result in two fingers from the same hand being detected as separate clusters. To address this, a distance threshold between cluster centroids is used to determine if the clusters represent separate hands. The threshold distance is measured relative to a min-max normalized value of the screen dimensions in pixels. A distance threshold of 0.18 was found to be effective through testing.

Initially, before any hands have been detected in the `TransformReplicaState` shown in Figure 4.2, the distinction between the primary and secondary hands is maintained by queuing fingers based on the order of their touch. The cluster containing the first detected finger represents the primary hand. Once both hands are detected in the `TransformReplicaHandState` and beyond, hands are updated each frame by reapplying the K-means algorithm. The distinction is then made by counting how many fingers from the previously detected primary hand are in each newly detected cluster; the cluster with the most fingers from the primary hand is associated with it. To update the hands, new fingers in the clusters are added to the corresponding hands, and previously detected fingers remain in their respective hands unless they have been removed from the touch surface. This approach allows left-handed and right-handed users to perform all Replico gestures easily.

4.4.2 Distinguishing Vertical Transform and Balloon Selection

The vertical transform and balloon selection gestures, described in Sections 3.2.1 and 3.2.2 respectively, can be easily confused with the pinch gesture required for scaling the replica. To aid in this distinction:

- **Vertical Transform:** At least one finger on the primary hand and at least two fingers on the secondary hand. The secondary hand must remain stationary for 0.2 seconds.
- **Balloon Selection:** Exactly one finger on the primary hand and exactly one finger on the secondary hand. Both hands must remain stationary for 0.2 seconds.

The vertical transform only checks if the second hand has moved, allowing the user to add the secondary hand while transforming with the first. The criterion for determining if a hand hasn't moved is that none of its fingers have moved past a threshold δ from the position where they were first placed. This threshold is measured relative to a min-max normalized value of the screen dimensions in pixels, with testing indicating 0.01 as an appropriate value. Once a finger has moved, it will be considered moved until it is removed.

4.5 Table Tracking

The user's table is tracked to a real-world table using a VR controller, as shown in Figure 4.3. The controller is positioned pointing toward the user instead of forward from the table so that the

tracked controller position matches the table's corner. If it pointed the other way, a translation based on the controller's length toward the table corner would be necessary, which is not feasible for all controller types.

Figure 4.3: Tracking the table using a VR controller.

When attaching a user to a virtual table, such as when the user joins a table or teleports, the orientation and position of the tracker must match the orientation and position of the table's attach point, which is an empty GameObject on the table's Prefab. To achieve this, the user's orientation is first updated using `MatchOriginUpCameraForward`, followed by updating the position with `MoveCameraToWorldLocation`, both functions from the OpenXR plugin.

The `MatchOriginUpCameraForward` function requires two parameters: an up vector and a forward vector. The up vector is set to match the attach point's up vector, assuming the user and controller are on flat ground. The forward vector, $\vec{v}_{\text{forward}}$, is calculated as shown in Equation 4.4. Here, $\mathbf{q}_{\text{tracker}}$ represents the quaternion rotation of the tracker, and $\mathbf{q}_{\text{attach}}$ represents the quaternion rotation of the table's attach point. $\Delta\theta$ is the rotation of the attach point relative to the tracker. The yaw component is isolated to ignore pitch and roll, preventing rotation along the x and z axes due to the controller rolling, assuming the user is on flat ground. $\mathbf{q}_{\text{target}}$ is the user's target rotation, combining the user's current yaw rotation with the relative rotation of the attachment point. Finally, the forward vector is obtained by multiplying $\mathbf{q}_{\text{target}}$ by the (0,0,1) vector. The `MoveCameraToWorldLocation` function requires a position parameter. This position is calculated using the equation $P = P_{\text{attach}} + P_{\text{user}} - P_{\text{tracker}}$.

Because the controller may fall accidentally while the user is interacting with the touch surface and cannot see it when using the VR headset, it is only used to track the table when the user first joins a table. To achieve this, the tracker's rotation relative to the user is calculated using $\mathbf{q}_{\text{local}} = \mathbf{q}_{\text{tracker}} \cdot \mathbf{q}_{\text{user}}^{-1}$ and stored. Additionally, the tracker's local position relative to the user is calculated using the user's `InverseTransformPoint` function and stored.

$$\begin{aligned}
\Delta\theta &= \mathbf{q}_{\text{tracker}}^{-1} \cdot \mathbf{q}_{\text{attach}} \\
\Delta\theta_Y &= \text{Quat}\,(0, \text{yaw}(\Delta\theta), 0) \\
\mathbf{q}_{\text{user}_Y} &= \text{Quat}(0, \text{yaw}(\mathbf{q}_{\text{user}}), 0) \\
\mathbf{q}_{\text{target}} &= \mathbf{q}_{\text{user}_Y} \cdot \Delta\theta_Y \\
\vec{v}_{\text{forward}} &= \mathbf{q}_{\text{target}} \cdot (0, 0, 1)
\end{aligned} \tag{4.4}$$

To convert the local rotation back to world rotation, the calculation $\mathbf{q}_{\text{tracker}} = \mathbf{q}_{\text{user}} \cdot \mathbf{q}_{\text{local}}$ is used. Similarly, the local position is converted back to world position utilizing the user's `TransformPoint` function. These world coordinates and rotations are then used in the previously described calculations.

4.6 Visual Feedback

The prototype uses various visual indicators as forms of feedback. These include a virtual touch frame with finger indicators described in Section 4.6.1, a visual representation of the touch frame limits detailed in Section 4.6.2, tables and points of interest visible in both the 3D model and the replica as described in Sections 4.6.3 and 4.6.4, and effects related to balloon selection in Section 4.6.5, among others.

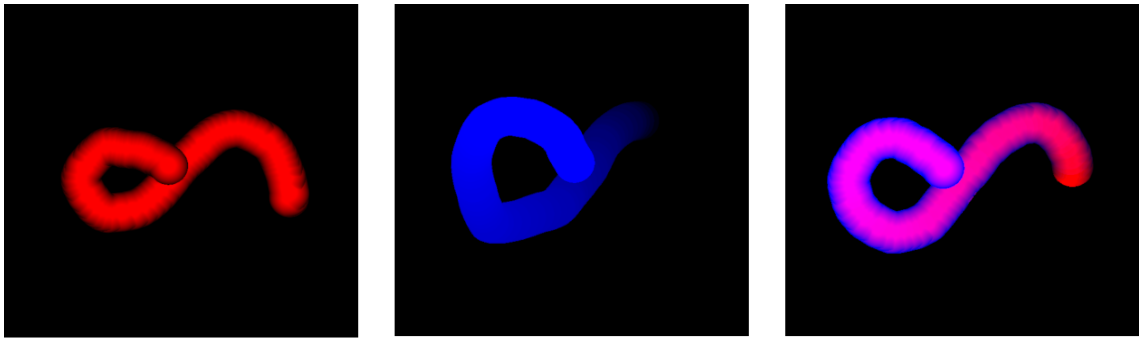
4.6.1 Virtual Touch Frame

In VR, users cannot see their hands or where their fingers are positioned. The prototype includes finger indicators within the virtual touch frame to address this issue, as depicted in Figure 4.4. Each finger is assigned a distinct color based on the order in which it was placed on the frame. Additionally, the finger trail shrinks from the current finger positions to previous positions, helping users understand their finger movements over time.

Figure 4.4: Touch indicators for four fingers on the touch frame.

This is implemented using a compute shader and a shader built with Unity's Shader Graph. A compute shader is a program that runs on the GPU outside of the normal rendering pipeline.[8] It is most useful for executing highly parallel algorithms. In this case, the compute shader processes finger positions, performs calculations, and stores results in render textures. The Shader Graph shader then uses these textures to render finger trail indicators in every frame.

Each render texture stores data for two fingers using two channels per finger per pixel. This results in five textures, each with dimensions matching the greatest nearest power of two between the screen width and height. One channel stores the reverse distance from the pixel to the center of the finger trail, ranging from 1 (closest to the center) to 0 (outside the trail radius), similar to a signed distance function. The other channel records the decay of the pixel, where 1 indicates the most recent position, and 0 indicates total decay. These channels are depicted in Figure 4.5, with red representing the distance to the trail's center and blue representing the decay.



a) b) c)

Figure 4.5: The different components stored on the render texture for each finger: a) reverse distance to center; b) decay; c) distance and decay combined.

A compute shader function is executed by several compute shader thread groups for each of the three dimensions: X, Y, and Z. In this case, the function defines for each group 8 threads for the X dimension, representing the pixel coordinate on the X axis, 8 threads for the Y dimension, representing the pixel coordinate on the Y axis, and 1 thread for the Z dimension, representing the render texture being processed. When the compute shader is executed, it runs using $\text{texture}_{\text{width}}/8$ groups on the X axis, $\text{texture}_{\text{height}}/8$ groups on the Y axis, and 5 groups on the Z axis, one for each render texture.

The compute shader takes several inputs: five different render textures, two `float4` structured buffers of size 5 (one for the current finger positions and one for the finger positions of the previous frame), a `float` structured buffer of size 10 that stores the average inclination of each finger trail, a linear decay rate δ, a finger radius in pixels r, and the time elapsed between the last frame and the current frame in seconds Δt. Each `float4` vector in the structured buffers represents the screen-space coordinates of two fingers, with the first two floats for one finger and the next two floats for another finger.

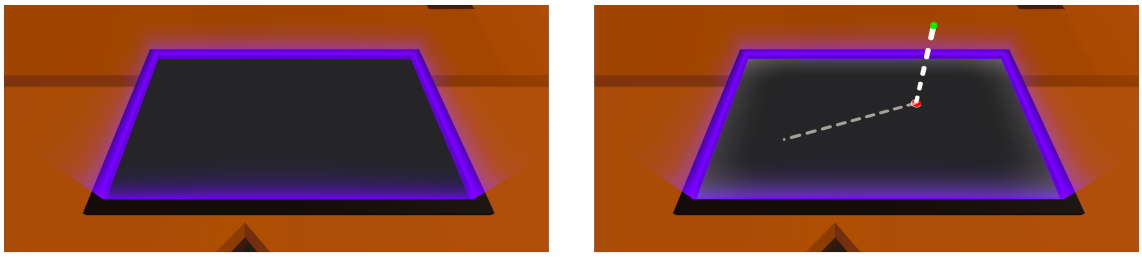
[8] https://docs.unity3d.com/Manual/class-ComputeShader.html

A simplified description of the algorithm for each finger and each pixel operates as follows: first, it calculates the decay λ_i of the pixel using the previous frame's decay value λ_{i-1} by computing $\lambda_{i-1} - \delta \cdot \Delta t$. Next, it calculates the distance d and reverse distance d_{rev} from the pixel to the line segment that starts at the finger's position in the previous frame and ends at the current finger position. This calculation, shown in Equation 4.5, creates a capsule-like shape. In this equation, A is the finger's position in the last frame, B is the current finger's position, r is the line's thickness, and P is the pixel's position.

$$\begin{aligned} \vec{AB} &= B - A \\ \vec{AP} &= P - A \\ h &= \text{clamp}\left(\frac{\vec{AP} \cdot \vec{AB}}{\vec{AB} \cdot \vec{AB}}, 0, 1\right) \\ d &= \frac{\|\vec{AP} - h \cdot \vec{AB}\|}{r} \\ d_{\text{rev}} &= \text{clamp}\,(1 - d, 0, 1) \end{aligned} \tag{4.5}$$

Based on how much the pixel is in front of the line segment, considering the average inclination of the finger trail, the algorithm linearly interpolates between $\max\left(d_{\text{rev}}, d_{\text{rev}_{i-1}}\right)$ and d_{rev} to calculate the value to store in the render texture. This approach allows the finger trail to overlap at the front and smoothly blend behind. The final decay value stored in the texture is λ_i, with an additional 1 added if $d \leq 1$, clamped between 0 and 1.

Finally, the fragment shader samples each render texture. It subtracts the decay value from the distance to create a shrinking effect, applies a border by comparing the resulting value with a threshold, and assigns a color based on the finger's order.



a) b)

Figure 4.6: Glow effect indicating gesture detection on the touch surface. (a) The initial state with no gesture detected. (b) The glow effect activates when a gesture is detected.

The virtual touch frame illuminates with a glowing effect whenever the system detects vertical transform or balloon selection gestures, signaling the user that their gesture has been recognized. This glow effect, demonstrated in Figure 4.6, is achieved using a shader that uses a rounded box signed distance function.[9] The strength of the glow is animated using the function $-\left(2\sqrt{t} - 1\right)^2 +$

[9] https://www.shadertoy.com/view/Nlc3zf

1, where t represents the elapsed time since the animation began. Initially, this function rises quickly until the result reaches 1 at $t = 0.25$, after which it gradually diminishes. The animation halts at $t = 0.8$ and resumes from that point when the gesture concludes.

4.6.2 Frame Limit Indicator

The balloon's position on the XZ axis during balloon selection is constrained by the boundaries of the virtual touch frame. To help users understand these boundaries, even when the frame is obscured by the replica, a purple illumination effect outlines the limits. This effect is shown in Figure 4.7.

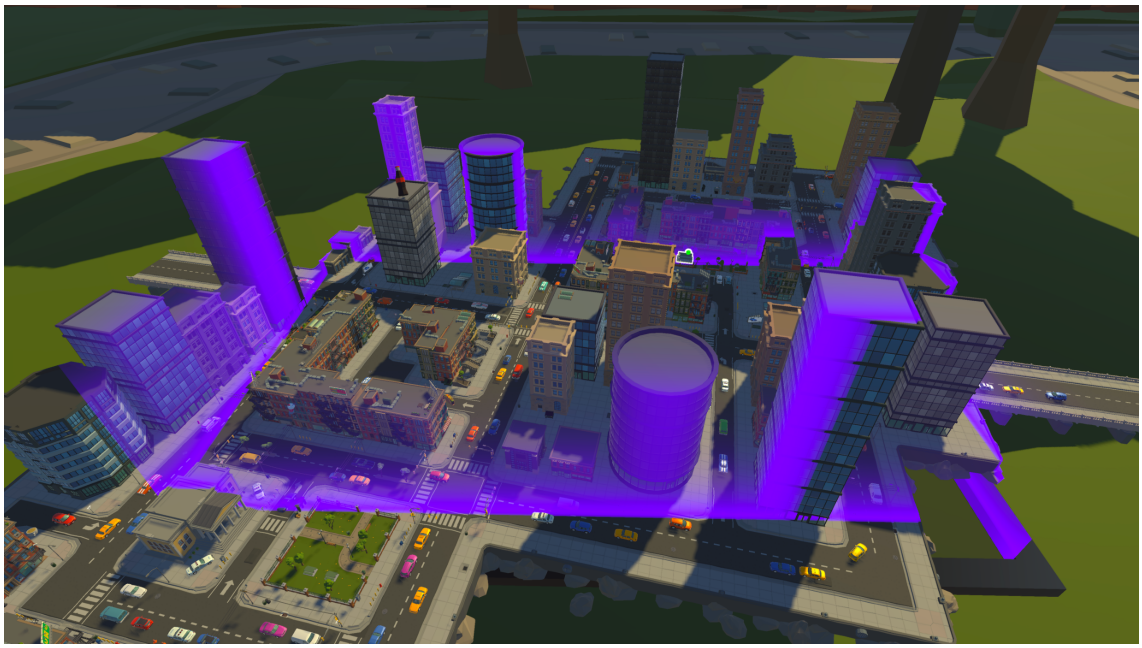


Figure 4.7: Illumination effect on the replica indicating the limits of the touch frame.

The illumination effect is achieved using a shader applied to a transparent rectangular prism extending from the touch frame's base. The shader's primary function is to gradually diminish the illumination effect as the distance from the prism increases. This is accomplished using a modified version of a shader initially designed for a stylized water effect[10], created using Unity's Shader Graph.

To calculate the distance d, a vector $\vec{CA}$ is obtained from the camera to the fragment's position on the prism using the View Vector node. This vector is then normalized to $\hat{v}$. The depth texture is sampled to obtain the distance from the camera to the point occluded by the prism, $|CB|$. The normalized vector is multiplied by this distance, resulting in $\vec{CB} = \hat{v} \cdot |CB|$. Adding this vector to the camera's position gives the position of the occluded point, $B = C + \vec{CB}$. The distance vector $\vec{BA}$ is obtained by subtracting the occluded point's position from the fragment's position on the

[10] https://ameye.dev/notes/stylized-water-shader/

prism's surface, $\vec{BA} = A - B$. Finally, the length of this vector is calculated to obtain the distance, $d = \|\vec{BA}\|$.

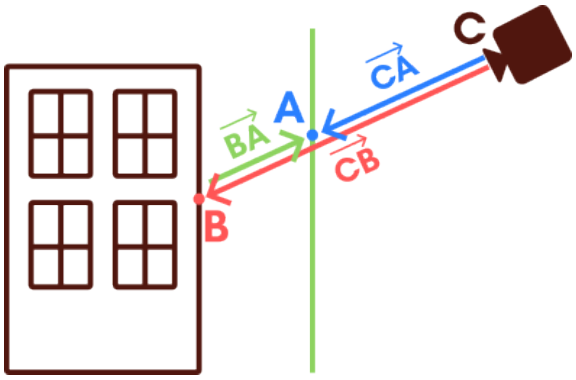


Figure 4.8: Diagram illustrating the steps to calculate d.

To achieve the gradual effect, the function $x = e^{-\frac{d}{0.05}}$ is applied. This ensures that when $d = 0$, the effect is at full power, declines rapidly, and then tapers off. This behavior is shown in graph a) of Figure 4.9. To soften the effect at the borders, the function described in Equation 4.6 and depicted in graph b) of Figure 4.9 is applied to x. This adjustment causes the effect to start at 0.2 power at the border, rise smoothly to 0.8 power, and then taper off as the distance increases. This progression is illustrated in graph c) of Figure 4.9.

The function $-37.5x^3 + 82.5x^2 - 60x + 15.2$ was derived from a cubic polynomial for creating a smooth curve between two points: (c,m) and $(k,m+b)$, shown in Equation 4.7.[11] In this case, the parameters are $c = 0.8$, $m = 0.8$, $k = 1$, $b = -0.6$, $p = 0$, and $q = -1.5$.

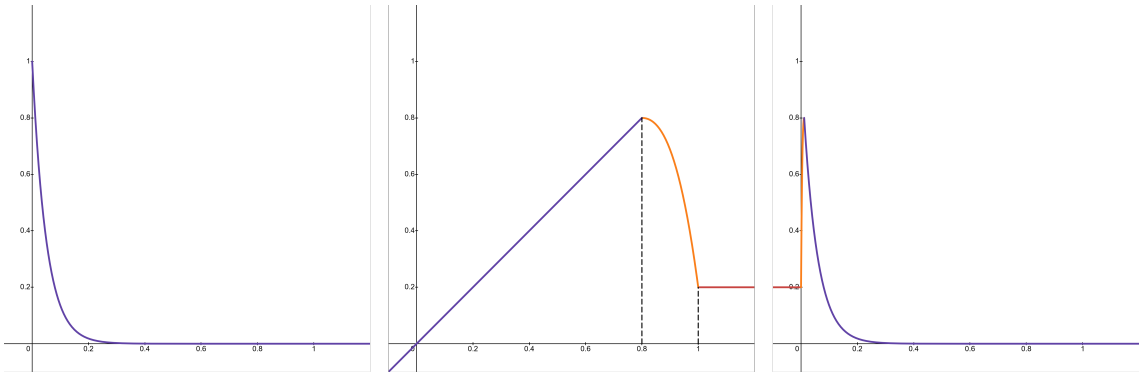
$$\alpha = \begin{cases} x & \text{if} \quad x \leq 0.8 \\ -37.5x^3 + 82.5x^2 - 60x + 15.2 & \text{if} \quad x > 0.8 \\ 0.2 & \text{if} \quad x > 1 \end{cases} \tag{4.6}$$

$$(p+q-2\cdot b)\cdot\left(\frac{x-c}{k-c}\right)^3 + (3\cdot b - 2\cdot p - q)\cdot\left(\frac{x-c}{k-c}\right)^2 + p\cdot\left(\frac{x-c}{k-c}\right) + m \tag{4.7}$$

4.6.3 Virtual Table

As shown in the literature [69, 59, 72], the presence of a virtual table can be helpful in presenting information. This prototype uses the virtual touch frame described in Section 4.6.1 to represent that information. While the table is useful for displaying information, it can obscure much of the to-scale model, especially if the user wants to look down. To make it less intrusive, the table

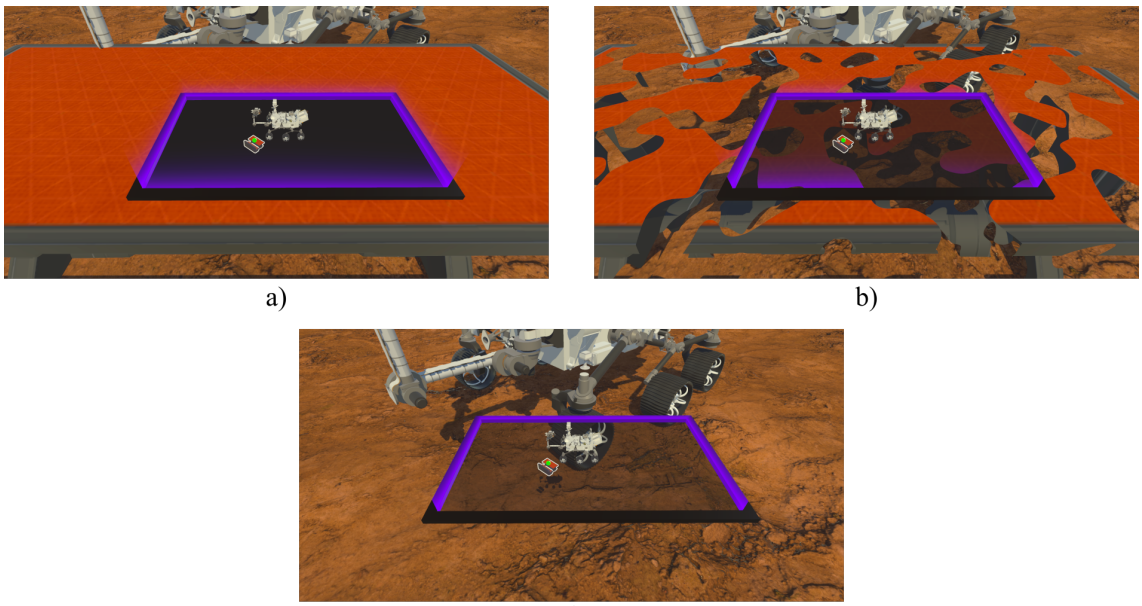
[11] https://math.stackexchange.com/a/2209953



a) b) c)

Figure 4.9: Graphs illustrating the functions used to modify the intensity of the limit illumination effect. Graph a) shows $e^{-\frac{d}{0.05}}$ where the horizontal axis represents distance d. Graph b) displays the function described in Equation 4.6, with the horizontal axis representing x. Graph c) depicts the function from Equation 4.6 with the horizontal axis representing distance d.

begins to fade and becomes invisible after 2 seconds of the touch surface not detecting any fingers, as shown in Figure 4.10.



a) b)

c)

Figure 4.10: The transition of the virtual table from: a) fully visible; b) half-visible; c) fully invisible.

This effect is achieved using a shader that takes the pixel's world space position and applies a simplex 3D noise function[12] to determine the pixel's alpha clip threshold. After 2 seconds of inactivity, the table's alpha value is reduced using a smooth animation curve. The alpha clip

[12] https://github.com/JimmyCushnie/Noisy-Nodes

threshold then determines whether a pixel is visible or invisible. Alpha blending was not used because it caused visual artifacts, making the table visible from behind itself.

User tables are also visible in the replica as miniatures, as shown in Figure 4.11 and described in Section 3.3. These miniatures feature an outline effect to help them stand out from the surrounding environment, using a free Unity package.[13] They also glow intermittently to draw attention, increasing the lightness of the table's color through an animation using a quadratic easing in-out function.[14] The miniatures display who is at the table by showing a user seated at it, as seen in image c) of Figure 4.11. These miniatures do not scale with the replica, keeping their size constant, similar to markers on a map.

They remain visible behind objects in the replica to help users quickly identify their and others' tables. This is achieved by rendering the table miniature in an additional render pass using the depth buffer to determine the appropriate material. If the table is behind an object, it appears slightly transparent and in a single color; otherwise, it uses the normal material.[15]

a) b) c)

Figure 4.11: The table miniature visible in the replica. Image (a) shows the table behind an object, image (b) shows the table within the replica, and image (c) shows two users at the table.

4.6.4 Points of Interest

As mentioned in Section 3.3, the points of interest reflect the appearance of their creators. Figure 4.12 demonstrates this: the first user's points of interest are green-striped spheres, while the second user's are purple checkerboard cubes. Each point of interest is marked with a number that rotates to face the camera and glows intermittently to draw attention, achieved using a Fresnel effect with a sine curve animation.

Similar to the miniature tables, points of interest are visible behind objects in the replica and do not scale with the replica. This is shown in image (a) of Figure 4.13, where a point of interest is visible behind a building in the replica with a muted color and slight transparency. However,

[13] https://assetstore.unity.com/packages/tools/particles-effects/quick-outline-115488

[14] https://assetstore.unity.com/packages/vfx/shaders/shader-graph-easing-193427

[15] https://docs.unity3d.com/Packages/com.unity.render-pipelines.universal@10.4/manual/renderer-features/how-to-custom-effect-render-objects.html

a) b)

Figure 4.12: Point of interest appearance based on the creator's appearance. Image (a) shows points of interest from the first user, and image (b) shows points of interest from the second user.

in the 3D model, the point of interest is not visible behind the building, as shown in image (b). This ensures that the points of interest do not distract or confuse users when looking at the replica. Points of interest in the 3D model scale with distance so users can see them from afar but do not become too large when close, preserving essential details.

Points of interest created by another user that are not yet acknowledged are marked with a vertical line, as shown in Figure 4.14. This line is capped with a symbol displaying the point of interest's identification number, resembling the point of interest's appearance to help users quickly identify unacknowledged points. The line and symbol always face the user using a vertical billboard effect. The line is also visible behind objects in the replica and scales with distance, as seen in image (c) of Figure 4.14, ensuring it can be seen from any angle. If the marker does not fit within the user's field of view, it flips upside down to remain visible, as shown in image (d) of Figure 4.14.

4.6.5 Balloon Selection

The balloon selection gesture is indicated by a set of dashed helper lines: one on the touch frame connecting the primary and secondary hands, and another vertical line from the primary hand to the balloon, using a vertical billboard effect, as shown in Figure 4.15. The balloon follows the appearance of the creator's points of interest. The helper lines and the balloon are visible behind objects in the replica with reduced opacity. When the secondary hand is inactive, the helper line on the touch frame loses its opacity, as shown in image (b) of Figure 4.15. If a segment of that line is behind an object and the secondary hand is inactive, that segment becomes invisible.

a) b)

Figure 4.13: Points of interest visibility. Image (a) shows a point of interest in the replica that is visible behind the building. Image (b) shows that the same point of interest is not visible in the 3D model behind the building.

The dashed lines are created using a shader that takes the positions of the hands, calculates the start and end positions of the dash segments using the percentages for each dash and gap, then draws the dash segments based on the calculations described in Equation 4.5. The dash segments at the ends are masked with a line from the start to the end position to ensure they have rounded ends.

The balloon and the vertical helper line are also visible in the 3D model, as shown in Figure 4.16, helping users understand the balloon's position in the real world. Both are not visible behind objects in the 3D model. The balloon scales with distance so users can see it from afar.

When the balloon intersects a point of interest, the point of interest becomes outlined, indicating it is selected, as shown in image (a) of Figure 4.17. When the balloon intersects a table, the table grows in size, as shown in image (b) of Figure 4.17.

4.7 Networking

The prototype utilizes Unity's Netcode for GameObjects library to handle networking, enabling synchronization of GameObjects and Scenes across clients. This simplifies the management of the prototype's networking components. The networking architecture follows a client-server topology,

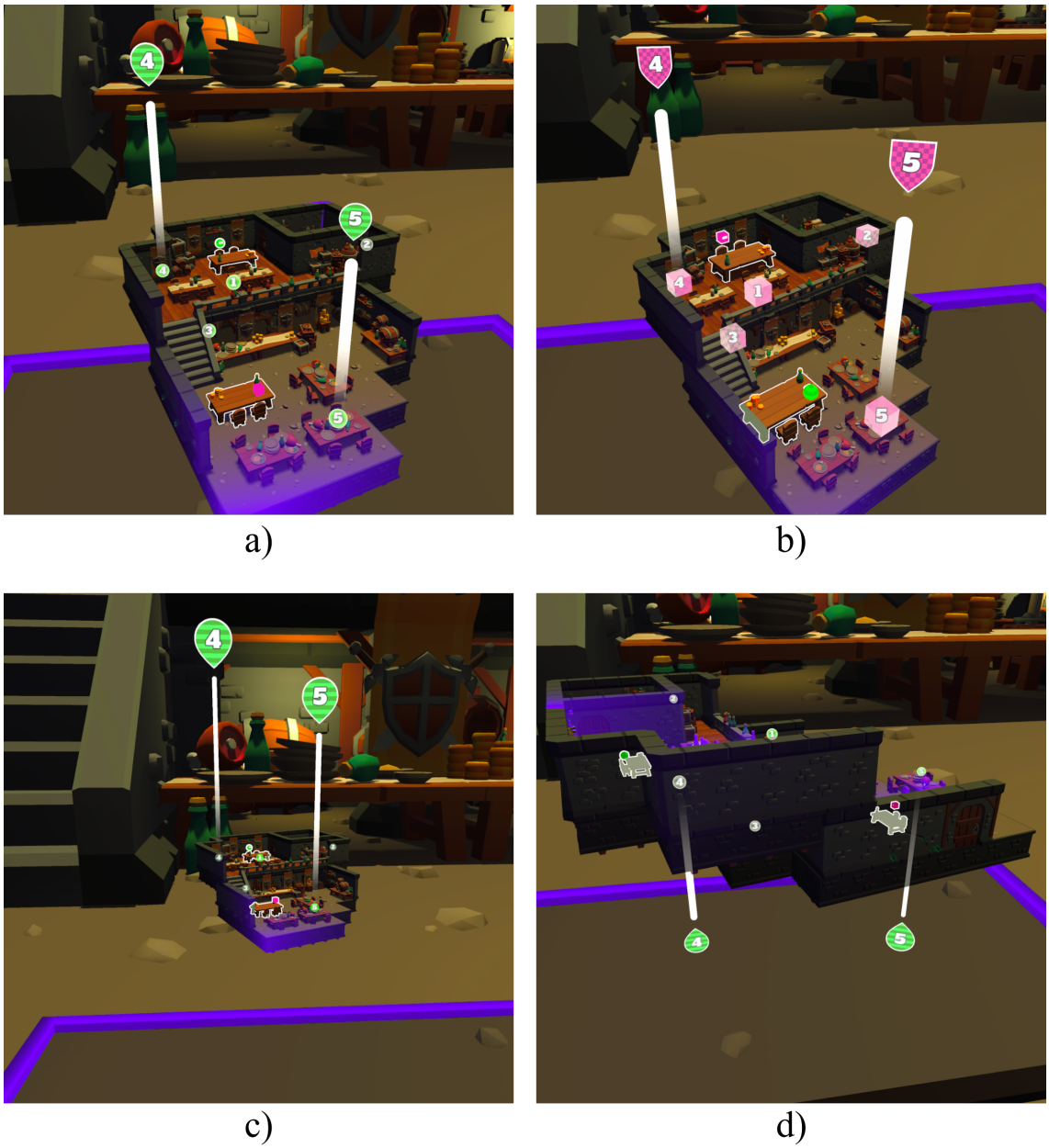
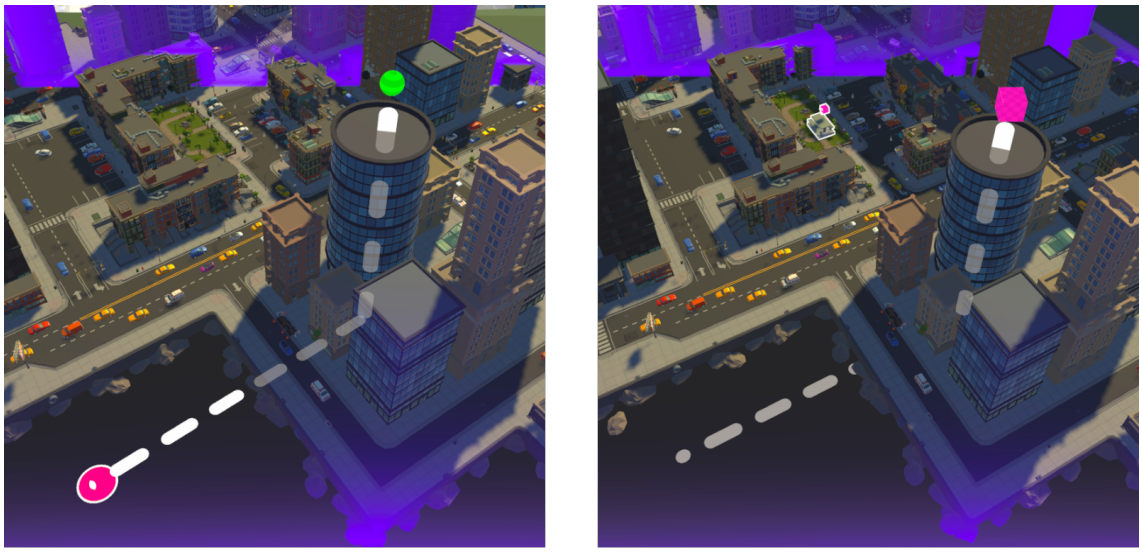


Figure 4.14: Point of interest markers. Image (a) shows markers for the first user, and image (b) shows markers for the second user. Image (c) demonstrates the scaling of the markers with distance. Image (d) displays the markers flipped upside down to ensure they are always visible.

illustrated in Figure 4.18. Specifically, it uses a client-hosted server named a host, where one client is also the server. This architecture was chosen for its simplicity in setup and management, and because it meets the prototype's requirements.

Each client has a user network object for each connected user and a table network object for each table in the scene. A NetworkObject is a GameObject that interacts with Netcode. Before an instantiated NetworkObject is synchronized across clients, it must be spawned. In the client-server architecture, only the server has the authority to spawn and despawn NetworkObjects. By default,



a) b)

Figure 4.15: Balloon selection helper lines. Image (a) shows the balloon for the first user, and image (b) shows the balloon for the second user with the secondary hand removed.

NetworkObjects are owned by the server and tied to its lifecycle. However, the user network objects are an exception. These are player NetworkObjects, that are automatically spawned by the server whenever a client connects, assigned to the client with ownership, and despawned when the client disconnects.[16]

NetworkObjects share data through network variables, synchronized with new clients when they join the server and with existing clients when the data changes. Network variables can have different permissions for reading and writing data. In a client-server architecture, the default setting is that the server has read-write permissions, while clients have read-only permissions.[17] Sections 4.7.1 and 4.7.2 describe the user and table network objects in detail.

The server manages the user and table network objects through the User Manager and Table Manager, respectively. These managers store data about connected users and tables, such as the association of user IDs to Netcode client IDs, the current point of interest ID, and the connected user IDs. They also handle the creation and destruction of network objects. Detailed functions of these managers are described in Sections 4.7.3 and 4.7.4.

The managers communicate with network objects through updates on network variables and remote procedure calls (RPCs). An RPC allows methods to be called on objects in another executable. In Netcode, RPCs execute methods on NetworkObjects across clients. A client can call an RPC on the server, and the server can call an RPC on a client. Clients can also call RPCs on other clients, though this passes through the server as a proxy.[18] Section 4.7.5 describes the sequence

[16] https://docs-multiplayer.unity3d.com/netcode/current/basics/networkobject/

[17] https://docs-multiplayer.unity3d.com/netcode/current/basics/networkvariable/

[18] https://docs-multiplayer.unity3d.com/netcode/current/advanced-topics/message-system/rpc/

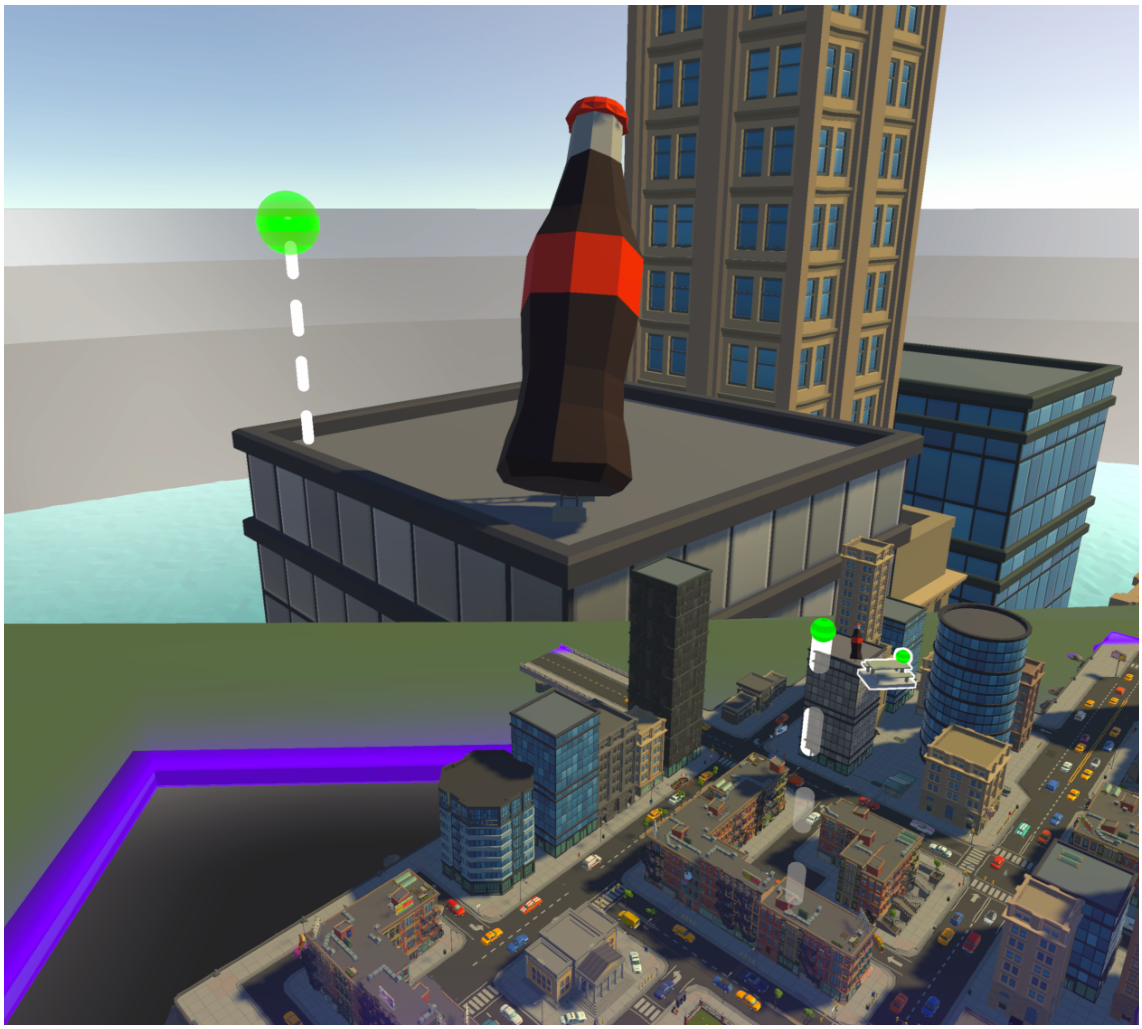


Figure 4.16: Balloon selection visible both in the replica and in the 3D model.

of events during the prototype's execution, including the RPCs used and other network-related events.

4.7.1 User Network Object

User network objects use network variables to store and synchronize data about the users. They have three network variables: user ID, a list of points of interest, and user transform data. The user ID identifies the user and determines their appearance. The list of points of interest contains the ID, position, and user ID of each point created by the user. The user transform data includes the user's position and rotation.

The user ID and points of interest can only be modified by the server, while the user transform data is updated by the client. This design choice simplifies the synchronization of the table tracking algorithm described in Section 4.5. The server assigns the user ID when the user joins, and updates the points of interest list as the user creates or removes points. The client updates the user transform data whenever the user moves.

a) b)

Figure 4.17: Balloon selection intersection with a point of interest in image a) and a table in image b).

When a user network object receives an update to the user ID it updates the user's appearance. If the client owns the user network object, it also updates the balloon's appearance of the balloon selection.

When a user creates a point of interest, a temporary point of interest is first created on the client, without an identification number, in both the replica and the 3D model. The temporary point of interest's position and user ID are sent to the user manager through an RPC. The user manager assigns an identification number to the point of interest and updates the appropriate user network object's points of interest list. Whenever this list is updated, the user network objects determine the type of update. If the client is the owner of the user network object and a point has been added, the temporary point of interest is updated with the identification number. If the client is not the owner, the replica controller is updated with the new point of interest, and marks it as unacknowledged. If a point of interest is removed, the user network object removes the point of interest from the replica controller. Late-joining users acquire all the spawned user network objects and update the WIM with the points of interest.

4.7.2 Table Network Object

Table network objects have three network variables: transform data, the user ID seated at the first seat, and the user ID seated at the second seat. The transform data includes the table's position and rotation. Only the server can modify these variables, managed by the table manager. When a table network object is spawned, clients create a table replica in the WIM with the defined transform and seat data. When the table is despawned, the table replica is destroyed. Late-joining users acquire all the spawned table network objects and create table replicas accordingly.

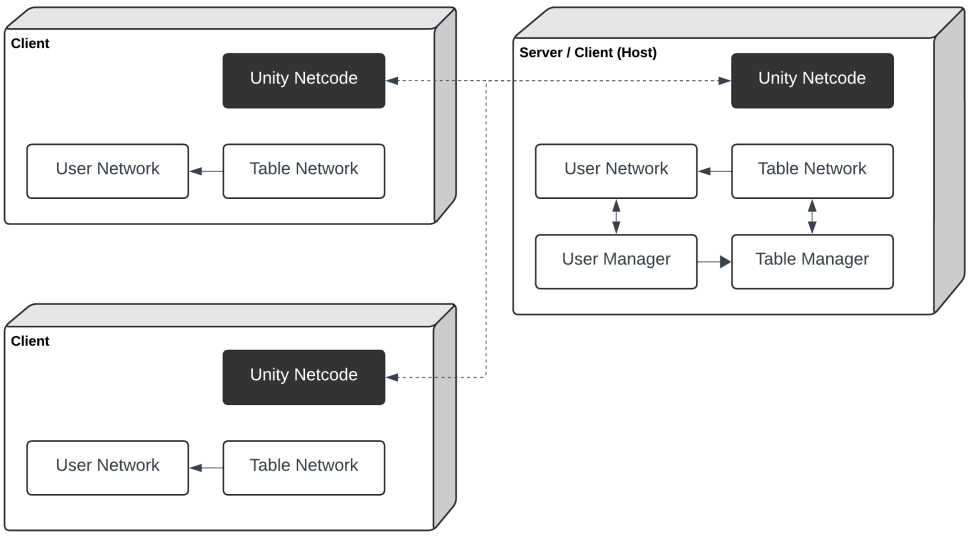


Figure 4.18: Networking architecture, using a client-server topology.

When the server updates the table network object's transform data, clients update the local table instance and the table replica's transform data. If a client is seated at the table, the client moves to the new position using the tracking data, as described in Section 4.5. When the server updates the table network object's seat data, clients update the table replica to reflect who is seated at each seat.

4.7.3 User Manager

The user manager is responsible for tracking and communicating with connected users. It maintains the association of user IDs to Netcode client IDs, the current point of interest ID, and the list of connected user IDs. When a user connects, Netcode triggers the `OnClientConnected` event on the user manager, which assigns a user ID and updates the list of connected user IDs. For this prototype, only two users can be connected, so the user manager has a list of available user IDs, 0 and 1, which it assigns to new users. After updating the user's ID network variable, the user manager requests the table manager to add the player to an available table, as described in Section 4.7.4. Once a table is assigned, the user manager sends an RPC `MoveUserToTableClientRpc` to the user network object to move the user to the table, as shown in Figures 4.19 and 4.21. When a user disconnects, the user manager removes the user ID from the list of connected user IDs and adds it back to the list of available user IDs. It then requests the table manager to remove the player from their table, as depicted in Figure 4.23. The spawning and despawning of user network objects are handled automatically by Netcode.

When a user creates a point of interest, the user's network object sends an `CreatePointOfInterestRpc` to the user manager. The user manager increments the current point of interest ID and updates the user network object's points of interest list. Because this list is a network variable, it is synchronized across all clients, and the user network object handles the update accordingly,

as described in Section 4.7.1. When a user removes a point of interest, the user network object sends an RPC `RemovePointOfInterestRpc` to the user manager. The user manager then removes the point of interest from the network object's list.

When a user teleports or joins a table, the user network object sends an `MoveUserToPositionRpc` or `MoveUserToTableRpc` to the user manager. The user manager then instructs the table manager to move the user to the new position or table, as outlined in Section 4.7.4. Afterward, the user manager sends an RPC `MoveUserToTableClientRpc` to update the user network object's position, as illustrated in Figures 4.20 and 4.22.

4.7.4 Table Manager

The table manager oversees the table network objects and assigns users to tables. Its primary functions include assigning newly connected users to available tables, managing user teleportation, handling users joining tables, and removing users from tables upon disconnection.

When assigning a user to a table, the table manager first checks for available seats at existing tables. If a seat is available, the user is assigned to it. If no seats are available, the table manager creates a new table and assigns the user to the first seat. These scenarios are depicted in Figures 4.19 and 4.21. When a user disconnects, the table manager removes them from their table. If the table becomes empty, it is despawned. This process is illustrated in Figure 4.23.

When handling user teleportation, the table manager first checks if the user is alone at their current table. If they are, it moves the table to the new position. If the user is not alone, a new table is created at the new position, and the user is assigned to the first seat. These scenarios are depicted in Figure 4.20.

When a user joins a table, the table manager removes them from their original table. If the original table becomes empty, it is despawned. The user is then assigned to the first available seat on the new table. This process is shown in Figure 4.22.

4.7.5 Sequence of Events

This section outlines a typical scenario that demonstrates the key interactions between users, tables, and the network. The diagrams illustrated only show the interaction between one client and the server at a time, from the perspective of the owner of the depicted user network object. Whenever a network variable is updated, the change is synchronized across all clients.

The scenario begins with a user creating a server, thereby becoming its host. In this case, RPCs are local and executed instantly since the server also functions as a client. The client then connects to the server, as shown in Figure 4.19. The user manager assigns the user the ID 0 by updating the user's network variable and the list of connected user IDs. Upon updating this ID, the user's model is updated to the correct appearance. The user manager then requests the table manager to assign the user to an available table. Since no other users are connected, the table manager creates a new table, Table Network 1, and assigns the user to the first seat. When the table is spawned, it

requests the local client to create the table replica on the WIM. The user manager then sends an RPC to the user network object, moving the user to the table.

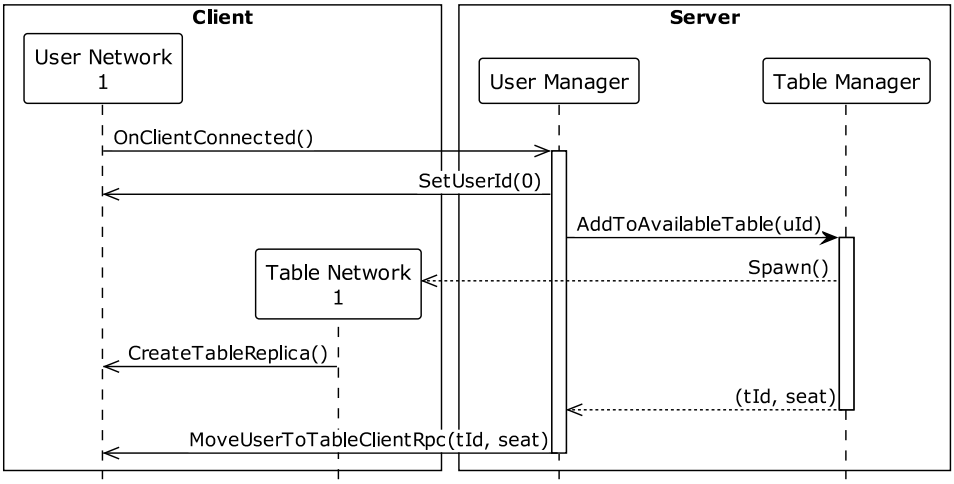


Figure 4.19: Sequence diagram of a user connecting and creating a table.

Next, the user creates a point of interest. Initially, a temporary point of interest is created on the client, followed by sending a `CreatePointOfInterestRpc` to the user manager. The user manager increments the current point of interest ID and updates the user network's points of interest list. When the user network object receives this update, it assigns the identification number to the temporary point of interest, as it is owned by the client.

The user then teleports to a new position. The user network object sends a `MoveUserToPositionRpc` to the user manager, which instructs the table manager to move the user to the new location. Since the user is alone at the table, the table is moved to the new position, updating its network variables accordingly. The table network object, upon receiving these updates, adjusts the table replica on the client, and moves the user to the new position.

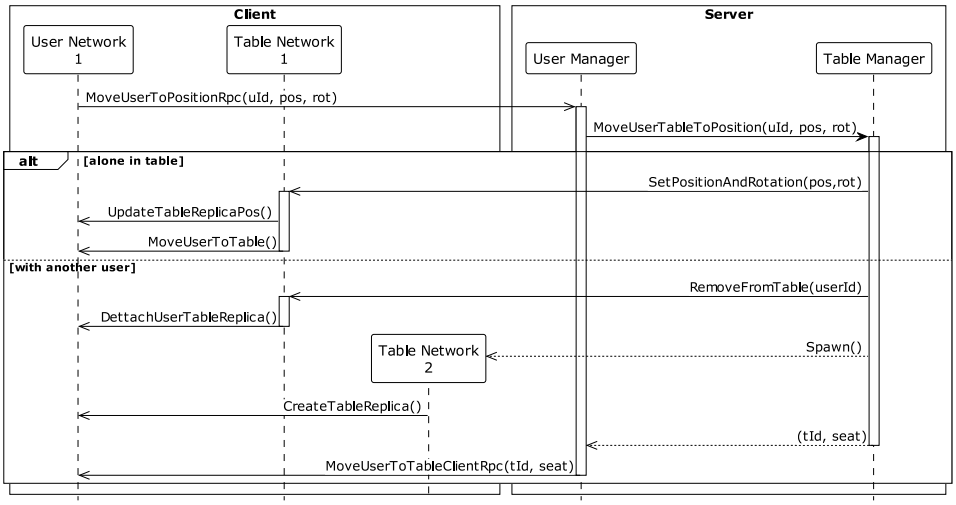


Figure 4.20: Sequence diagram of a user teleporting.

Later, another user connects to the server, as illustrated in Figure 4.21. The user manager assigns this user the ID 1 and updates the list of connected user IDs. Consequently, the user network object is updated with the new user ID and the corresponding appearance. The table manager assigns the new user to the second seat of Table Network 1, updating its seat network variables. The table network object then updates the table replica on each client, displaying the new user in the second seat. Following this, the user manager sends an RPC to the user network object to position the user at the table.

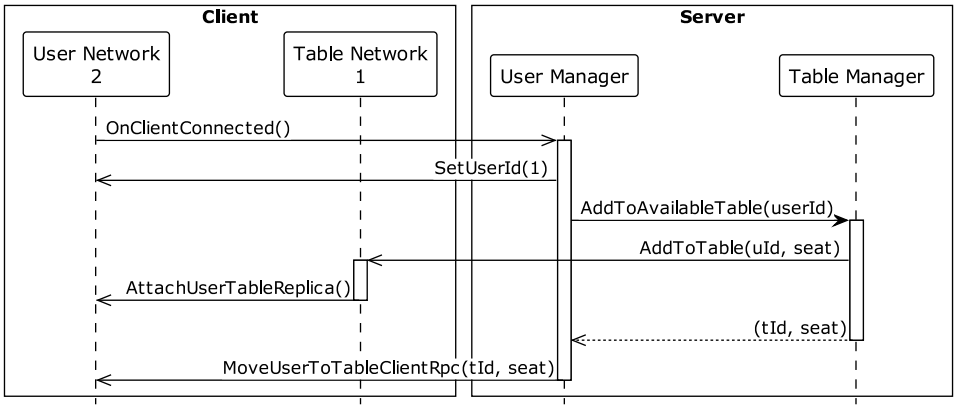


Figure 4.21: Sequence diagram of a user connecting and joining an existing table.

Upon joining, the second user gathers information about the first user's points of interest by loading the point of interest data from the first user's network object, marking them as unacknowledged. Since acknowledgment data is local, no network communication is required. The second user would also load table data from other table network objects if any additional tables existed.

Next, the first user moves to a new position, as depicted in Figure 4.20. The user network object sends a `MoveUserToPositionRpc` to the user manager, which instructs the table manager to relocate the user. Because the user is not alone at the table this time, the table manager removes the user from Table Network 1 and spawns a new table, Table Network 2, assigning the user to the first seat. The new table network object updates the table replica on each client to reflect the user's new position, while the old table network object removes the user from the replica. Finally, the user manager sends an RPC to the user network object, moving the user to the new table.

Following this, the second user joins the second table, as shown in Figure 4.22. Their network object sends a `MoveUserToTableRpc` to the user manager, which instructs the table manager to move the user to the second table. The table manager removes the user from Table Network 1 and assigns them to the second seat of Table Network 2. Since Table Network 1 is now empty, it is despawned. The table network objects update the table replicas on each client accordingly. The user manager then sends an RPC to the user network object, positioning the user at the new table.

Finally, the first user disconnects, as depicted in Figure 4.23. The user manager removes the user ID from the list of connected user IDs and returns it to the list of available user IDs. It then instructs the table manager to remove the user from their table. The table network object

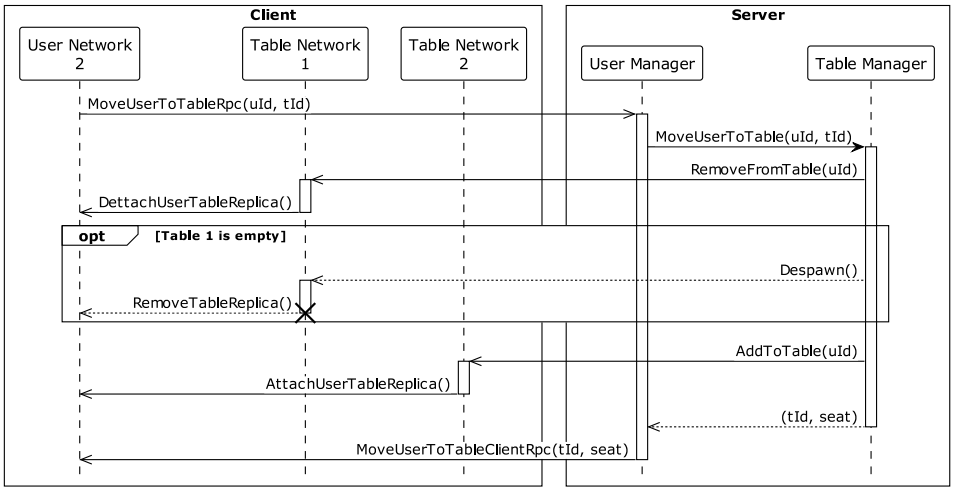


Figure 4.22: Sequence diagram of a user joining a table.

updates the table replica on each client to reflect the user's removal. The user network object is automatically despawned by Netcode.

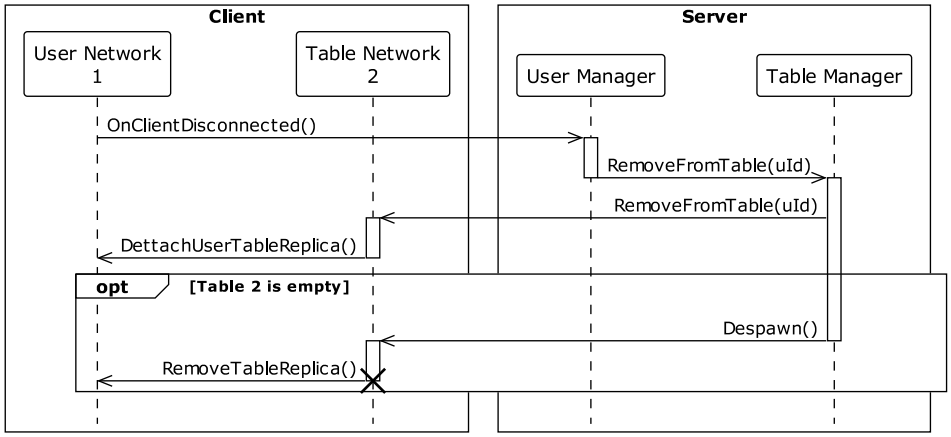


Figure 4.23: Sequence diagram of a user disconnecting.

4.8 Summary

This chapter details the implementation of a prototype for Replico. It starts with an overview of the system architecture and the hardware and software environment utilized, including Unity, Netcode for GameObjects, OpenXR, two HTC Vive Pro 2 headsets, and two multi-touch surfaces.

The chapter continues by describing the state machine used to manage the prototype's states. These states range from the initial state, to `TransformReplicaState` where users can manipulate the replica, and `BalloonSelectionInitialState` where users can create, delete, and acknowledge points of interest, teleport, and join tables.

Furthermore, the chapter details the calculations required to use the multi-touch input to transform the replica. Then, it details the gesture detection mechanisms used, including the utilization

of the K-means clustering algorithm implemented via the ML.NET library for hand detection. Additionally, it outlines the use of VR controllers to track the user's interactions with virtual tables.

The chapter describes the visual indicators implemented in the prototype. These include visual cues such as finger trails on the virtual touch frame, glow effects signaling gesture detection, illumination effects marking the touch frame's boundaries on the replica, and the visual representation of tables, table replicas, points of interest, and their corresponding visual appearances.

The networking implementation is explained, describing the architecture, user and table network objects, user and table managers, network variables, and Remote Procedure Calls (RPCs). The section ends with an outline of a typical sequence of events that unfold during the prototype's operation.

Chapter 5

Evaluation

A user study was conducted to evaluate Replico's efficiency, effectiveness, and user-friendliness. The main goal was determining how easily users could communicate points of interest within the virtual environment using Replico. Secondary goals included assessing ease of use and overall user experience. The study aimed to answer four key research questions:

- **RQ1**: How efficiently can users create a point of interest on a given object?
- **RQ2**: How effectively does Replico notify users when a point of interest is created?
- **RQ3**: How useful is the world-in-miniature metaphor for communicating points of interest? How useful is the representation of user locations on the replica for understanding intent?
- **RQ4**: How user-friendly is Replico, and how much physical effort is required to use it?

5.1 Setup

The user study was conducted at FEUP, in the GIG laboratory in room I220. The setup included two VR-ready computers connected to a local network. One computer was equipped with an Intel Core i9-13900KF CPU, an NVIDIA GeForce RTX 4090 GPU, and 32 GB of RAM, while the other had an Intel Core i9-11900F CPU, an NVIDIA GeForce RTX 3080 GPU, and 32 GB of RAM. Each computer had an HTC Vive Pro 2 headset and a VR controller for table tracking. Two touch surfaces – a 32-inch infrared frame and a 47-inch capacitive Displax Skin Ultra touchscreen – were placed on opposite tables within the central VR play space. Participants, in pairs, were seated in front of each touch surface with their backs facing each other, as shown in Figure 5.1.

One computer served as the host, while the other connected as a client. The setup process involved a starting screen where the moderator could select the IP address of the host computer. The roles of each computer did not change throughout the study to simplify the setup process and avoid confusion.

a) b)

Figure 5.1: Setup for the user study. In image (a) one participant is seated in front of the Displax Skin Ultra. In image (b) the participant is seated in front of the infrared touch frame.

5.2 Methodology

The study was conducted in pairs, with initial tasks performed individually and later tasks performed collaboratively. Each session consisted of three main parts: an introduction, a training session, and the main tasks. Each session lasted approximately 60 minutes. After each session, participants received a chocolate bar as a token of appreciation.

Before the study began, participants were introduced briefly to the study's purpose and the Replico system. They then completed a consent form and a profiling questionnaire. Following this, they watched a video presentation explaining Replico's features, usage, and the tasks they would perform.

During the training session, participants familiarized themselves with the system by experimenting with all of Replico's features. Once comfortable with the solo interactions, a set of points of interest and a simulated player were added to the environment, allowing users to practice acknowledging points of interest and joining the other user's table. After becoming comfortable with these interactions, they proceeded to the main tasks.

The main tasks were performed using two different 3D models: a city and the Perseverance rover, described in Section 5.2.1. The order in which the models were used alternated between pairs to avoid bias. These tasks aimed to assess the efficiency and effectiveness of Replico's features, as described in Section 5.2.2. Metrics for each task were collected as detailed in Section 5.2.3. Participants completed a questionnaire on the tasks they performed between each test scenario, described in Section 5.2.4.

5.2.1 Test Scenarios

Three scenarios were used during the study, two for the main tasks, as shown in Figure 5.2. The first scenario, used for training, is a small dungeon tavern built with the free version of the KayKit Dungeon Remastered Pack from itch.io[1] and the Modular Asset Staging Tool (MAST) for Unity[2]. The second scenario is a city from Synty's POLYGON City Pack[3], obtained through Unity's student plan. The third scenario features the Perseverance rover, obtained from NASA's 3D model repository[4], with the surrounding environment created using Unity's terrain tools and a tinted sand texture from Polyhaven[5].

a) b) c)

Figure 5.2: The three test scenarios used in the user study. From left to right: the dungeon tavern, the city, and the Perseverance rover.

Two scenarios were selected for the main tasks to evaluate how well the approach works with different 3D models. The city model was chosen for its large, complex structure, allowing users to immerse themselves within it. The Perseverance rover model was selected for its smaller size but sufficient detail, enabling users to view the entire model at once without being part of it. The dungeon tavern was used for practice, providing a small, enclosed environment distinct from the main task scenarios.

In all scenarios, the virtual world includes surrounding environment elements to provide context and a sense of scale. For example, the city is bordered by an ocean, the rover by the Martian surface, and the dungeon tavern by some of its walls. The WIM does not replicate these environmental elements, as they are not part of the primary scenario.

[1] https://kaylousberg.itch.io/kaykit-dungeon-remastered
[2] https://fertile-soil-productions.itch.io/mast
[3] https://assetstore.unity.com/packages/3d/environments/urban/polygon-city-low-poly-3d-art-by-synty-95214
[4] https://nasa3d.arc.nasa.gov/detail/perseverance-glb
[5] https://polyhaven.com/a/sand_01

5.2.2 Tasks

Participants performed five main tasks during the study, using each of the two main scenarios. The first three tasks were done individually, while the last two were collaborative. Each task evaluated different aspects of Replico's features and user experience. The tasks are as follows:

- **Task 1**: Create a point of interest on six different predefined objects in the scene. This task evaluates how efficiently users can create points of interest.
- **Task 2**: Acknowledge five out of twelve predefined points of interest created by a simulated user. This task assesses how effectively Replico notifies users of new points of interest.
- **Task 3**: Teleport to four different predefined zones and orient themselves to face a specific object. This task measures how well users can navigate the environment using Replico.
- **Task 4 & 5**: One user must show another user a specific object in the scene without verbal communication. The other user then verbally confirms the object they believe the first user is referring to. Once confirmed, the roles are reversed, and the process is repeated with a different object. This task evaluates how well users can communicate points of interest using Replico

Between each task, the environment was reset to its initial state. Any points of interest created during the previous task were removed, and participants were placed back at their starting positions. This was done to ensure that each task was performed under the same conditions for all participants.

In the first task, the predefined objects were chosen with different sizes and at various heights, visible in Figure 5.3. The goal was to evaluate how efficiently users could create points of interest, not to test their ability to identify objects. To draw attention to them, these objects glowed with an expanding and contracting effect in the WIM. They increased in size and changed their outline color from white to green when the balloon intersected with them. Creating a point of interest during this state allowed users to progress to the next object. The objects appeared one at a time, immediately after the user created a point of interest on the previous object. The order in which the objects appeared was consistent for all participants.

In the second task, the predefined points of interest were placed around the environment simultaneously, as shown in Figure 5.4. The goal was to evaluate how effectively Replico notifies users of new points of interest and helps distinguish acknowledged from unacknowledged points of interest. The task was complete when the user acknowledged the five unacknowledged points of interest. Participants could acknowledge the points of interest in any order.

In the third task, participants navigated through four predefined zones in the environment, as shown in Figure 5.5. The goal was to assess how well users could navigate using Replico. Participants were instructed to teleport to each zone and orient themselves to face a specific object. Zones changed color from white to green when the balloon intersected with them, and the object's glow effect also changed from white to green when the user was correctly oriented. The moderator

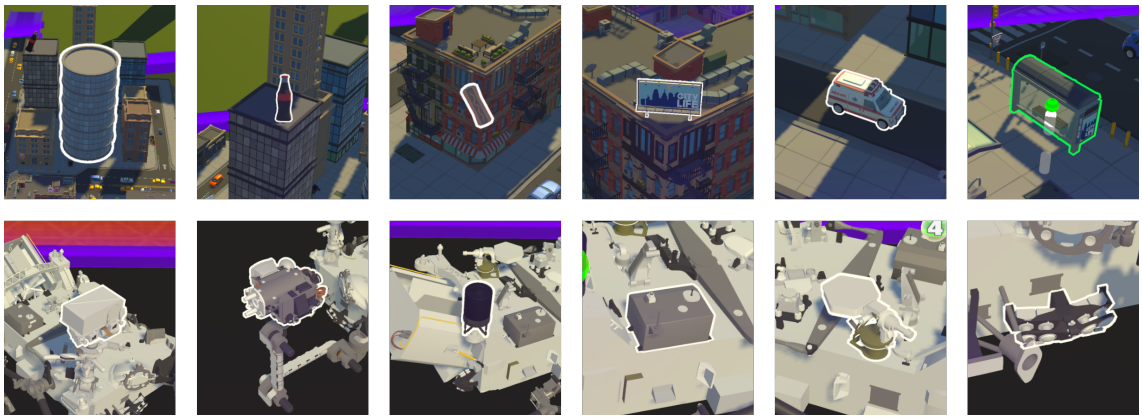


Figure 5.3: The six predefined objects used in Task 1 for each scenario. They are ordered from left to right, with the top row showing the objects in the city scenario while the bottom row shows the objects in the Perseverance rover scenario. The top-right image shows the green outline effect when the balloon intersects with the object.

a) b)

Figure 5.4: The created points of interest in Task 2. Image (a) shows the points of interest in the city scenario, while image (b) shows the points of interest in the Perseverance rover scenario.

advanced the task manually, allowing participants to take their time exploring the environment. The sequence of zones was the same for all participants.

In the fourth and fifth tasks, participants communicated objects of interest without verbal communication. Figure 5.6 shows the objects used in these tasks. The goal was to assess how effectively users could communicate using Replico. The selected objects were small and difficult to identify from a distance, encouraging the use of the WIM, points of interest, and teleportation for effective communication. For the Perseverance rover model, since most people are unfamiliar with its components, the objects chosen were a small hidden star and a small hidden heart.

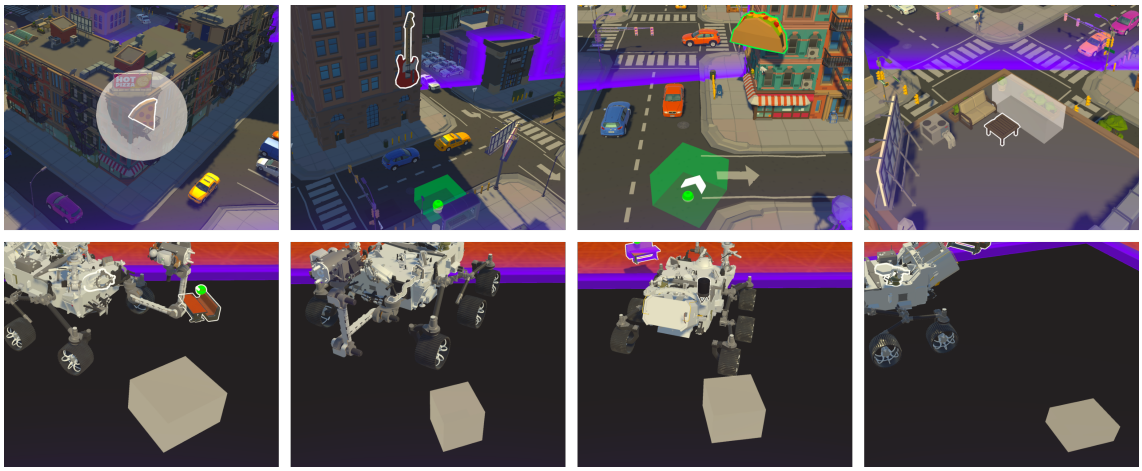


Figure 5.5: The four predefined zones used in Task 3 for each scenario. They are ordered from left to right, with the top row showing the zones in the city scenario while the bottom row shows the zones in the Perseverance rover scenario.

5.2.3 Metrics

During each task, several metrics were collected to evaluate the efficiency and effectiveness of Replico's features. These metrics, recorded automatically by the system, included the time in seconds taken to complete the task, the number of successful task steps, time spent transforming the replica, time spent in vertical transformation, and time spent in balloon selection. Additionally, the system tracked the number of times the transform gesture was detected (on entering `TransformReplicaState`), the number of times vertical transform was detected (on entering `TransformReplicaVerticalState`), and the number of times the balloon selection gesture was detected (on entering `BalloonSelectionInitialState`). It also recorded the number of points of interest created, deleted, and acknowledged, the number of teleportations, the number of table joins, the number of touches on the touch surface, the cumulative sum of finger movement in pixels, the cumulative sum of head rotation in angles, cumulative head translation in meters, cumulative sum of replica rotation in angles, cumulative sum of replica translation in meters, and cumulative sum of replica scaling in meters.

The time taken to complete each task was measured from the start to the end of the task. In Task 3, while the participant waited for the moderator to advance to the next zone, all metrics were paused. The number of successful task steps was the number of steps completed correctly by the user. For example, in Task 1, a successful step was creating a point of interest on a predefined object. In Task 2, a successful step was acknowledging a point of interest. In Task 3, a successful step was teleporting to a zone and orienting oneself to face a specific object. In Tasks 4 and 5, it referred to the number of guesses until the correct object was identified.

Additionally, data was collected and timestamped for each frame that showed a change. This included finger data, replica transform data, and head transform data.

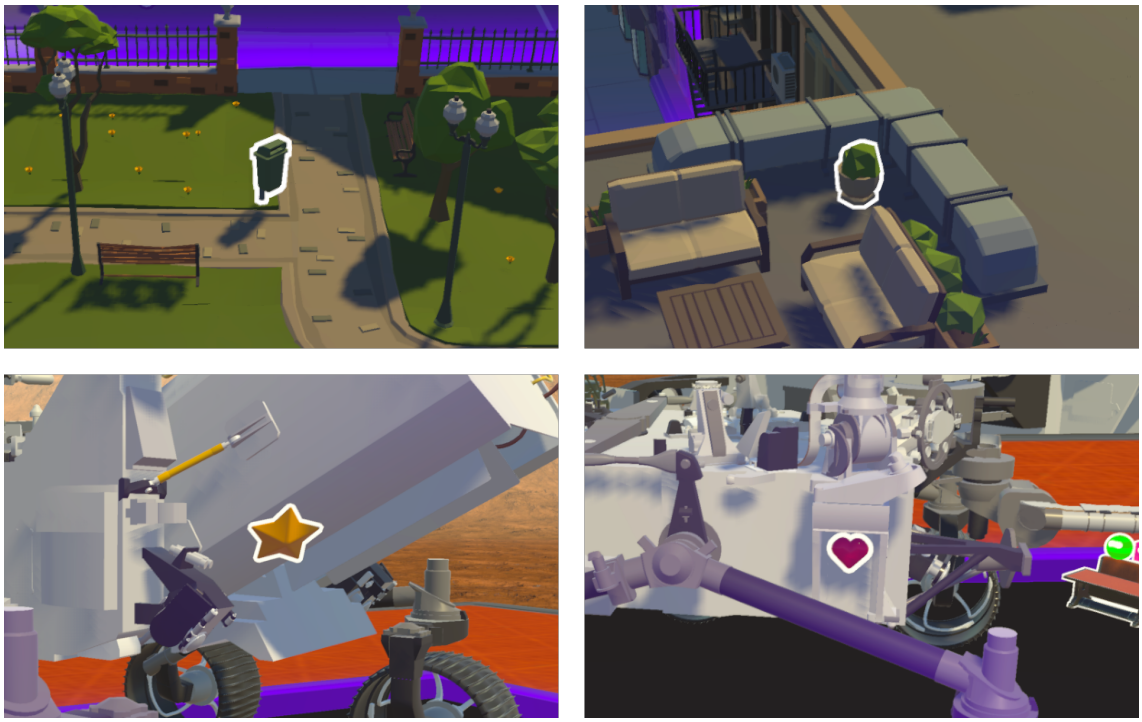


Figure 5.6: The objects used in Tasks 4 and 5 for each scenario. The top row shows the objects in the city scenario, while the bottom row shows the objects in the Perseverance rover scenario.

5.2.4 Qualitative Data

After each test scenario, participants completed a questionnaire to provide qualitative feedback on their experience. The questionnaire included six questions from the NASA Task Load Index (NASA-TLX) [24] for each task, merging tasks 4 and 5 into a single task, along with three additional questions at the end of the questionnaire, each rated on a 5-point Likert scale. The NASA-TLX questions were:

- How mentally demanding was the task?
- How physically demanding was the task?
- How hurried or rushed was the pace of the task?
- How successful were you in accomplishing what you were asked to do?
- How hard did you have to work to accomplish your level of performance?
- How insecure, discouraged, irritated, stressed, and annoyed were you?

The final three questions were:

- How nauseous did you feel during the task?
- How useful was the world-in-miniature metaphor for communicating points of interest?

- How useful was the representation of user locations on the replica for understanding intent?

Participants were also encouraged to provide additional comments or suggestions for improvement. The aim was to gather feedback on the user experience and identify areas for improvement in the system.

5.3 Participants

A total of 20 participants, forming ten pairs, took part in the study. Among them, eleven were male, and nine were female. Most participants, 19, were aged between 21 and 30 years old, with one participant aged between 31 and 40. Nineteen participants were right-handed, and one was left-handed. Seventeen participants were students, of whom six also worked, while three were exclusively workers. Fifteen participants had completed a bachelor's degree, two had a master's degree, and three had a high school diploma. Regarding VR experience, 11 had used VR once before, three had used it in the past year, one used it frequently, and five had never used VR before.

5.4 Results

The results of the user study are presented in this section. Two samples were collected for each task metric, one for each test scenario. Appropriate statistical tests were performed to determine the significance of the results, with the significance level set at the conventional alpha value of 5%.

5.4.1 Metrics

Using descriptive data analysis, outliers were identified and removed from the dataset. The Shapiro-Wilk test [54] was used to determine if the data followed a normal distribution. Since the analysis involved testing two related samples, either the paired t-test or the Wilcoxon signed-rank test [66] was used to determine whether the samples' differences were statistically significant. The paired t-test was used for normally distributed data, while the Wilcoxon signed-rank test was used for non-normally distributed data.

For the metrics of task 4 and task 5, except for task completion time (which was the same for both participants), the metrics were split into two separate tasks: one for the seeker (TaskSeek) and one for the shower (TaskShow). A two-way repeated measures ANOVA with Bonferroni correction was used to determine if there were significant differences between two factors: the object (task 4 or task 5) and the scenario (city or rover). If the sphericity assumption was violated according to Mauchly's test of sphericity, the Greenhouse-Geisser correction was applied. If the interaction effect was significant, a simple main effects analysis was performed to identify the specific differences. The studentized residuals were checked for outliers, and the data was checked for normality to ensure the assumptions of the ANOVA were met.

Five metrics were selected for analysis: task completion time, active time, finger movement, replica translation, and head translation. The results for each metric are presented in the following sections.

5.4.1.1 Task Completion Time

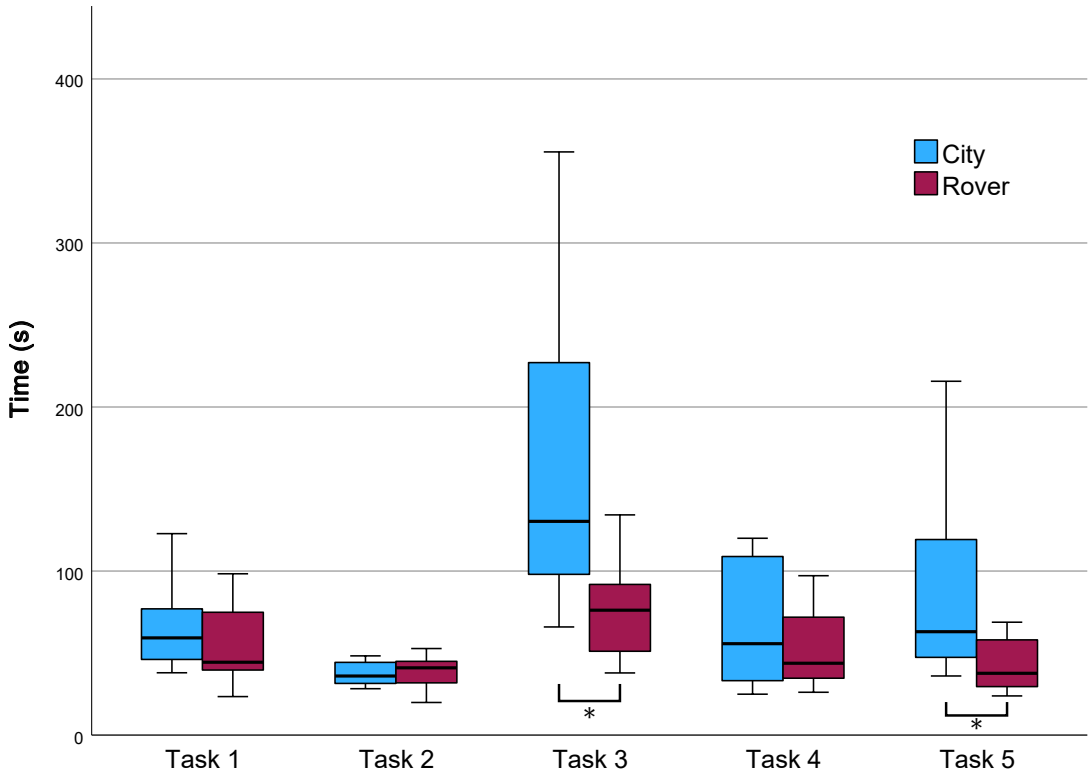


Figure 5.7: Box-plot of the time taken to complete each task for each scenario. The symbol $*$ indicates a significant difference between the city and rover scenarios.

Task completion time, recorded in seconds, measures how long it took participants to complete each task. The results for each task are shown in Figure 5.11. For the first task, there was no significant difference in completion time between the city and rover scenarios ($t(14) = 1.440,\ p = 0.172$). Similarly, the second task showed no significant difference in completion time between the two scenarios ($t(10) = 0.134,\ p = 0.896$). However, for the third task, there was a significant difference in completion time between the two scenarios ($Z = -3.248,\ p = 0.001$), with participants taking longer in the city scenario. The fourth task also showed no significant difference in completion time between the two scenarios ($Z = -1.415,\ p = 0.157$). The fifth task had a significant difference in completion time between the two scenarios ($Z = -3.154,\ p = 0.002$), with participants taking longer in the city scenario.

Overall, task completion times were similar between the city and rover scenarios, except for tasks 3 and 5. Task 3 took longer in the city scenario, likely due to the difficulty of locating the teleportation zones. These zones were more spread out and smaller, making balloon selection

harder and requiring participants to be more precise. Although participants could scale the replica to aid selection, they generally did not. Task 5 also took longer in the city scenario because the shower object was small and hidden among other objects, making it harder for the seeker to find.

In contrast, in the rover scenario, the objects were easier to locate, with some participants finding them by accident during previous tasks. Additionally, points of interest in the WIM are visible behind objects, which can completely obscure small objects. Teleportation would help in this case, as points of interest in the to-scale model aren't visible behind objects and become smaller as the user gets closer to them. However, participants generally did not use it.

5.4.1.2 Active Time

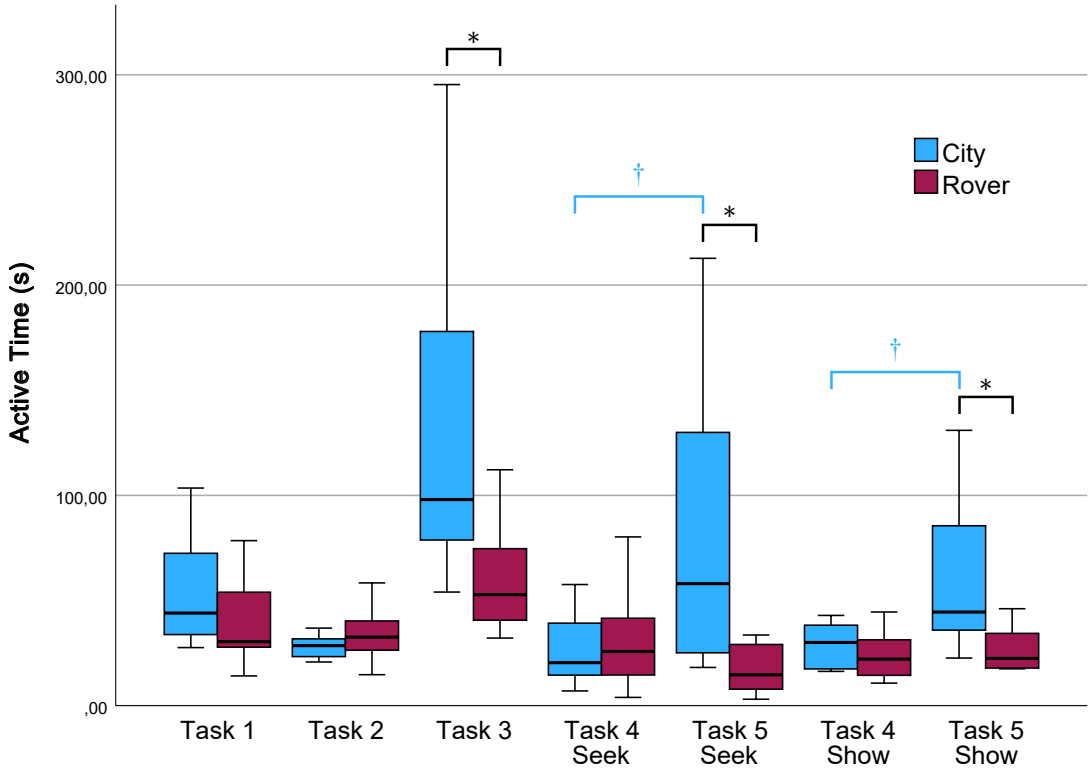


Figure 5.8: Box plot showing the active time for each task in both scenarios. The asterisk ($*$) indicates a significant difference between the city and rover scenarios. The dagger (†) shows significant differences between objects in TaskSeek and TaskShow.

Active time, measured in seconds, represents the duration participants spent actively touching the touch surface. The results for each task are shown in Figure 5.8. In the first task, there was no significant difference in active time between the city and rover scenarios ($t(12) = -1.913,\ p = 0.056$). The second task also showed no significant difference in active time between the two scenarios ($t(12) = -1.431,\ p = 0.178$). However, for the third task, participants spent significantly more active time in the city scenario ($Z = -3.051,\ p = 0.002$).

A repeated measures ANOVA with Greenhouse-Geisser correction revealed no significant difference in active time between the different scenarios in TaskSeek ($F(1,9) = 4.680,\ p = 0.059$), and no significant difference between the different objects ($F(1,9) = 2.468,\ p = 0.151$). However, the scenarios and objects had a significant interaction effect ($F(1,9) = 12.630,\ p = 0.006$). Simple main effects analysis showed that, in the city scenario, there was a significant difference in active time between the different objects ($p = 0.035$), with the second object having a significantly higher active time. In the rover scenario, there was no significant difference in active time between the different objects ($p = 0.075$). For the first object, there was no significant difference in active time between the city and rover scenarios ($p = 0.767$). However, for the second object, there was a significant difference in active time between the city and rover scenarios ($p = 0.017$).

A repeated measures ANOVA with Greenhouse-Geisser correction also determined no significant difference in active time between the different scenarios in TaskShow ($F(1,6) = 4.472,\ p = 0.079$), and no significant difference between the different objects ($F(1,6) = 3.217,\ p = 0.123$). However, the scenarios and objects had a significant interaction effect ($F(1,6) = 16.370,\ p = 0.007$). Simple main effects analysis revealed that, in the city scenario, there was a significant difference in active time between the different objects ($p = 0.032$), with the second object having a significantly higher active time. In the rover scenario, there was no significant difference in active time between the different objects ($p = 0.429$). For the first object, there was no significant difference in active time between the city and rover scenarios ($p = 0.935$). However, for the second object, there was a significant difference in active time between the city and rover scenarios ($p = 0.025$).

Compared to task completion time, active time was mainly comparable between the city and rover scenarios, except for Task 3, Task 5 Seek, and Task 5 Show, where participants spent more active time in the city scenario. This analysis highlights the increased difficulty of the second object compared to the first object in the collaborative tasks for the city scenario. Both objects were equally challenging to find in the rover scenario, with participants spending similar active time on both.

5.4.1.3 Finger Movement

Finger movement, measured in meters, represents the cumulative distance traveled by participants' fingers on the touch surface. Each touch surface's pixels per inch (PPI) was calculated using its dimensions and screen resolution to convert pixel data to meters. The pixel finger movement data was then divided by the PPI to convert it to inches, and this value was subsequently converted to meters. The results for each task are shown in Figure 5.9.

In the first task, there was a significant difference in finger movement between the city and rover scenarios ($t(14) = 2.976,\ p = 0.010$), with participants moving their fingers more in the city scenario. The second task showed no significant difference in finger movement between the two scenarios ($t(15) = 1.030,\ p = 0.320$). However, for the third task, there was a significant difference in finger movement between the city and rover scenarios ($Z = -3.285,\ p = 0.001$), with participants again moving their fingers more in the city scenario.

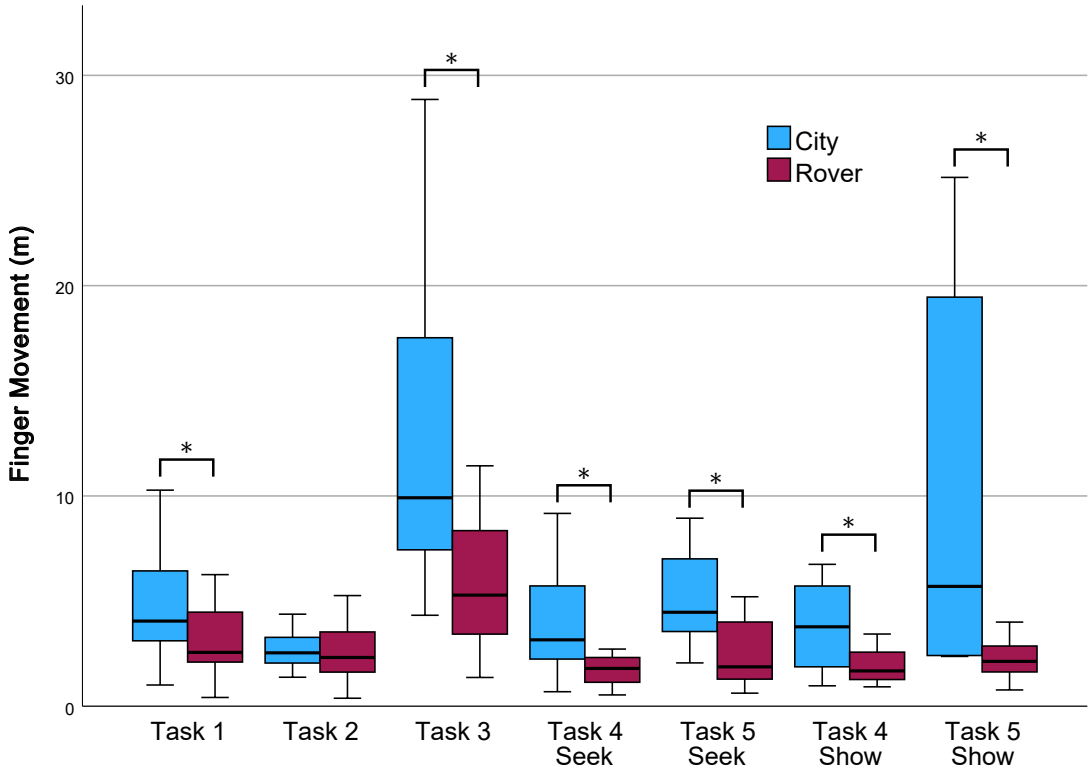


Figure 5.9: Box plot showing the cumulative sum of finger movement in meters for each task in both scenarios. The asterisk ($*$) indicates a significant difference between the city and rover scenarios.

A repeated measures ANOVA with Greenhouse-Geisser correction revealed a significant difference in finger movement between the different scenarios in TaskSeek ($F(1,5) = 9.907,\ p = 0.025$), with participants moving their fingers more in the city scenario. There was no significant difference in finger movement between the different objects ($F(1,5) = 2.048,\ p = 0.212$), and no significant interaction effect between the scenarios and objects ($F(1,9) = 0.976,\ p = 0.368$). For TaskShow, there was a significant difference in finger movement between the different scenarios ($F(1,6) = 6.409,\ p = 0.045$), with participants moving their fingers more in the city scenario. There was no significant difference in finger movement between the different objects ($F(1,6) = 4.822,\ p = 0.070$), and no significant interaction effect between the scenarios and objects ($F(1,6) = 5.897,\ p = 0.050$).

Participants moved their fingers more in the city than in the rover scenario, except for Task 2. This increased movement is likely due to the city's more extensive and complex environment, where objects of interest are more spread out, necessitating more exploration and finger movement for transform gestures. The minimal difference in finger movement in Task 2 might be attributed to the fact that points of interest do not scale with the replica, making them easier to recognize without scaling, as they remain the same size.

5.4.1.4 Replica Translation

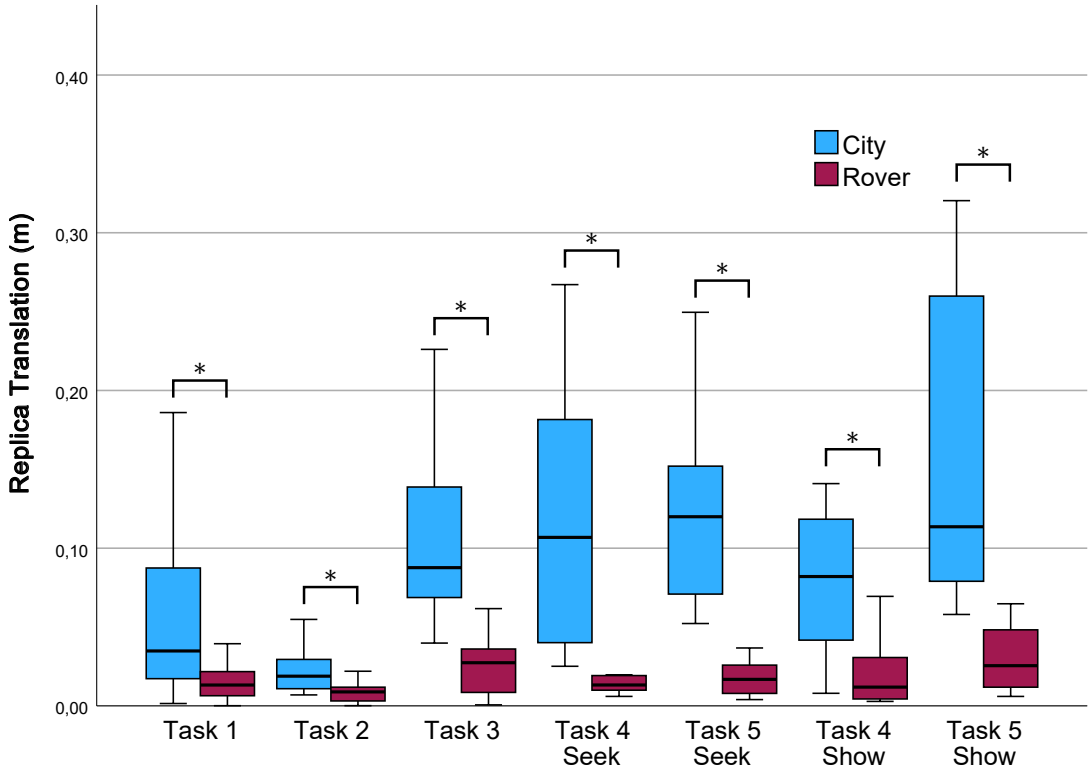


Figure 5.10: Box-plot of the cumulative sum of replica translation in meters for each task for each scenario. The asterisk (∗) indicates a significant difference between the city and rover scenarios.

Replica translation, measured in meters, represents the cumulative distance participants moved the replica. The results for each task are shown in Figure 5.11. All tasks showed a significant difference in replica translation between the city and rover scenarios, with participants moving the replica more in the city scenario (Task 1: $Z = -3.058,\ p = 0.002$; Task 2: $t(16) = 3.594,\ p = 0.002$; Task 3: $t(16) = 6.724,\ p < 0.001$).

A repeated measures ANOVA with Greenhouse-Geisser correction revealed a significant difference in replica translation between the different scenarios in TaskSeek ($F(1,4) = 34.498,\ p = 0.004$), with participants moving the replica more in the city scenario. There was no significant difference in replica translation between the different objects ($F(1,4) = 4.246,\ p = 0.108$) and no significant interaction effect between the scenarios and objects ($F(1,4) = 3.742,\ p = 0.125$). For TaskShow, there was a significant difference in replica translation between the different scenarios ($F(1,7) = 21.834,\ p = 0.002$), with participants moving the replica more in the city scenario. There was no significant difference in replica translation between the different objects ($F(1,7) = 3.039,\ p = 0.125$) and no significant interaction effect between the scenarios and objects ($F(1,7) = 4.955,\ p = 0.065$).

Like finger movement, participants moved the replica more in the city than in the rover scenario. However, unlike finger movement, this difference was evident across all tasks, including

Task 2. This discrepancy may be due to the primary interaction in Task 2 being the acknowledgment of points of interest, which requires the use of two fingers for balloon selection. In contrast, replica translation can be done with one finger, meaning replica translation may not have impacted finger movement as much as acknowledging points of interest did.

5.4.1.5 Head Translation

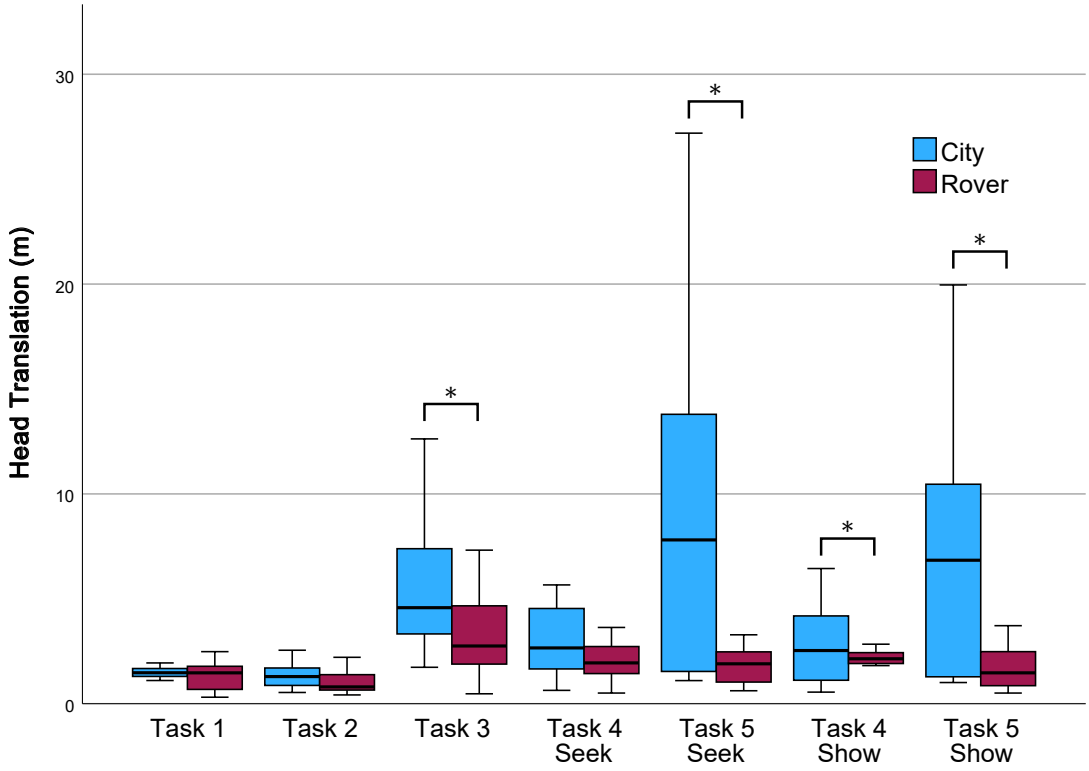


Figure 5.11: Box-plot of the cumulative head translation in meters for each task for each scenario.

Head translation, measured in meters, indicates the total distance participants moved their heads. Figure 5.11 displays the results for each task. In the first task, there was no significant difference in head translation between the city and rover scenarios ($t(9) = 0.550,\ p = 0.298$). Similarly, the second task showed no significant difference ($Z = -1.870,\ p = 0.062$). However, in the third task, participants moved their heads significantly more in the city scenario compared to the rover scenario ($Z = -2.427,\ p = 0.015$).

A repeated measures ANOVA with Greenhouse-Geisser correction identified a significant difference in head translation across scenarios in TaskSeek ($F(1,5) = 7.679,\ p = 0.039$), with more significant head movement observed in the city scenario. There was no significant difference in head translation between different objects ($F(1,5) = 4.505,\ p = 0.087$), but an interaction effect between scenarios and objects was significant ($F(1,5) = 7.237,\ p = 0.043$). To interpret this interaction, simple main effects analyses were conducted: head translation did not significantly differ across objects in the city scenario ($p = 0.064$) or the rover scenario ($p = 0.799$). Specifically, the

head movement did not significantly differ between the city and rover scenarios for the first object ($p = 0.106$) but did for the second object ($p = 0.040$).

In TaskShow, a repeated measures ANOVA indicated a significant difference in head translation between scenarios ($F(1,5) = 10.181,\ p = 0.025$), with more head movement observed in the city scenario. There was no significant difference in head translation across different objects ($F(1,5) = 4.445,\ p = 0.089$), nor was there a significant interaction effect between scenarios and objects ($F(1,5) = 6.043,\ p = 0.057$).

Overall, participants generally moved their heads more in the city scenario compared to the rover scenario, except for tasks 1 and 2, and Task 4 Seek.

5.4.2 Qualitative Data

The results from the questionnaires are discrete, ordinal data, as opposed to the continuous data from the metrics. As such, the Wilcoxon signed-rank test was used to determine if significant differences existed between the responses for each question across scenarios.

Table 5.1 displays the results for the first task. No significant differences were found between the city and rover scenarios for any of the questions. This suggests a similar task load in both scenarios. Participants generally did not report feeling mentally or physically strained, rushed, or insecure. They also felt successful in accomplishing their tasks and believed their performance required a lower-to-moderate level of effort.

Table 5.1: Median (IQR) and Wilcoxon Signed Ranks test results for each of the NASA-TLX questions for the first task in each scenario.

Question	Median (IQR)		Wilcoxon Signed Ranks
	City	Rover	
How mentally demanding was the task?	2 (1)	1 (1)	$Z = -0.977,\ p = 0.329$
How physically demanding was the task?	1 (1)	1 (0.75)	$Z = -0.322,\ p = 0.748$
How hurried or rushed was the pace of the task?	2 (1.75)	1 (2)	$Z = -1.578,\ p = 0.115$
How successful were you in accomplishing what you were asked to do?	5 (1)	5 (0.75)	$Z = -1.100,\ p = 0.271$
How hard did you have to work to accomplish your level of performance?	2 (1.75)	2 (2)	$Z = -0.844,\ p = 0.399$
How insecure, discouraged, irritated, stressed, and annoyed were you?	1 (1)	1 (0.75)	$Z = -0.551,\ p = 0.582$

Table 5.2 presents the results for the second task. The city scenario was statistically significantly more demanding than the rover scenario regarding the effort required to achieve the performance level ($Z = -2.055,\ p = 0.040$). No other significant differences were observed between the two scenarios. Participants did not indicate feeling mentally or physically strained, rushed, or insecure in either scenario. They reported feeling successful in completing their tasks. They

perceived a moderate level of effort required for the city scenario and a low level of effort for the rover scenario.

Table 5.2: Median (IQR) and Wilcoxon Signed Ranks test results for each of the NASA-TLX questions for the second task in each scenario. The asterisk ($*$) indicates a significant difference between the city and rover scenarios.

Question	Median (IQR)		Wilcoxon Signed Ranks
	City	Rover	
How mentally demanding was the task?	2 (1)	1 (1)	$Z = -1.311,\ p = 0.190$
How physically demanding was the task?	1 (1)	1 (0.75)	$Z = -1.265,\ p = 0.206$
How hurried or rushed was the pace of the task?	2 (1.75)	1.5 (1)	$Z = -0.905,\ p = 0.366$
How successful were you in accomplishing what you were asked to do?	5 (1)	5 (1)	$Z = -0.905,\ p = 0.366$
How hard did you have to work to accomplish your level of performance? $*$	2 (1.75)	1.5 (1)	$Z = -2.055,\ p = 0.040$
How insecure, discouraged, irritated, stressed, and annoyed were you?	1 (1)	1 (0)	$Z = -0.921,\ p = 0.357$

Table 5.3 presents the results for the third task. Interestingly, no significant differences were found between the city and rover scenarios for any of the questions despite metrics indicating significant differences in completion time and active time. Participants reported experiencing moderate mental and physical strain in both scenarios, with a moderate level of effort required to achieve their performance. They did not feel rushed or insecure in either scenario but felt moderately successful in completing their tasks.

Table 5.3: Median (IQR) and Wilcoxon Signed Ranks test results for each of the NASA-TLX questions for the third task in each scenario.

Question	Median (IQR)		Wilcoxon Signed Ranks
	City	Rover	
How mentally demanding was the task?	2.5 (1)	2 (2)	$Z = -0.819,\ p = 0.413$
How physically demanding was the task?	2 (2)	1.5 (1)	$Z = -1.311,\ p = 0.190$
How hurried or rushed was the pace of the task?	2 (2)	1.5 (1)	$Z = -1.394,\ p = 0.163$
How successful were you in accomplishing what you were asked to do?	4 (1)	4 (1)	$Z = -0.262,\ p = 0.794$
How hard did you have to work to accomplish your level of performance?	3 (2)	2 (2.5)	$Z = -1.279,\ p = 0.201$
How insecure, discouraged, irritated, stressed, and annoyed were you?	1 (1)	1 (1)	$Z = -0.431,\ p = 0.666$

Table 5.4 presents the results for the fourth task. Participants reported that the city scenario was significantly more mentally demanding ($Z = -2.365,\ p = 0.018$) and physically demanding

($Z = -2.310,\ p = 0.021$) than the rover scenario. They also felt significantly more successful in accomplishing their tasks in the rover scenario ($Z = -2.581,\ p = 0.010$). However, no significant differences were found between the two scenarios for the other questions. Participants did not feel insecure in either scenario and reported a low level of effort required to achieve their performance and a low level of feeling rushed.

Table 5.4: Median (IQR) and Wilcoxon Signed Ranks test results for each of the NASA-TLX questions for the fourth task in each scenario. The asterisk ($*$) indicates a significant difference between the city and rover scenarios.

Question	Median (IQR)		Wilcoxon Signed Ranks
	City	Rover	
How mentally demanding was the task? $*$	2.5 (1)	2 (1)	$Z = -2.365,\ p = 0.018$
How physically demanding was the task? $*$	2 (1)	1 (0.75)	$Z = -2.310,\ p = 0.021$
How hurried or rushed was the pace of the task?	2 (1.75)	1.5 (2)	$Z = -1.604,\ p = 0.109$
How successful were you in accomplishing what you were asked to do? $*$	4 (1)	5 (0)	$Z = -2.581,\ p = 0.010$
How hard did you have to work to accomplish your level of performance?	2 (1)	1 (1.75)	$Z = -1.787,\ p = 0.074$
How insecure, discouraged, irritated, stressed, and annoyed were you?	1 (1)	1 (0)	$Z = -1.150,\ p = 0.250$

Table 5.5 presents the results for the general questions. No significant differences were found between the city and rover scenarios for any of the questions. Participants generally did not feel nauseous during the tasks, found the WIM metaphor helpful for communicating points of interest, and found the representation of user locations on the replica useful for understanding intent. Two participants reported nausea levels above 2 in the city scenario, while none did in the rover scenario.

Table 5.5: Median (IQR) and Wilcoxon Signed Ranks test results for each of the general evaluation questions in each scenario.

Question	Median (IQR)		Wilcoxon Signed Ranks
	City	Rover	
How nauseous did you feel during the task?	1 (1)	1 (0)	$Z = -0.636,\ p = 0.525$
How useful was the world-in-miniature metaphor for communicating points of interest?	5 (0.75)	5 (0.75)	$Z = -0.333,\ p = 0.739$
How useful was the representation of user locations on the replica for understanding intent?	4 (1)	5 (1)	$Z = -0.093,\ p = 0.926$

Overall, the qualitative data analysis revealed that participants generally found the city scenario comparable to the rover scenario, except in the second and fourth tasks, where the city

scenario had a higher task load. Participants reported that the tasks were mentally and physically undemanding, requiring a low to moderate effort to achieve their performance. They generally did not feel rushed or insecure and felt successful in completing their tasks. However, the third and fourth tasks were considered more challenging regarding mental and physical demands and the required effort.

There were discrepancies between the qualitative data and the collected metrics. For instance, in the third task, participants reported similar task loads in both scenarios despite metrics indicating significant differences in completion time, active time, finger movement, replica translation, and head translation. This discrepancy may be due to participants' subjective perceptions, which may not always align with objective metrics. Participants answered the questions after completing each scenario, potentially basing their responses on their general feeling of the task rather than their actual experience.

5.4.3 Observations

During each test session, participants were observed to identify any issues or challenges they encountered while performing the tasks. They were encouraged to think aloud and provide general feedback on the usability of the approach.

One common issue across all tasks was the hardware quirks of the different touch surfaces. The infrared touch frame functions differently from what people expect: it does not track fingers directly on the table but uses infrared sensors positioned slightly above it. This can lead to many accidental touches, as users need to lift their hands higher than expected, which is tiring. Inevitably, users would forget, lower their hands, or rest their arms on the touch frame. This resulted in many accidental points of interest and premature teleportation with incorrect orientation. Much of the time spent on tasks, especially task 3, involved users accidentally creating points of interest and then trying to delete them.

Additionally, the touch frame's relatively low resolution made it challenging to select small objects or teleport to small zones, an issue that could be mitigated by scaling the replica. However, this was not intuitive for some users. The testing table with the infrared touch frame, shown in image b) of Figure 5.1 was placed near the limits of the tracking area, which caused the headset to lose tracking when users approached the table.

The Displax touch table also presented challenges, particularly with losing finger tracking when moving fingers due to high friction, making it hard to slide fingers across the surface. This would sometimes cancel gestures or switch finger recognition from one hand to the other. Fingers also needed to be more spread out; otherwise, the table would consider them one finger. Furthermore, the table's highly reflective surface caused the headset tracking to jitter when users approached the table.

Many participants found the teleporting gesture challenging to understand. The long press was particularly difficult, leading to numerous accidental points of interest. The rotation gesture for orienting the balloon was also confusing. Some participants did not grasp the rotation motion correctly; they rotated their fingers around themselves or used their wrists to rotate their hands.

Some even rotated the balloon counter-clockwise when a clockwise rotation required less effort, indicating a lack of understanding of the gesture. Another issue was that the balloon's position did not match the user's head but was instead aligned with the bottom of the table. This wasn't very clear for some users, as they expected to be positioned with their heads at the teleport destination. In retrospect, this was a design flaw, as the teleport destination should have been aligned with the user's head. Due to the challenges with understanding teleportation, many participants chose not to use it, even when it would have been beneficial in the collaborative tasks.

The tutorial video was one reason for the difficulty in understanding the teleportation gesture. The video was long and not interactive, making it hard for users to remember all the steps. Additionally, the video was shown at the beginning of the test session, and participants did not have the opportunity to rewatch it. During the training session, ensuring that users understood the teleportation gesture was challenging, especially since explaining it to two people simultaneously was difficult.

Task 3 was particularly challenging for participants, as some didn't understand that they had to orient themselves to face the object or that the zones changed color when they were inside them. Additionally, users did not instinctively scale the replica to aid in teleportation, causing the balloon to be too large for them to see if they were inside the teleportation zones. A slight oversight in the design of the teleportation zones further aggravated this issue. When using balloon selection, the system checks which interactable object the balloon is intersecting and prioritizes the teleportation zones over points of interest. At first glance, this approach seemed correct. However, if a point of interest was accidentally created near a zone, which happened frequently, it could obscure the zone. When this happened, users had difficulty deleting the point of interest because the system would prioritize the zone over the point of interest, making it hard to see if they were correctly inside it.

Two participants experienced difficulty focusing on objects in the city scenario. They struggled to understand the depth of the objects, which was discouraging for them. One of these participants, the only one to fail the task, could not show the second object in Task 5 due to difficulty determining the object's position relative to the balloon. This issue could be attributed to the imperfect implementation of the illumination effect at the touch surface's limits. The glow effect was always directed from the user's viewpoint instead of outwards or inwards from the touch surface's edges. For users with less depth perception or VR experience, seeing a single line of the limits made it difficult to discern whether it represented the farther or closer edge of the table.

Some general observations include that participants rarely used the vertical transform gesture, as they felt no need to. The second scenario during test sessions usually performed better than the first, as participants became accustomed to the system.

Participants provided several suggestions for improvement. They recommended replacing the video tutorial with a step-by-step interactive guide with small videos for each gesture. Additionally, they proposed implementing a feature to indicate the user's gaze direction by creating a point of interest wherever the user looked in the to-scale model or by displaying a ray. Another suggestion was to enable the simultaneous deletion or acknowledgment of multiple points of interest by

increasing the balloon's size. Since the balloon selection technique described in [5] supports resizing the balloon, this enhancement could be implemented using the same method. Furthermore, participants suggested incorporating a help feature to guide users on performing gestures. They also expressed a need for a distinct way to identify their color and appearance aside from their balloon, as they found it challenging to distinguish themselves from the other users.

Regarding positive feedback, participants appreciated being able to see each other, as it made the experience more social. They found balloon selection easy to use and intuitive. They also liked the general visual design and presentation of the system.

5.4.4 Discussion

The evaluation results indicate that the city scenario was comparable to the rover scenario regarding task completion time and active time across most tasks, with exceptions noted in tasks 3 and 5. This suggests that the approach is efficient for larger and smaller 3D models. However, participants faced more significant challenges in the city scenario, particularly regarding finger movement, replica translation, and head translation throughout all tasks. The increased movement in these metrics can be attributed to the city's larger and more complex environment, where objects of interest are more widely dispersed, requiring more exploration and physical interaction. Consequently, the city scenario appears to impose greater physical demands on participants than the rover scenario.

Qualitative data analysis corroborates that participants did not perceive significant differences in physical strain, mental effort, feelings of being rushed, or insecurity between the city and rover scenarios for tasks 1, 2, and 3. However, task 2 revealed that the city scenario required significantly more effort to achieve the performance level than the rover scenario. Similarly, for task 4, the city scenario was found to be notably more mentally and physically demanding than the rover scenario. Participants also reported feeling significantly more successful in completing tasks in the rover scenario. Responses to general questions indicated that participants generally did not experience nausea during the tasks, found the WIM (World-in-Miniature) metaphor effective for communicating points of interest, and found the representation of user locations on the replica useful for understanding intent.

The three tasks identified as most challenging for users, 3, 4, and 5, were characterized by metrics and qualitative feedback. Qualitative data suggests these tasks are moderately physically and mentally demanding, requiring moderate effort to achieve performance and moderate success in task completion. Metrics further demonstrate that participants spent more time on these tasks, increased their finger, replica, and head movements. Task 3 was primarily challenging due to difficulties understanding the teleportation gesture, while tasks 4 and 5 were particularly pronounced in the city scenario, where finding objects was more difficult. Interestingly, despite Task 5 in the city scenario being particularly challenging regarding completion time and active time, the metrics for finger transform, replica transform, and head transform do not indicate significant differences between Task 4 and Task 5. This suggests that the increased completion time experienced in Task 5

was not primarily due to heightened physical effort or manipulation of objects. Taking everything into account, the research questions can be addressed as follows:

RQ1 How efficiently can users create a point of interest on a given object? The results indicate that participants were able to create points of interest efficiently in both scenarios, with no significant differences in task completion time ($t(14) = 1.440,\ p = 0.172$) and active time ($t(10) = 0.134,\ p = 0.896$) between the city and rover scenarios for the first task. For the city scenario, the mean completion time was 64.92 ± 23.59 seconds. In comparison, the rover scenario had a mean completion time of 54.98 ± 24.30 seconds. Although the city scenario showed a slightly higher mean completion time, the difference was not statistically significant.

Similarly, for active time, the mean for the city scenario was 51.99 ± 22.08 seconds, while for the rover scenario, it was 39.68 ± 18.89 seconds. The active time metric provides a more precise measurement of user engagement and interaction with transform and balloon selection techniques. The lack of significant differences in task completion time and active time between the city and rover scenarios indicates that users can create points of interest efficiently regardless of the complexity or scale of the environment.

RQ2 How effectively does Replico notify users when a point of interest is created? The results suggest that Replico effectively notifies users when a point of interest is created, as all participants could identify all unacknowledged points of interest in both scenarios during the second task. The effectiveness was further assessed through user feedback collected via questionnaires. Participants reported that the second task was mentally undemanding, with median scores of $2(1)$ for the city and $1(1)$ for the rover scenario. The level of effort required was reported as low to moderate, with median scores of $2(1.75)$ for the city and $1.5(1)$ for the rover. Participants did not generally feel irritated, with median irritation scores of $1(1)$ for the city and $1(0)$ for the rover. They also felt very successful in accomplishing the task, with median success scores of $5(1)$ for both scenarios. Despite the generally positive feedback, the city scenario was found to be significantly more demanding in terms of effort compared to the rover scenario ($Z = -2.055,\ p = 0.040$). This suggests that Replico effectively notifies users when a point of interest is created in 3D models of varying complexity. Still, larger and more complex environments may require more effort to achieve the same performance level.

RQ3 How useful is the world-in-miniature metaphor for communicating points of interest? How useful is the representation of user locations on the replica for understanding intent? Qualitative data analysis revealed that participants found the world-in-miniature (WIM) metaphor useful for communicating points of interest. The median usefulness score was $5(0.75)$ for both the city and rover scenarios. Similarly, the representation of user locations on the replica was found to help understand intent, with median scores of $4(1)$ for the city and $5(1)$ for the rover, with no statistically significant differences between the

two. However, the teleportation feature was not as widely utilized as expected, making it challenging to assess the usefulness of the user location representation fully.

Despite this, the task completion and active time metrics for the collaborative tasks (Tasks 4 and 5) suggest that the WIM metaphor and user location representation are more effective in the rover scenario. For task completion time, Task 5 in the city scenario took significantly longer than Task 4 in the rover scenario ($Z = -3.154,\ p = 0.002$). Regarding active time, the city scenario had significantly higher active times than the rover scenario for the second object in both TaskSeek and TaskShow ($p = 0.017$) and ($p = 0.025$), respectively. This indicates that the WIM metaphor and user location representation are more effective in smaller, less complex environments where users can more easily navigate and interact with the replica.

Improvements could include making the teleportation gesture more intuitive to incentivize its use, enhancing the illumination effect at the touch surface's limits to improve depth perception, and addressing the issue of points of interest obscuring the objects they highlight.

RQ4 How user-friendly is Replico, and how much physical effort is required to use it? The qualitative results for the first two tasks indicate that participants generally found them physically and mentally undemanding, requiring a low to moderate level of effort to achieve their performance, and reached high levels of success while not feeling rushed or insecure. However, the third task was more challenging. Participants found it moderately mentally demanding ($2.5(1)$ in the city scenario and $2(2)$ in the rover scenario) and moderately physically demanding ($2(2)$ in the city scenario and $1.5(1)$ in the rover scenario). It required a moderate level of effort to accomplish ($3(2)$ in the city scenario and $2(2.5)$ in the rover scenario), and participants felt slightly less successful ($4(1)$ in both scenarios). In general, participants did not feel nauseous during the tasks in either scenario ($1(1)$ in the city and $1(0)$ in the rover).

Observations revealed hardware quirks with the touch surfaces, such as accidental touches and high friction on the Displax touch table, which made teleportation more difficult. Participants also found the teleportation gesture challenging to understand, particularly the long press and rotation gestures, leading to many accidental points of interest and premature teleportation. The teleportation destination also did not match the user's expectations, as it was aligned with the bottom of the table instead of the user's head.

Metrics for finger movement, replica translation, and head translation indicate that the city scenario required significantly more physical effort from participants than the rover scenario. Specifically, finger movement and replica translation were significantly higher in the city scenario for all tasks except task 2 for finger movement. This discrepancy between qualitative and quantitative data may be due to participants' subjective perceptions not always aligning with objective metrics. Additionally, qualitative data was collected after task completion, potentially influencing participants' responses.

Participants generally found the creation, acknowledgment, and deletion of points of interest user-friendly and physically undemanding. However, the teleportation gesture was challenging to understand, with participants often forgetting how to perform it correctly due to the number of techniques they were required to remember. Additionally, larger 3D models may require more physical effort to use Replico effectively.

Several improvements can be considered to improve user-friendliness. An interactive step-by-step gesture guide could help users understand and remember the necessary actions. Additionally, reworking the teleportation gesture to be more intuitive would likely reduce the frequency of accidental inputs and enhance user understanding. Addressing hardware issues such as unintentional touches and high friction on touch surfaces by using more sophisticated finger tracking could also improve user experience.

5.5 Summary

This chapter details the evaluation process for Replico, which addresses the research questions. The evaluation involved pairs of participants completing five tasks in two distinct scenarios: a large, complex city and a smaller, detailed Mars Perseverance rover. The first three tasks were performed individually, focusing on specific Replico techniques, while the last two tasks were collaborative, examining the system's performance in a collaborative setting. After each scenario, participants completed a NASA-TLX-based questionnaire to assess the task load.

The results were analyzed using both quantitative and qualitative methods. The quantitative analysis focused on metrics such as completion time and active time for measuring efficiency, finger movement, replica translation, and head translation for measuring physical effort. The qualitative analysis centered on responses from the NASA-TLX questionnaire and general observations. The findings indicated that users could efficiently create points of interest in both scenarios. Replico effectively notified users of new points of interest, although larger models demanded more effort from users.

The World-in-Miniature metaphor and the representation of user locations helped communicate points of interest, with greater effectiveness observed in the rover scenario. Overall, participants found Replico to be user-friendly and physically undemanding. However, the teleportation gesture proved challenging to understand, and the city scenario required more physical effort than the rover scenario.

Chapter 6

Conclusions

Collaboration is an integral part of human life, and it is essential to understand how to effectively enable it in digital tools to ensure social information is preserved and users remain aware of it. Virtual Reality (VR) holds great promise in this regard, as it facilitates more natural interactions with the digital world and allows users to share the same space. However, prolonged VR use can be physically tiring, limiting user engagement duration. DeskVR addresses this issue by allowing users to remain seated at their desks while fully immersed in a virtual environment, thus reducing the physical strain of VR use. This dissertation aimed to explore and design a collaborative approach that enables seated users to interact with a virtual environment, considering the constraints of a seated position, and to evaluate how this approach impacts user experience and collaboration.

The literature review explored concepts of social and workspace awareness, concurrency control, and existing work in the field of DeskVR. Although the study on concurrency control did not suggest specific avenues for exploration, it introduced the idea of personal workspaces, leading to an investigation of the world-in-miniature concept. Research in DeskVR highlighted the value of touch-based interactions in enhancing user comfort and reducing the physical strain of VR. The study on social and workspace awareness showed the importance of these concepts in collaboration and identified their key components. With this understanding, the requirements for the proposed approach were defined: to enable users to communicate collaboratively and effectively about objects or areas of interest in 3D models while minimizing physical effort and providing workspace awareness, all while remaining seated.

This dissertation proposes Replico, a collaborative approach for DeskVR that allows users to communicate about 3D models in a virtual environment using the world-in-miniature metaphor. Replico enables users to explore the 3D model using touch controls to manipulate a personalized miniature replica and to create points of interest using the Balloon Selection metaphor. These points of interest are replicated in both the miniature and the 3D model. Users are anchored to virtual tables representing their real-life counterparts and can join other users' tables to share their perspectives. Additionally, users can teleport around the 3D model to explore it from different angles in true scale. The miniature displays social information about users, such as their positions,

and identifies them by appearance. Points of interest are appearance-coded to correspond with the user who created them.

A prototype was developed using Unity to evaluate the proposed approach. Gesture detection was implemented using a state machine and uses the K-means clustering algorithm to track the user's hands. The system tracked the user's table with a VR controller to align with the real-life table. Various forms of visual feedback, such as finger trails and touch frame limits, were provided to enhance user interaction. Networking functionality was implemented using Unity's Netcode for GameObjects.

The approach was evaluated through a user study with 20 participants, forming ten pairs. Participants were required to perform five tasks to assess the efficacy and efficiency of several Replico techniques. These tasks were conducted in two scenarios to test the approach's applicability to various 3D models: a large city and a small but detailed Mars Perseverance rover. After each scenario, participants filled out forms based on the NASA-TLX to evaluate the usability of the approach and their experience with it.

The results showed that participants could efficiently create points of interest in both scenarios, with no significant differences in task completion and active time. Replico effectively notified users of new points of interest, though larger models required more effort from users. The world-in-miniature metaphor proved useful for communicating points of interest, especially in the smaller rover scenario. While users generally found the approach user-friendly and not physically demanding, they struggled with the teleportation gesture, indicating a need for a more intuitive solution.

These findings suggest that the approach is practical for collaborative interaction in VR, enabling efficient and user-friendly creation and acknowledgment of points of interest. The study highlights the potential and versatility of the world-in-miniature metaphor for enhancing workspace awareness and social information sharing in large, complex environments and smaller, detailed models. Additionally, it demonstrates the value of touch-based interactions and DeskVR in reducing physical strain and improving user comfort. However, the complexity of the environment increases the effort required from users, and the teleportation gesture needs further refinement. These insights point to areas for future improvement and further development.

6.1 Future Work

The evaluation of Replico revealed several areas for improvement and avenues for future work. One key area for enhancement is the teleportation gesture, which was found to be confusing and difficult to execute. Future work could explore alternative gestures or methods for teleportation that are more intuitive and user-friendly. Specifically, the teleportation orientation gesture could be refined. Research on designing tactile interfaces for older users suggests that touches should not have multiple functions based on duration or speed of press, nor should they include double taps. This principle could be applied to the teleportation gesture, which currently uses a long press to initiate teleportation. Improving finger tracking to be more robust against hardware limitations

or making gestures more resilient to faulty tracking and accidental touches could further enhance this feature. Additionally, incorporating a step-by-step interactive guide and a help menu could assist users in learning and utilizing the gestures more effectively.

Improving feedback on gestures is another area for future work. Color-coding finger trails to provide visual feedback on user actions, similar to the visual feedback in the implementation of SIT6 [2], could be beneficial. For instance, fingers from one hand could be red, while fingers from the other could be blue. Audio feedback when a gesture is recognized, a finger is detected, or a point of interest is created could also improve user experience, as suggested by studies on auditory feedback in tabletop interfaces [37]. Haptic feedback, such as vibrotactile feedback when a point of interest is created, could enhance the interaction experience.

Additionally, the teleportation behavior could be improved by matching the user's head position to the balloon's position, providing a better frame of reference for where the user will be teleported. Currently, it is ambiguous where the user's vertical position will be after teleportation.

Another point of improvement is addressing the issue of points of interest obscuring the objects they are meant to highlight. Future work could explore making points of interest more transparent or allowing users to adjust their transparency to improve visibility.

Users also reported difficulty determining when they were within the touch frame limits. Providing more explicit feedback on the touch frame limits, such as fading in from either inside or outside the frame limits, could address this issue. This would require a more complex shader implementation but could be a valuable addition to the system.

New interactions could be added to the approach to enhance collaboration and communication. For example, users could create zones of interest encompassing an area rather than a single point. Allowing users to simultaneously delete and acknowledge multiple points by increasing the balloon size could also be beneficial. This feature could also be used to create larger or smaller points of interest, prioritizing some points. Additional transformations, such as rotation on other axes, could be introduced. Another interaction could be the ability to lock the replica to prevent accidental movement or reset it to its original position. Users could also hide the replica from sight to focus on the 3D model.

Other potential interactions include showing the user's gaze direction to help others understand what the user is looking at, which could be achieved by displaying a ray from the user's head to the object they are looking at. With this interaction, users could even create points of interest by looking at an object and performing a gesture. Allowing users at the same table to communicate through audio could further enhance collaboration.

References

[1] Marilyn Jager Adams, Yvette J. Tenney, and Richard W. Pew. Situation Awareness and the Cognitive Management of Complex Systems. *Human Factors*, 37(1):85–104, March 1995.

[2] Diogo Almeida, Daniel Mendes, and Rui Rodrigues. SIT6: Indirect touch-based object manipulation for DeskVR. *Computers & Graphics*, 117:51–60, December 2023.

[3] Guilherme Amaro, Daniel Mendes, and Rui Rodrigues. Design and Evaluation of Travel and Orientation Techniques for Desk VR. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 222–231, March 2022.

[4] Steve Benford and Lennart Fahlén. A Spatial Model of Interaction in Large Virtual Environments. In Giorgio de Michelis, Carla Simone, and Kjeld Schmidt, editors, *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW '93*, pages 109–124. Springer Netherlands, Dordrecht, 1993.

[5] 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*, March 2007.

[6] Ray L. Birdwhistell. *Introduction to Kinesics: (An Annotation System for Analysis of Body Motion and Gesture)*. Department of State, Foreign Service Institute, 1952.

[7] 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*, I3D '97, pages 35–ff., New York, NY, USA, April 1997. Association for Computing Machinery.

[8] W. Broll. Interacting in distributed collaborative virtual environments. In *Proceedings Virtual Reality Annual International Symposium '95*, pages 148–155, March 1995.

[9] Herbert H. Clark. *Using Language*. 'Using' Linguistic Books. Cambridge University Press, Cambridge, 1996.

[10] Florian Daiber, Eric Falk, and Antonio Krüger. Balloon selection revisited: Multi-touch selection techniques for stereoscopic data. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 441–444, New York, NY, USA, May 2012. Association for Computing Machinery.

[11] Yufei Ding, Yue Zhao, Xipeng Shen, Madanlal Musuvathi, and Todd Mytkowicz. Yinyang k-means: A drop-in replacement of the classic k-means with consistent speedup. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 579–587, Lille, France, 07–09 Jul 2015. PMLR.

[12] Alan Dix, Janet E. Finlay, Gregory D. Abowd, and Russell Beale. *Human-Computer Interaction*. Pearson, 3 edition, September 2003.

[13] Christophe Domingues, Frederic Davesne, Malik Mallem, and Samir Otmane. Collaborative 3D Interaction in Virtual Environments: A Workflow-based Approach. In *Virtual Reality*, chapter 3. IntechOpen, January 2011.

[14] Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work*, CSCW '92, pages 107–114, New York, NY, USA, December 1992. Association for Computing Machinery.

[15] Mica R. Endsley. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1):32–64, March 1995.

[16] Thomas Erickson and Wendy A. Kellogg. Social translucence: An approach to designing systems that support social processes. *ACM Transactions on Computer-Human Interaction*, 7(1):59–83, March 2000.

[17] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[18] Sebastian Friston, Elias Griffith, David Swapp, Simon Julier, Caleb Irondi, Fred Jjunju, Ryan Ward, Alan Marshall, and Anthony Steed. Consensus Based Networking of Distributed Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, 28(9):3138–3153, September 2022.

[19] David M. Gaba, Steven K. Howard, and Stephen D. Small. Situation Awareness in Anesthesiology. *Human Factors*, 37(1):20–31, March 1995.

[20] William W. Gaver. Sound Support For Collaboration. In Liam Bannon, Mike Robinson, and Kjeld Schmidt, editors, *Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW '91*, pages 293–308. Springer Netherlands, Dordrecht, 1991.

[21] Saul Greenberg and David Marwood. Real time groupware as a distributed system: Concurrency control and its effect on the interface. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, pages 207–217, New York, NY, USA, October 1994. Association for Computing Machinery.

[22] Carl Gutwin and Saul Greenberg. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3):411–446, September 2002.

[23] O. Hagsand. Interactive multiuser VEs in the DIVE system. *IEEE MultiMedia*, 3(1):30–39, 1996.

[24] 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.

[25] Christian Heath, Marina Jirotka, Paul Luff, and Jon Hindmarsh. Unpacking collaboration: The interactional organisation of trading in a city dealing room. *Computer Supported Cooperative Work (CSCW)*, 3(2):147–165, June 1994.

[26] E. Hutchins. The Technology of Team Navigation. In Jolene Galegher, Robert E. Kraut, and Carmen Egido, editors, *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work*, page 552. Routledge, 1 edition, June 1990.

[27] Eunjee Kim and Gwanseob Shin. User discomfort while using a virtual reality headset as a personal viewing system for text-intensive office tasks. *Ergonomics*, 64(7):891–899, July 2021.

[28] Leslie Lamport. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM*, 21(7):558–565, July 1978.

[29] Jun Lee, Mingyu Lim, HyungSeok Kim, and Jee-In Kim. Supporting Fine-Grained Concurrent Tasks and Personal Workspaces for a Hybrid Concurrency Control Mechanism in a Networked Virtual Environment. *Presence*, 21(4):452–469, November 2012.

[30] Jason Leigh, Andrew Johnson, and Thomas Defanti. CAVERN: A distributed architecture for supporting scalable persistence and interoperability in collaborative virtual environments. *Virtual Reality: Research, Development and Applications*, 2:217–237, December 1997.

[31] John M. Linebarger and G. Drew Kessler. Concurrency Control Mechanisms for Closely Coupled Collaboration in Multithreaded Peer-to-Peer Virtual Environments. *Presence*, 13(3):296–314, June 2004.

[32] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982.

[33] J. MacQueen. Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symp. Math. Stat. Probab., Univ. Calif. 1965/66, 1, 281-297 (1967)., 1967.

[34] David Margery, Bruno Arnaldi, and Noël Plouzeau. A General Framework for Cooperative Manipulation in Virtual Environments. In Michael Gervautz, Dieter Schmalstieg, and Axel Hildebrand, editors, *Virtual Environments '99*, Eurographics, pages 169–178, Vienna, 1999. Springer.

[35] Susan E. McDaniel and Tom Brinck. Awareness in Collaborative Systems: A CHI 97 Workshop. In *The SIGCHI Bulletin*, volume 29, October 1997.

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

[37] Daniel Mendes, Sofia Reis, João Guerreiro, and Hugo Nicolau. Collaborative tabletops for blind people: The effect of auditory design on workspace awareness. *Proc. ACM Hum.-Comput. Interact.*, 4(ISS), nov 2020.

[38] Jesper Mortensen, Vinoba Vinayagamoorthy, Mel Slater, Anthony Steed, Benjamin Lok, and Mary Whitton. Collaboration in Tele-Immersive Environments. pages 93–101, January 2002.

[39] 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*, VRIC '13, pages 1–10, New York, NY, USA, March 2013. Association for Computing Machinery.

[40] Donald A. Norman. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Addison-Wesley Longman Publishing Co., Inc., USA, 1993.

[41] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. Virtual Replicas for Remote Assistance in Virtual and Augmented Reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pages 405–415, New York, NY, USA, November 2015. Association for Computing Machinery.

[42] Vasco Pereira, Teresa Matos, Rui Rodrigues, Rui Nóbrega, and João Jacob. Extended Reality Framework for Remote Collaborative Interactions in Virtual Environments. In *2019 International Conference on Graphics and Interaction (ICGI)*, pages 17–24, November 2019.

[43] Márcio S. Pinho, Doug A. Bowman, and Carla M. Dal Sasso Freitas. Cooperative object manipulation in immersive virtual environments: Framework and techniques. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '02, pages 171–178, New York, NY, USA, November 2002. Association for Computing Machinery.

[44] Marcio S. Pinho, Doug A. Bowman, and Carla M. Dal Sasso Freitas. Cooperative object manipulation in collaborative virtual environments. *Journal of the Brazilian Computer Society*, 14(2):53–67, June 2008.

[45] Pei-Luen Patrick Rau, Jian Zheng, Zhi Guo, and Jiaqi Li. Speed reading on virtual reality and augmented reality. *Computers & Education*, 125:240–245, October 2018.

[46] David Roberts and Robin Wolff. Controlling Consistency within Collaborative Virtual Environments. In *Eighth IEEE International Symposium on Distributed Simulation and Real-Time Applications*, pages 46–52, October 2004.

[47] David Roberts, Robin Wolff, Oliver Otto, and Anthony Steed. Constructing a Gazebo: Supporting Team Work in a Tightly Coupled, Distributed Task in Virtual Reality. *Presence Teleoperators & Virtual Environments*, 12:644–657, December 2003.

[48] David J. Roberts, Robin Wolff, and Oliver Otto. Supporting a Closely Coupled Task between a Distributed Team: Using Immersive Virtual Reality Technology. *COMPUTING AND INFORMATICS*, 24(1):7–29, 2005.

[49] D.J. Roberts and P.M. Sharkey. Maximising concurrency and scalability in a consistent, causal, distributed virtual reality system, whilst minimising the effect of network delays. In *Proceedings of IEEE 6th Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pages 161–166, June 1997.

[50] Roy A. Ruddle, Justin C. D. Savage, and Dylan M. Jones. Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments. *ACM Transactions on Computer-Human Interaction*, 9(4):285–308, December 2002.

[51] Roy A. Ruddle, Justin C. D. Savage, and Dylan M. Jones. Levels of Control During a Collaborative Carrying Task. *Presence*, 12(2):140–155, April 2003.

[52] Tony Salvador, Jean Scholtz, and James Larson. The Denver model for groupware design. *ACM SIGCHI Bulletin*, 28(1):52–58, January 1996.

[53] Leon D. Segal. Effects of checklist interface on non-verbal crew communications. Technical Report NASA-CR-177639, May 1994.

[54] S. S. SHAPIRO and M. B. WILK. An analysis of variance test for normality (complete samples)†. *Biometrika*, 52(3-4):591–611, December 1965.

[55] Alejandro Jarillo Silva, Omar A. Domínguez Ramirez, Vicente Parra Vega, and Jesus P. Ordaz Oliver. PHANToM OMNI Haptic Device: Kinematic and Manipulability. In *2009 Electronics, Robotics and Automotive Mechanics Conference (CERMA)*, pages 193–198, September 2009.

[56] Adalberto L. Simeone, Eduardo Velloso, and Hans Gellersen. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3307–3316, New York, NY, USA, April 2015. Association for Computing Machinery.

[57] Gurminder Singh, Luis Serra, Willie Png, and Hern Ng. BrickNet: A Software Toolkit for Network-Based Virtual Worlds. *Presence: Teleoperators and Virtual Environments*, 3(1):19–34, February 1994.

[58] Markus Sohlenkamp and Greg Chwelos. Integrating communication, cooperation, and awareness: The DIVA virtual office environment. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, pages 331–343, New York, NY, USA, October 1994. Association for Computing Machinery.

[59] Maurício Sousa, Daniel Mendes, Soraia Paulo, Nuno Matela, Joaquim Jorge, and Daniel Simões Lopes. VRRRRoom: Virtual Reality for Radiologists in the Reading Room. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 4057–4062, New York, NY, USA, May 2017. Association for Computing Machinery.

[60] Richard Stoakley, Matthew J. Conway, and Randy Pausch. Virtual reality on a wim: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, page 265–272, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[61] Sven Strothoff, Dimitar Valkov, and Klaus Hinrichs. Triangle cursor: Interactions with objects above the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 111–119, New York, NY, USA, November 2011. Association for Computing Machinery.

[62] Un-Jae Sung, Jae-Heon Yang, and Kwang-Yun Wohn. Concurrency control in CIAO. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pages 22–28, March 1999.

[63] FORGY E. W. Cluster analysis of multivariate data : efficiency versus interpretability of classifications. *Biometrics*, 21:768–769, 1965.

[64] R. Waters, D.B. Anderson, and D.L. Schwenke. Design of the Interactive Sharing Transfer Protocol. In *Proceedings of IEEE 6th Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pages 140–147, June 1997.

[65] Richard C. Waters, David B. Anderson, John W. Barrus, David C. Brogan, Michael A. Casey, Stephan G. McKeown, Tohei Nitta, Ilene B. Sterns, and William S. Yerazunis. Diamond Park and Spline:Social Virtual Reality with 3D Animation, Spoken Interaction, and Runtime Extendability. *Presence: Teleoperators and Virtual Environments*, 6(4):461–481, August 1997.

[66] Frank Wilcoxon. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[67] Jeonghwa Yang and Dongman Lee. Scalable prediction based concurrency control for distributed virtual environments. In *Proceedings IEEE Virtual Reality 2000 (Cat. No.00CB37048)*, pages 151–158, March 2000.

[68] F. Yeung. Internet 2: Scaling up the backbone for R&D. *IEEE Internet Computing*, 1(2):36–37, March 1997.

[69] Daniel Zielasko, Marcel Krüger, Benjamin Weyers, and Torsten W. Kuhlen. Menus on the Desk? System Control in DeskVR. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1287–1288, March 2019.

[70] Daniel Zielasko and Bernhard E. Riecke. To Sit or Not to Sit in VR: Analyzing Influences and (Dis)Advantages of Posture and Embodied Interaction. *Computers*, 10(6):73, June 2021.

[71] Daniel Zielasko, Benjamin Weyers, Martin Bellgardt, Sebastian Pick, Alexander Meibner, Tom Vierjahn, and Torsten W. Kuhlen. Remain seated: Towards fully-immersive desktop VR. In *2017 IEEE 3rd Workshop on Everyday Virtual Reality (WEVR)*, pages 1–6, March 2017.

[72] Daniel Zielasko, Benjamin Weyers, and Torsten W. Kuhlen. A Non-Stationary Office Desk Substitution for Desk-Based and HMD-Projected Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1884–1889, March 2019.

Appendix A

Finger Trail Compute Shader

The following code snippet is the compute shader used to generate the finger trails in the virtual touch frame.

```
#pragma kernel decay_finger_history

// This is the compute shader that calculates the decay of the finger history
// based on the current finger positions and the previous finger history.
// Each texture can store the history of two fingers, so we have 5 textures
// X is the inverse distance to the line, Y is the decay or longevity of the finger
// Z is the inverse distance to the line for the second finger, W is the decay or
    longevity of the finger for the second finger
RWTexture2D<float4> finger_history;
RWTexture2D<float4> finger_history_2;
RWTexture2D<float4> finger_history_3;
RWTexture2D<float4> finger_history_4;
RWTexture2D<float4> finger_history_5;

RWStructuredBuffer<float4> finger_positions;
RWStructuredBuffer<float4> last_finger_positions;
RWStructuredBuffer<float> average_incline;

float delta_time;
float linear_decay_rate;
float quadratic_decay_rate;
float finger_radius;

float distance_to_line (const float2 p, const float2 a, const float2 b)
{
    const float2 ab = b - a;
    const float2 ap = p - a;

    const float h = saturate(dot(ap, ab) / dot(ab, ab));
    return length(ap - h * ab) / finger_radius;
}
```

```
31
32 float point_in_front (const float2 p, float2 a, const float incline)
33 {
34     const float2 b = float2(a.x + finger_radius * cos(incline), a.y + finger_radius
           * sin(incline));
35
36     const float2 ab = b - a;
37     const float2 ap = p - a;
38
39     return (dot(ap, ab) / dot(ab, ab));
40 }
41
42 float2 get_finger_radius(const float2 p, const float2 a, const float2 b, const
       float average_incline, const float decay, const float previous_radius)
43 {
44     const float finger_distance = distance_to_line(p, a, b);
45     const float in_front = point_in_front(p, a, average_incline);
46
47     const float inverse_finger_distance = saturate(1.0 - finger_distance);
48     const float previous_inverse_distance = previous_radius * (1.0 - step(1.0, 1.0
           - decay));
49
50     return float2(
51         lerp(
52             previous_inverse_distance > 0.0 ? max(inverse_finger_distance,
                   previous_inverse_distance) : inverse_finger_distance,
53             finger_distance <= 1.0 ? inverse_finger_distance :
                   previous_inverse_distance,
54             saturate(in_front + 0.5)
55         ),
56         saturate(decay + 1.0 * (1.0 - step(1.0, finger_distance)))
57     );
58 }
59
60 float calculate_decay(const float decay)
61 {
62     return decay - linear_decay_rate * delta_time - quadratic_decay_rate *
           delta_time * delta_time;
63 }
64
65 float4 calculate_finger_history(const float2 p, float4 previous_history, float4
       last_positions, float4 current_positions, float2 average_incline)
66 {
67     const float finger_1_decay = calculate_decay(previous_history.y);
68     const float2 a1 = last_positions.xy;
69     const float2 b1 = current_positions.xy;
70
71     const float2 finger1_radius = get_finger_radius(p, a1, b1, average_incline.x,
           finger_1_decay, previous_history.x);
```

```
72
73      const float finger_2_decay = calculate_decay(previous_history.w);
74      const float2 a2 = last_positions.zw;
75      const float2 b2 = current_positions.zw;
76
77      const float2 finger2_radius = get_finger_radius(p, a2, b2, average_incline.y,
            finger_2_decay, previous_history.z);
78
79      return float4(
80          finger1_radius.x,
81          finger1_radius.y,
82          finger2_radius.x,
83          finger2_radius.y
84      );
85  }
86
87  [numthreads(8,8,1)]
88  void decay_finger_history (uint3 id : SV_DispatchThreadID)
89  {
90      const float2 p = float2(id.xy);
91      float4 history;
92
93      switch (id.z)
94      {
95          case 0:
96              history = finger_history[id.xy];
97              const float4 first_finger_history = calculate_finger_history(p, history
                    , last_finger_positions[id.z], finger_positions[id.z], float2(
                    average_incline[id.z * 2], average_incline[id.z * 2 + 1]));
98
99              finger_history[id.xy] = first_finger_history;
100             break;
101         case 1:
102             history = finger_history_2[id.xy];
103
104             const float4 second_finger_history = calculate_finger_history(p,
                    history, last_finger_positions[id.z], finger_positions[id.z],
                    float2(average_incline[id.z * 2], average_incline[id.z * 2 + 1]));
105             finger_history_2[id.xy] = second_finger_history;
106             break;
107         case 2:
108             history = finger_history_3[id.xy];
109
110             const float4 third_finger_history = calculate_finger_history(p, history
                    , last_finger_positions[id.z], finger_positions[id.z], float2(
                    average_incline[id.z * 2], average_incline[id.z * 2 + 1]));
111             finger_history_3[id.xy] = third_finger_history;
112             break;
113         case 3:
```

```
114             history = finger_history_4[id.xy];
115
116             const float4 fourth_finger_history = calculate_finger_history(p,
                    history, last_finger_positions[id.z], finger_positions[id.z],
                    float2(average_incline[id.z * 2], average_incline[id.z * 2 + 1]));
117             finger_history_4[id.xy] = fourth_finger_history;
118             break;
119         default:
120             history = finger_history_5[id.xy];
121
122             const float4 fifth_finger_history = calculate_finger_history(p, history
                    , last_finger_positions[id.z], finger_positions[id.z], float2(
                    average_incline[id.z * 2], average_incline[id.z * 2 + 1]));
123             finger_history_5[id.xy] = fifth_finger_history;
124             break;
125     }
126 }
```