LONDON'S GLOBAL UNIVERSITY
Department of Computer Science

Blended Environment in Mixed Reality Remote Collaboration: User Interaction Experiment

Pruetikorn Chirattitikarn

Supervisors: Prof. Anthony Steed

Faculty of Engineering

Department of Computer Science

University College London

This report is submitted as part requirement for Computer Graphics, Vision and Imaging MSc at University College London. It is substantially the result of my own work except where explicitly indicated in the text. The report may be freely copied and distributed provided the source is explicitly acknowledged.

6^{th} September 2024

Abstract

Recently, mixed reality technology has been integrated into remote collaboration systems, with numerous researchers developing different approaches to enable interactions between local and remote users within a mixed reality platform. These studies have facilitated collaborative scenarios by aligning objects from remote environments with those in the user's local environment to enhance accessibility. This study aims to determine the most effective method of blending environments from two users' rooms across three different conditions. The first condition, *Solid Blending*, presents a realistic virtual model of the remote environment; the second, *Transparent Blending*, renders remote objects as transparent; and the third, *Partial Blending*, only incorporates object models relevant to specific tasks. The study evaluates user interaction and behavior by tasking participants with finding matching Rubik's cubes and collecting positional and audio data, as well as responses to questionnaires on presence, talkativeness, and cooperation. The results reveal significant differences in interaction and behavior between the *Solid Blending* and *Partial Blending* conditions, with fewer notable differences observed between the *Solid Blending* and *Transparent Blending* conditions, as well as between the *Partial Blending* and *Transparent Blending* conditions.

Keywords— mixed reality, augmented reality, remote collaboration

I dedicate this to my mentors, family and friends. I am grateful for all your encouragement and supports. Your faith in me and your contributions to my academic journey have made this work possible.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor, Professor Anthony Steed, for all his support and kindness all along this project.

My sincere gratitude also go to Nels Numan and Ziwen Lu, postgraduate students from UCL computer science, PhD students under supervision of Professor Anthony Steed during my project time who had assisted and supported my project's procedure.

Declaration

I, Pruetikorn Chirattitikarn, declare that the thesis has been composed by myself and that the work has not be submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included and referenced. The report may be freely copied and distributed provided the source is explicitly acknowledged.

	01/09/2024
Signature	*Date*

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D: 2 dimensions

3D: 3 dimensions

API: Application Programming Interface

AR: Augmented Reality

AV: Augmented Virtuality

CPU: Computer Processing System

HCI: Human Computer Interaction

HMD: Head Mount Device

MR: Mixed Reality

VR: Virtual Reality

1 | Introduction

With the advancement of technology, the efficiency of communication and collaboration has increased dramatically, such as phone calls and internet. These mentioned methods have effectively connected people regardless of geographical distances via computers and smartphones. Consequently, remote collaboration has become increasingly important in a world that grows rapidly. However, these technologies still face limitations, particularly in providing access to local environments that require specialized equipment beyond standard communication devices. To address this issue, mixed reality (MR) has been introduced, overlaying virtual objects on physical scenes to enhance collaborative capabilities. Many technology companies and academic institutes research on remote collaboration with mixed reality to create an environment that users can feel others' presences and communicate seamlessly.

Over the past decade, the field of MR remote collaboration has focused on identifying the most effective user interfaces. Researchers have developed systems by adapting scenes that allow remote users to access local environments. Teo and Lawrence [2] introduced a system integrating both 360-degree video and 3D reconstruction to provide a real-time view of the surroundings. This approach offers complete environmental information to the remote user and is referred to as the *Solid Blending Condition* in this study. Grønbæk [6] developed a technique that merges only specific parts of each user's workspace, while Herskovitz [7] designed a toolkit for distributed collaboration using multiple methods such as shared anchors, portals, and world-in-miniature models. These examples represent the *Partial Blending Condition.* Additionally, Slater [8] explored user behavior in virtual environments with different rendering techniques, referred to in this study as the *Transparent Blending Condition.*

Previous works have conducted surveys related to collaborative MR. Fidalgo and Yan [9] focused on surveying the remote collaboration in MR environment, but solely for training and assistance. Numan [10] gave a result on asymmetric collaboration system where one user use VR HMD, while another AR HMD, which they collaborated in AR user's room only. Pidel and Ackermann [11] worked on reviewing collaborative system in AR and VR, but their work did not include merging two environments.

1.1 Aims and Objectives

The topic was initiated from allowing remote users access local user's environment fully while they were not colliding with their own surrounding. As previously stated, various approaches have integrated MR for collaborative scene, each has different advantages and limitations. This study focuses on user behavior and interaction within a blended virtual environment, created by merging the local environments of each user into a single scene. Additionally, the environment was presented in several styles to determine which approach was most effective for MR remote collaboration. The experiment were conducted with three conditions, *Solid Blending Condition*, *Transparent Blending Condition*, and *Partial Blending Condition*, which were based from the previous works, and compared the results from behavioral and statistical analysis.

Figure 1.1: An image presenting the experiment where two virtual rooms (Left, Right) reconstructing from physical place are merged into single space (Middle). This image is captured from Unity, which simulate the study.

1.2 Structure of the Report

In this thesis, Chapter 2 give an explanation of mixed reality and remote collaboration, including the past works which inspired this study in term of methodology and conditions. Then, the procedure throughout the project is demonstrated in Chapter 3, which covers designing, application platform, and measurement, and Chapter 4 provides the data analysis and observation result. Finally, the discussion and conclusion are reported in Chapter 5 and 6 respectively.

2 | Background

This chapter is divided into three parts, each explaining the background of this study. The first two sections cover the definition of the main topics. The first part discusses remote collaboration, the second part explains mixed reality technology, and the final part reviews past works that inspired this project.

2.1 Mixed Reality Technology

The concept of virtuality continuum was first explained in the work of Paul Milgram and Fumio Kishino [1]. The virtuality continuum, shown in Fig.2.1, is a scale of blending real environment and virtual environment based on visual perception. Each component shares similarities with the components next to it, and is distinctly different from those further away. Mixed reality, which is firstly introduced in their paper, is a segment which blend Augmented Reality (AR) and Augmented Virtuality (AV), creating an experience where the physical and digital environments coexist and interact seamlessly.

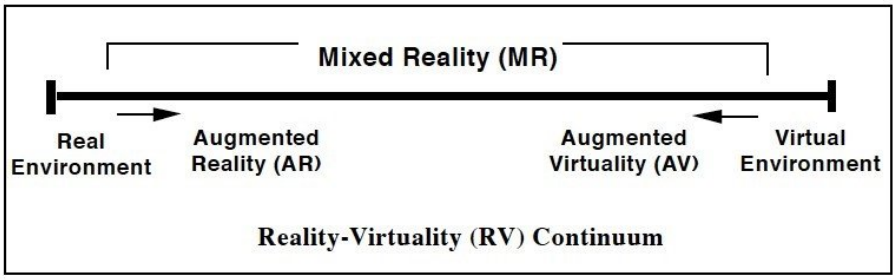


Figure 2.1: Reality-Virtuality Continuum [1]

To generate a fully immersive mixed reality display, systems or computers must accurately interpret environmental input and human perception. [12] In a mixed reality system, environmental input includes physical locations, object recognition, and users' body movements. The computer processing unit (CPU) then converts these inputs to create a co-existent scene that seamlessly integrates reality and virtual elements. [13]

2.2 Remote Collaboration

Remote collaboration, also known as remote assistance, is a process which multiple people communicating and collaborating regardless of their physically distant location. It contains several benefits to collaboration, such as reducing travel time and costs, and providing flexible workplaces and working hours. Simple problems can often be resolved with a phone call, while more complex issues requiring visual representation can be addressed through video calls or online meetings. However, some tasks are too complex for video calls and online meetings alone to handle effectively. [14, 15]

The emergence of AR/VR technology has significantly enhanced remote collaboration experiences. Many researchers have proposed seamlessly and synchronously combining remote collaboration with 3D display technology to improve both user actions and interactions. In tasks requiring physical interaction for communication, AR/VR, as well as mixed reality, can provide a richer experience. [16, 17]

For example, when workers need to perform maintenance tasks that require expert guidance, traditionally, experts must be on-site to analyze the problem. Phone calls or online meetings allow remote users to see the local surroundings, but they have limited fields of view because the scene is presented through a single camera. Additionally, local users may struggle to perceive direction and location accurately due to 2D displays lacking depth indicators. Mixed reality technology overcomes these limitations by allowing experts to view the scene and precisely indicate positions and directions. Presenting information in 3D enables local users to understand the instructions clearly, while remote users can comprehensively investigate the scene.

2.3 Examples of Study

This section covers the previous works that are relevant to this study, and based from above sections in this chapter. It is divided into 3 parts, XR collaboration, mutual awareness, user behavioral study.

2.3.1 XR Collaboration

Extended reality (XR) technologies, including augmented reality (AR), virtual reality (VR), and mixed reality (MR), overlays virtual objects on physical environments and both types of objects coexist seamlessly. Numerous applications support remote collaboration in VR, such as MeetinVR [18] and VRChat [19], as well as in AR/MR, like Spatial [20]. Research into collaboration within virtual environments has been ongoing for over two decades. With advancements in technology, these systems are now capable of supporting increasingly complex remote collaboration scenarios. [6, 7, 21]

Recently, researchers have increasingly adopted reconstructing the environment in 3D for remote collaboration. [22, 23, 24, 25] Nuernberger et al. [26] developed an application for 3D navigation within a reconstructed scene generated from a series of photographs. Their approach involved snapping the user's view to a photo to facilitate 3D reconstruction. Piumsomboon et al. [17] introduced a mixed reality system, CoVAR, which enables remote collaboration between users in augmented reality (AR) and augmented virtuality (AV) environments. This system employs 3D reconstruction using images captured by HoloLens. Both of these projects focus on creating 3D reconstructed scenes prior to sharing the view. Recent studies have shifted towards real-time 3D reconstruction for mixed reality remote collaboration. Their work demonstrates the ability to create accurate 3D meshes in real time. Additionally, Tian et al. [27] combined both static and live 3D reconstruction to create virtual replicas of specific objects. In their virtual environment, manipulable objects are tracked by three depth cameras, while static objects are clearly represented, resulting in an immersive experience for remote user interaction.

Teo et al. [2] advanced mixed reality remote collaboration by combining 360-degree video with static 3D reconstruction. This approach leverages the strengths of both systems to mitigate their respective weaknesses. Their study included an experiment comparing the benefits of each individual system, as well as their combined use. In their subsequent work, Teo et al. [28] further integrated these approaches by enabling users to share visual cues through a mixed reality display device. This setup allows remote users to track the presence of local users and provides visual cues—reflecting the remote user's viewpoint to the local user.

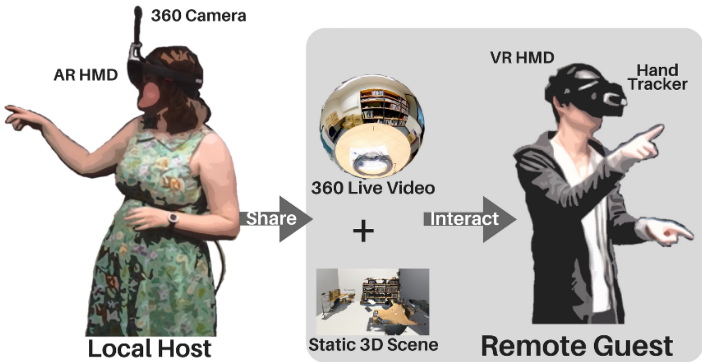


Figure 2.2: MR remote collaboration system overview. [2]

This study is inspired by these systems to use 3D reconstruction for mixed reality remote collaboration, including procedure and evaluation method for the experiment setup. The difference is that the system of this study merge 3D reconstructed scene from remote environment with the physical environment. As mentioned in section 1.1, it is developed for user to navigate the remote space with colliding the physical object. Moreover, in the future, the experiment will be implemented with 360 camera or depth cameras to reconstruct the scene in real time, and enable object tracking algorithm.

Herskovitz et al. [7] introduced an augmented reality (AR) toolkit designed for developing co-space collaboration applications. Their research explored various methods for creating a shared AR space, such as using physical objects as reference points, portals, and world-in-miniature models. One notable approach was mesh cropping, which allowed users to share only partial or contextually relevant sections of their collaborative spaces with others. Another significant contribution is the 'Blending Spaces' framework, presented by Cools et al. [3]. This work integrated multiple modalities and immersive scenes, bridging VR and AR through five interaction techniques. The blended space enabled users to experience merged environments combining elements of both VR and AR. This study adopts these approaches by creating a condition where only relevant objects are merged, while irrelevant context is removed, to compare interactions with a fully merged environment (Fig.2.3). When merging environments, another crucial aspect to consider is the rendering method or representation style. This consideration was inspired by Steptoe et al. [29], who examined interactions using various rendering techniques. The condition used in this experiment involved a transparent, fully merged environment.

Figure 2.3: The Blended Space which mixed augmented environment and virtual environment. [3]

2.3.2 Mutual Awareness

Awareness of both individual and group activities is crucial in collaborative scenarios and has been a key topic in research on remote collaborative virtual environments [7, 30, 31]. Gutwin et al. [5] defined the elements of workspace awareness, dividing them into two categories: "what is happening" and "where it is happening." For instance, in a sports team, each player must be aware of the positions and actions of their teammates on the field.

It was crucial to carefully consider how to collect information about individual interactions within the workspace and how to present this information to other participants. The framework was developed to illustrate the mechanisms of workspace awareness, a fundamental concept in designing groupware activities.

Element	**Relevant Questions**
Presence	Who is participating in the activity?
Location	Where are they working?
Activity Level	How active are they in the workspace?
Actions	What are they doing?
Intentions	What will they do next?
Changes	What changes are they making,and where?
Objects	What objects are they using?
Extents	What can they see? How far can they reach?
Abilities	What can they do?
Sphere of Influence	Where can they make changes?
Expectations	What do they need me to do next?

Table 2.1: Element of workspace awareness [5]

The design of MR collaboration seeks to enhance interactions with digital objects within physical environments, offering capabilities beyond those found in traditional collaboration. However, for these enhanced interactions to be effective, they must integrate a shared sense of space within users' environments. This study aims to identify the optimal conditions for collaborative experiences, which are essential for designing a system that supports mutual awareness.

2.3.3 User Behavioral Study

Prior research conducted a survey on remote collaborative system, which each has different measurement techniques and subjects. Numan et al. [10] explored an asymmetric collaborative mixed reality, where one user used VR device and another user used AR device. During the trials, their behavior were measured based on audio data and position in a shared environment, and they were interviewed afterward. Fidalgo et al. [9] analysed past work related to mixed reality remote collaboration which individual's work had different aspects to compare, for instance, degree of collaboration, input modality, visual display.

These studies influenced the hypotheses and questionnaire procedure of this project. The purpose of this study is to explore the most effective condition of blended environment. In other words, It is to observe each individual's interaction with blended environment's condition, and inspect the visual and spatial awareness of each user. In addition, the task that would be conducted in this study was inspired from Axelsson et al. [32], which studied on collaboration in immersive scene, and also could be referred the method of measuring behavior.

3 | Methodology

Mixed reality can enhance user experience on remote collaboration, since it can present the information in 3D space. [33] The aim is to observe user interaction on blending environment where the surrounding of remote user is merged into local user's surrounding, and to analyse the best condition of merging based on user interaction and satisfaction.

3.1 Research Questions

Regarding to previous works, this study explored the behavior of users engaging with merged environment, which was applied with various blending techniques. These conditions were divided into 3 techniques (Fig.3.1):

- CONDITION A: **Solid Blending.** In this condition, the virtual objects were fully opaque. This approach may affect how users perceive spatial relationships, and their ability to interact with both virtual and real-world objects. [2, 17, 26]
- CONDITION B: **Transparent Blending.** The virtual environment were rendered with transparency which had alpha value of 216 out of 255. It could make virtual objects appear more distinct of the physical environment, which could influence how users engage with these objects. [29]
- CONDITION C: **Partial Blending.** This condition involved blending only relevant context of virtual environment with the real world. In other words, this technique involved with presenting the virtual objects related to the tasks only. This method aimed to focus user attention on specific virtual elements, possibly changing the dynamics of interaction by emphasizing certain aspects of the MR experience. [3, 7]

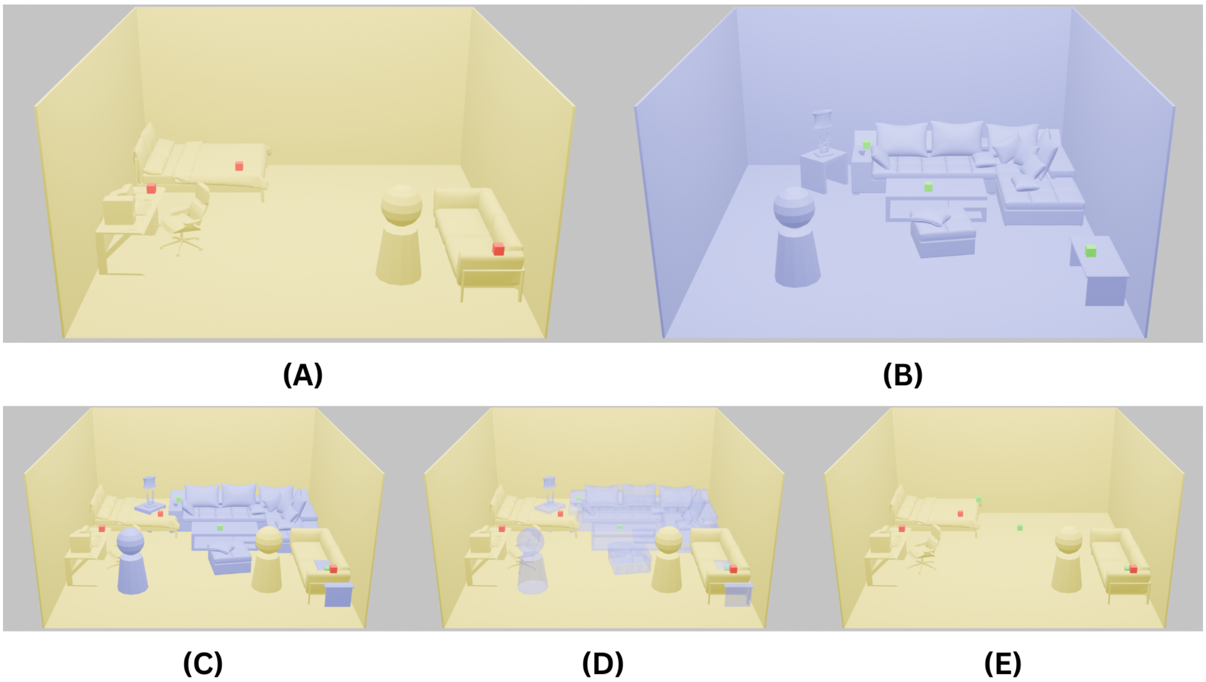


Figure 3.1: A visual concept images representing blending techniques. (A) A simulation model of *Room 1* with first user where yellow objects are the surrounding and red objects are context for trial. (B) A simulation model of *Room 2* with second user where purple objects mean the surrounding and green objects are context for trial. The next three images show the conditions from the perspective of first user, which environment of *Room 2* is virtual information: (C) *Solid Blending* (D) *Transparent Blending* (E) *Partial Blending*.

Within the context of this study, it was aimed to observe the different user interaction among these conditions by these research questions and find the answers of these questions.

Research Question 1. *Is there a difference in interaction between Solid Blending condition (A) and Transparent Blending condition (B)?*

This question sought to determine whether the transparency of virtual objects affects user interaction in an MR environment by comparing a fully opaque blending approach with an integrated transparent method. Steptoe et al. [29] found that different rendering conditions can influence how users perceive and distinguish between virtual and physical objects, which informed the sub-questions and hypotheses in this study. Additionally, the analysis of movement paths indicated that the *virtualized* approach, which displays objects differently from their real-world counterparts, resulted in distinct navigation patterns compared to other methods.

HYPOTHESIS 1. *Users will confuse with virtual objects in condition A, while users will not confuse with virtual objects in condition B.*

HYPOTHESIS 2. *Users will avoid both virtual and physical objects in condition A, while users will avoid only physical objects in condition B.*

HYPOTHESIS 3. *Users will travel faster in condition B than in condition A.*

Based on the previous works [10, 32], there were different level of activity in distinct modalities. Therefore, at the same level of modality, it could be assumed that:

HYPOTHESIS 4. *Users will have same participant-rated talkativeness score in both conditions.*

HYPOTHESIS 5. *Users will have same participant-rated presence score in both conditions.*

Research Question 2. *Is there a difference in interaction between Solid Blending condition (A) and Partial Blending condition (C)?*

The variation in approaches may impact talkativeness and user behavior. Previous studies [8, 10, 32, 34] have shown that interface asymmetry can influence factors like leadership and talkativeness. As a result, the questions in this study were designed based on the assumption that the differing number of virtual objects could affect user behavior in terms of interaction and conversation.

HYPOTHESIS 1. *Users will have more participant-rated talkativeness score in condition C than in condition A.*

HYPOTHESIS 2. *Users will speak longer in condition C than in condition A.*

In addition, in condition C, as there were less virtual objects in the rooms, it could be hypothesised that:

HYPOTHESIS 3. *Users will travel further in condition A than in condition C.*

HYPOTHESIS 4. *Users will travel faster in condition C than in condition A.*

HYPOTHESIS 5. *Users will complete the task faster in condition C than in condition A.*

Research Question 3. *Is there a difference in interaction between Transparent Blending condition (B) and Partial Blending condition (C)?*

Since there is a similarity of modalities between *Solid Blending* condition and *Transparent Blending* condition, it can be concluded the difference of interaction as same as second research question.

HYPOTHESIS 1. *Users will have more participant-rated talkativeness score in condition C than in condition B.*

HYPOTHESIS 2. *Users will speak longer in condition C than in condition B.*

HYPOTHESIS 3. *Users will travel further in condition A than in condition B.*

HYPOTHESIS 4. *Users will travel faster in condition B than in condition A.*

HYPOTHESIS 5. *Users will complete the task faster in condition B than in condition A.*

3.2 Experimental Design

The design of the experiment must consider the method to measure variable for research questions or hypotheses in section 3.1. The system needed to convey user behavior that answer these research questions effectively.

3.2.1 Apparatus

The system operated on remote collaboration with MR (Mixed Reality) users. The head-mounted display devices selected for this study had to support mixed reality APIs and enable a pass-through screen, allowing the user to view their surroundings.

The device used in this experiment was the Meta Quest 3, a standalone head-mounted display featuring two RGB cameras and depth detection sensors. It is an upgraded version of the Meta Quest 2 [35], offering a colorful pass-through and supporting a mixed reality experience. Additionally, it supports hand tracking and spatial mapping using SLAM (Simultaneous Localization and Mapping) technology [36]. The Meta Quest 3 also benefits from several libraries that extend its capabilities and large communities for developing and debugging mixed reality applications.

Figure 3.2: Meta Quest 3 [4]

3.2.2 Experimental Space

The experiment was conducted in two separate rooms on the same floor of a university building, referred to in the study as *Room 1* and *Room 2*, as shown in Fig.3.3. The rooms were selected for their similar sizes to facilitate the study's focus on blending environments of comparable dimensions.

Figure 3.3: Images of experimental rooms. (Left) *Room 1* and (Right) *Room 2*.

3.2.3 Tasks

The primary goal of the task was to allow users to navigate freely within the experimental space described in Section 3.2.2 and facilitate collaboration to complete the trials. The task design was adapted from previous studies on mixed reality remote collaboration [10, 8, 34], which involved different object-matching scenarios. In those studies, participants solved puzzles using virtual posters. For this experiment, the task was modified to involve searching for matching Rubik's cubes, inspired by Axelsson et al. [32]. This modification aimed to capture more detailed data on navigation paths and user movement. Using Rubik's cubes as the primary objects required participants to observe them from multiple angles to recognize their color patterns, leading to increased movement. Additionally, the complexity of the Rubik's cube patterns heightened communication between users, as they needed to assist each other and exchange information to successfully complete the task.

In the experimental setup, each room contained five Rubik's cubes, each with a unique color pattern and was positioned differently in the rooms (as shown in Fig.3.4). Additionally, three larger virtual cubes, referred to as task cubes, were displayed floating in mid-air and were three times the size of the real cubes. Above each virtual cube, number signs were placed as indicators for matching. (Fig.3.5) Participants were instructed to locate the Rubik's cubes with appearances matching the task cubes by placing the corresponding signs near the matched cubes. During the trial, both users can see ten Rubik's cubes, which five cubes were located in their own room physically, and other five cubes were presented from remote room virtually.

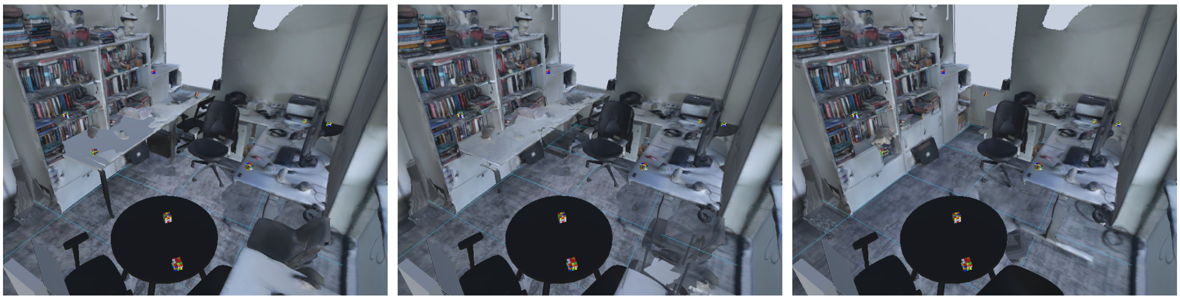


Figure 3.4: Images of Merged Environments Simulated in Unity Editor, which are viewed from *Room 1's* perspective. (Left) *Solid Blending* (Mid) *Transparent Blending* (Right) *Partial Blending*.

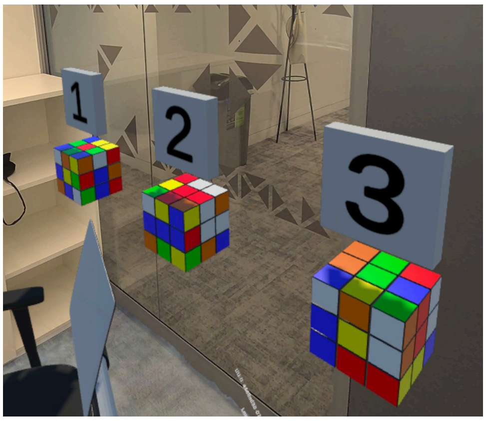


Figure 3.5: An image of task cubes appearing in the merged environment

3.3 System Implementation

This section describes how the program was developed based on the design outlined in Section 3.2. It details the process from creating models that simulate the rooms' environments to executing the system in the game engine. The system is built using Unity 3D Game Engine (version 2022.3.16f1) [37]. It is designed to perform mixed reality (MR) scenes by enabling a pass-through camera view, allowing users to see the physical world, and providing spatial awareness to overlay virtual objects onto reality. Unity offers plugins and library packages that support these requirements and allow integration with the Meta Quest 3. The code can be accessed in Appendix A.

3.3.1 Dependencies and APIs

To implement a mixed reality system for remote collaboration, it was necessary to install libraries that support mixed reality integration and apply a social platform framework. *Meta MR Utility Kit* [38] is a Unity package developed by the Oculus team that enables users to integrate mixed reality scenes on Meta Quest headsets. *Ubiq* [39] is an open-source social framework specifically designed for VR and is only available for Unity. By combining these two packages, the system can facilitate remote collaboration with an integrated mixed reality experience.

XR Interaction Toolkit [40] is a framework that facilitates 3D and UI interactions in AR, VR, and MR systems using Unity input events. It is a dependency for the Ubiq library and includes an asset called *XR Device Simulator*, which allows for simulating VR systems within the Unity Editor. This package enables users to connect to MR remote collaboration via a PC, but they can only join the system using a VR setup, not an MR system.

When building mixed reality application, it is essential to consider users' health and ensure the safety of the experiment for design, based on the guidelines. [41] As mentioned earlier, the system overlays digital objects onto physical, which means virtual display blocks a physical scene. (Fig.) It can cause a danger since the users cannot perceive the physical environment and might collide with it. An important asset for Unity's MR integration is *Depth API* [42]. This feature enhances mixed reality experiences by creating realistic scenes where virtual objects are accurately occluded by the physical environment based on depth perception. *Depth API* achieves this by rendering only the parts of virtual objects that are visible to the system, while omitting sections that are blocked by physical objects. By removing occluded parts, users can avoid physical objects smoothly.

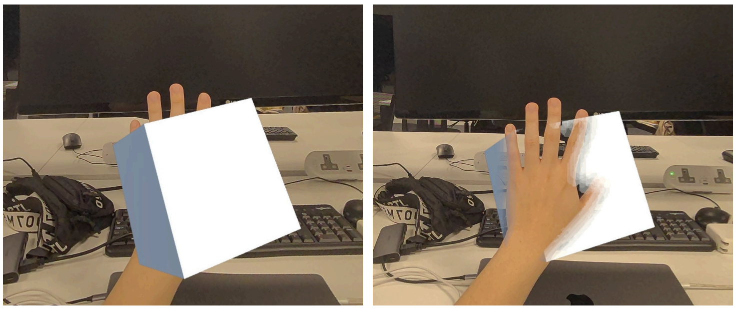


Figure 3.6: Images of virtual cubes without Depth API (Left) and with Depth API (Right).

3.3.2 Spatial Anchors

In mixed reality system, it is crucial to set a fixed position and rotation for the virtual object according to physical scene. Without configuration or marked position. the virtual objects will appear at a different position rather than the desired. An anchor is a fixed reference point in the real world that set the position and orientation of virtual point. It is one the features provided by *Meta MR Utility Kit* that locks the virtual object by physical environment. [43] This application used anchors to settle the origin point of the environments, which was approximately the centers of the rooms at floor level.

The code of managing spatial anchors was referenced from Black Whale Studio - XR Tutorials. [44] In *SpatialAnchorManager.cs*, the anchors are managed by the left controller, which the role of each button is determined in Table 3.1, and the name of each button was defined in Map Controller by Meta. [45] The position of anchor is fixed on the position of left-hand controller when *Button.PrimaryIndexTrigger* is pressed. In Unity, the game object has a component called Globally Unique Identifier (GUID), which can be used to reference an object, even when the scene, where contains the object, is not loaded. [46] Later, the anchor was saved by function provided by *OVRSpatialAnchor* class and its identifier was saved by *PlayerPrefs*, a Unity class that stores a data set by player between game scenes. [47] *AnchorLoader.cs* was a C# script that loaded anchors and was called when *Button.PrimaryThumbstick* was triggered. It generated the saved anchors based from GUID and local position according to physical scene.

3.3.3 Hand Controller Functions

For the task described in Section 3.2.3, participants were instructed to move the number signs to the matching cubes. They were required to use only the primary hand trigger on the right-hand controller to move the signs. In contrast, the left-hand controller was used to position the spatial anchors (Section 3.3.2).

Switch	Role
Button.PrimaryIndexTrigger	Create an anchor
Button.One	Save the last created anchor
Button.Two	Unsave the last created anchor
Button.PrimaryHandTrigger	Unsave all anchors
Button.PrimaryThumbstick	Load all saved anchors

Table 3.1: The function of each button on left controller

To configure the position of virtual environment's origin, it required four spatial anchors placing on the ground, and calculated the average position of these four anchors. The orientation was set with head-mount device's rotation, which set the direction from the anchor at the back to the one in front of user view. In order to set the center of the room accurately, it should be located at the bottom center of the wall. This method was written in *SceneConfiguration.cs*.

3.3.4 User Interface Screen

Room Selection Screen. (Fig.3.7) At the start, the user saw the screen with three buttons. This interface allowed users to choose the virtual environment that would be merged with physical scene by button at the left and center. The right button was created for spectator role, which teleported user, who used PC device, to the screen for room configuration (Fig.3.8).

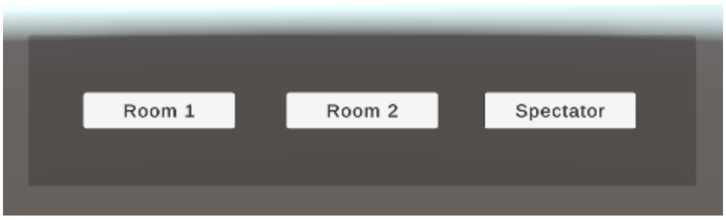


Figure 3.7: An interface screen for selecting the room or spectator role

Room Configuration Screen. (Fig.3.8) The interface enabled user to change the blending technique of the environments. At the top row, the buttons referred to the condition of solid blending, transparent blending, and merging with rubrics only from left to right respectively. The button at the bottom row would teleport player back to the room to have a conversation with others.

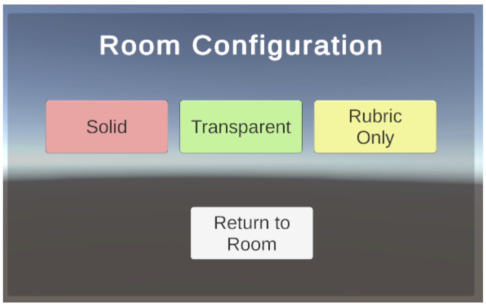


Figure 3.8: An interface screen for switch the blending condition

Ubiq Networking Screen. [39] (Fig.3.9) It was a screen to create a room server, and allow remote user to join the room. The host would create a room server by pressing a 'New' button and entering a room name. On the other hands, clients joined the room by pressing 'Join' button and entering a room code, which was generated after the room was hosted.

Figure 3.9: An interface screen of *Ubiq* networking component

3.3.5 Object Models

Sign Number Model. The task, explained in section 3.2.3, needed indicators to determine the matching cubes. In this experiment, it used a cube attached with text canvas showing the number. The sign's position and rotation relative to the parent transform was required to be shared to every user through Ubiq network component, which was programmed in *SignNumber.cs*.

Figure 3.10: The plate showing a number

Rubik's cube model. To ensure the comprehensive and accurate presentation of Rubik's cubes, they were modeled using Blender instead of scanning existing cubes. In this study, 10 virtual Rubik's cubes was created, each based on physical counterparts and featuring unique patterns to represent different problems. (Fig.3.11)

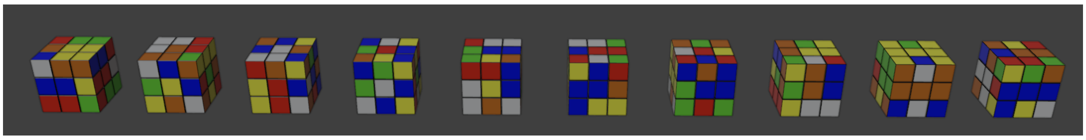


Figure 3.11: The models of Rubik's cubes created in Blender. The index of each cube is indicated by a number starting from 0 (Left) to 9 (Right).

Avatar model. The users, inside MR scene, were represented by random avatars provided by Ubiq avatar module. These avatars' shape is replicated from parts of human, which have only head, torso, and floating hands. Each part of body has a simple shape and a cartoon-like skin. (Fig.3.12) The position and orientation of the bodies in the experiment were determined by system devices. The floating hands were handled by Meta Quest 3 controllers and the head followed the movement of head-mount device. Avatar movement was processed through the networking module, which connected to a local network or Wi-Fi. The users, by a spatialized audio connection, communicated through voice over IP (VoIP) provided by Ubiq.

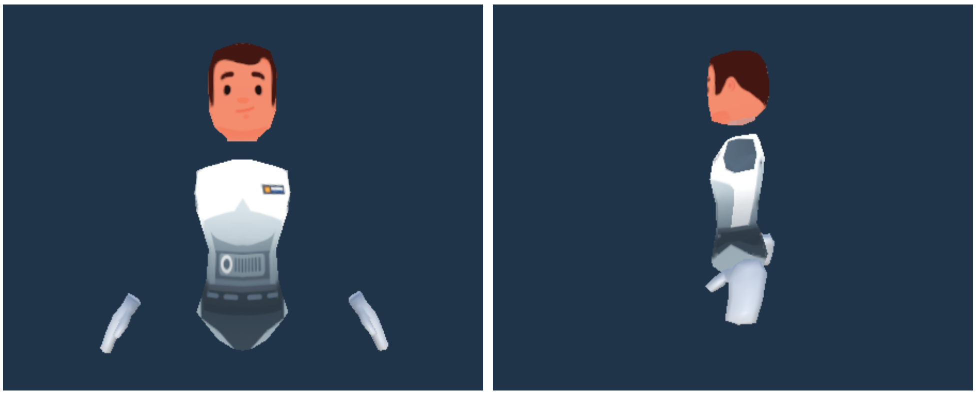
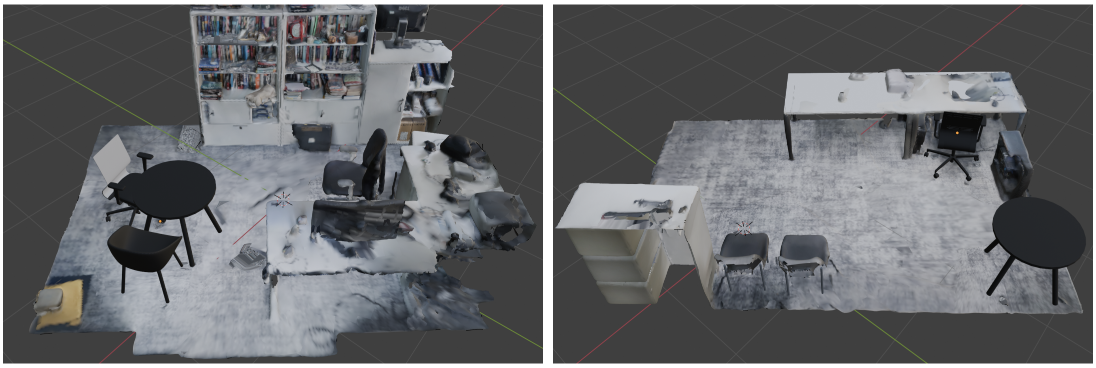


Figure 3.12: Example images of Ubiq avatar from the front (Left) and the side (Right).

Room model. The rooms, described in Section 3.2.2, were created using a 6.7-inch iPhone 13 Pro Max running iOS 17 and equipped with LiDAR. The scanning application used was the *Scaniverse* app [48]. After scanning, the room models were post-processed in Blender [49]. This involved filling in missing holes and removing excessive vertices. Additionally, some furniture was replicated using 3D models downloaded from CGTrader [50, 51] and modeled in Blender (Fig.3.13). The inclusion of synthetic models aims to examine the differences in user interaction between scanned data and synthetic data.



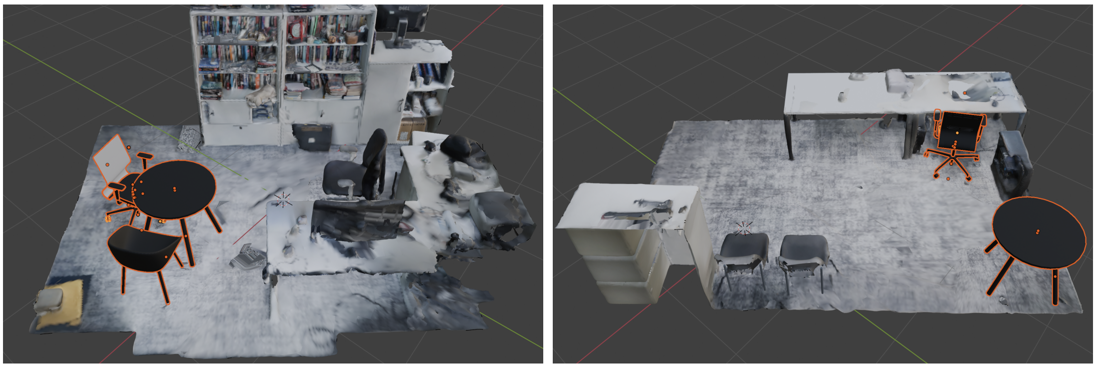


Figure 3.13: The post-processed models without the wall of *Room 1* (Left) and *Room 2* (Right) in Blender. The lower images shows the imported furniture which are indicated with orange edges.

3.4 Procedure

In the study, there were two participants, who were assigned to complete the tasks, and one experimenter. Before starting the trial, experimenter explained the information of the study to participants, and asked them to fill out the pre-study questionnaire with consent form. Then, participants were led to experimental rooms, where each user was assigned to a different room and was instructed to use HMD. Once both participants were ready, the experimenter took place in a distinct space and ran the system via PC, and explained verbally the tasks to participants. The participants, represented as random avatars, started the first trial, once they understood the detail. After each trial, users could ask to take a break before continuing the next trial, and experimenter changed the condition of blending technique to perform the same task. The Rubik's cubes that were needed to find the matching ones were different for each task. After completing all trials, participants were asked to fill the post-study questionnaire and do an unstructured interview for further detail. The participants received a £15 voucher for their time.

Figure 3.14: Images of experimental space before merging (Left) and after merging (Right).

3.4.1 Measurement

During the study, motion and audio data of the participants were recorded using *Ubiq*'s logging system and OBS Studio [52]. Logging system of *Ubiq* allowed experimenter collect the head-mount devices' position and orientation in real time, and OBS Studio recorded participants' audio and the simulation of the trial presented in Unity Editor Desktop app. These measures provided insight into user's level of activity, which displayed user behavior interacting with physical environment and virtual objects:

- *Spatial frequency* was visualized as a heat map by mapping the user path of both users. It explained how frequent each user walked in each area of the room.
- *Distance traveled* was measured based on the total movement of each participant within the experimental spaces. This measure revealed the amount of user activity depending on navigation.
- *Avatar head traveled speed* was calculated based on the value dividing distance traveled and the time interval spending during the trial. The data was delivered to *Ubiq* logging system, including the number of ticks of this instance. (i is an index

of instant speed, n is a number of tick time, x and z is horizontal axis position, t is tick time at the moment, and s is an instant speed. For i = 1,2,3,...,n)

$$s_i = \frac{\sqrt{(x_i - x_{i-1})^2 + (z_i - z_{i-1})^2}}{t_i - t_{i-1}}$$

- *Avatar head rotation velocity* was calculated based on user head movement captured by sensors on the devices. The data was delivered to *Ubiq* logging system, including the number of ticks of this instance.
- *Number of spoken words* was measured by transcribing the speech of each participant, and counting the words. This calculation result gave a level of talkativeness, which was involved in mixed reality collaboration. [8, 10, 32, 34]

3.4.2 User Study

Pre-study questionnaire. The questionnaire consisted of three sections. The first section was on demographics including age, gender, native language and others. It was to examine the variables that could influenced to the result. The second section inquired the familiarity of mixed reality and HMD, which might influenced user interaction with the interface and group activity. Finally, the last section asked about intimacy of the user to his/her partner to distinct the sample between completing task with acquaintance and with stranger.

Post-study questionnaire. After completing all tasks, participants were asked to fill out a questionnaire. The questions assessed various aspects of interaction, including activity level, communication, accessibility, presence, co-presence, and synergy. The questionnaire was adapted from prior research on mixed reality remote collaboration [10, 53]. To align with the specifics of this study, certain modifications were made, such as replacing references to a "puzzle" with "Rubik's cube" and omitting questions related to leadership from [10], as this experiment focused on collaboration among participants at an equal level.

To evaluate interaction during the collaborative tasks, participants were asked questions in three sections. The first section focused on activity level, where participants rated both their own and their partner's levels of activity and communication on a scale from 1 to 5 after each trial. The second section involved rating presence, co-presence, and the ability to distinguish between virtual and physical objects, also on a scale from 1 to 5, where '1' represented the lowest level and '5' the highest. The final section assessed cooperativeness, using the same rating scale.

4 | Results and Analysis

In this section, it presents the results relevant to the hypotheses, which referred to the differences across the conditions. To be more precise, the results showed the differences of user behavior and interaction between Solid blending condition and Transparent blending condition, and between Solid blending condition and Partial blending condition. The next part displays outcome was the relations between measures. Lastly, it reports the spatial behavior that happened during the study.

These results were concluded from 8 participants, or 4 dyads in total. Participants were recruited through posters spread across the university, mailing, and flyers given to the university's staffs and students. Participants were in the range of 18-24 years old (N=6) and 25-34 years (N=2), with a gender distribution of 2 women and 6 men. All of participants were Asian university students, and did not speak English natively (N=8). Half of participants never knew their partners (N=4) before the study. Furthermore, most participants have experienced with VR or MR system before (N=7).

Some data were not included in the analysis. Firstly, the video recording of the first dyad were missing due to an error in saving files. Secondly, because of a system error, avatar head rotation were not recorded in the first two trials.

Variable	A		B		C		df	F-stat	F-crit	p
	mean	std	mean	std	mean	std				
Talkativeness	3.125	0.464	3.469	0.632	4.344	0.785	2	5.036	3.467	.016
Presence	3.25	1.524	3.344	2.420	3.141	3.234	2	0.274	3.044	.759
Cooperativeness	3.651	1.004	3.813	1.073	4.036	1.152	2	3.869	3.023	.022
Spoken Words	596	397891	208.667	44916.333	94	12391	2	1.368	5.143	.324
Completion Time	417	98839	200.333	12280.333	115.333	1297.333	2	1.937	5.143	.224
Distance traveled	164.682	7816.253	95.72	2624.325	54.945	620.601	2	6.676	3.467	.006
Avatar Speed	0.187	0.193	0.248	0.231	0.273	0.24	2	18525.652*	5.991**	$<$.001
Roll	-0.168	128.615	6.208	156.441	7.385	201.97	2	703.742*	5.991**	$<$.001
Pitch	0.06	34.037	-0.109	41.878	-0.711	46.324	2	0.048*	5.991**	.976
Yaw	-0.096	146.414	-0.925	179.07	3.039	214.488	2	6.717*	5.991**	.034

Table 4.1: Overview of descriptive statistics between conditions. A: *Solid Blending* and B: *Transparent Blending*, and C: *Partial Blending*. *H-statistic. **H-critical.

4.1 Statistical Analysis

Building on the research questions outlined in Section 3.1, this section presents an analysis comparing talkativeness and interaction across different conditions, specifically focusing on the differences across 3 conditions consisting of *Solid Blending* and *Transparent Blending*, and *Partial Blending*. After the experiment, spatial data of participants' movements were collected through a logging system, capturing avatar position and orientation from the HMD. This data was then processed to calculate speed and visualized using heat maps and navigation paths.

4.1.1 Avatar traveled Speed

Travel speed was calculated by dividing the difference between two positions by the time difference when those positions were recorded. The analysis involved averaging the speeds of participants' avatars and comparing them across different conditions. Additionally, outliers—defined as data points with a z-score greater than 3 or less than -3—were identified and removed before conducting the tests.

Before comparing the speed data, it was necessary to assess whether the data was normally distributed across the sample. The Shapiro-Wilk test was conducted, and the results showed that the *Solid Blending*, *Transparent Blending*, and *Partial Blending* conditions had p-values less than 0.05 ($p < .001$). Since these p-values were all less than 0.05, the null hypothesis that the data followed a normal distribution was rejected. Additionally, the boxplot analysis (Fig.4.1) further confirmed that the data was not normally distributed.

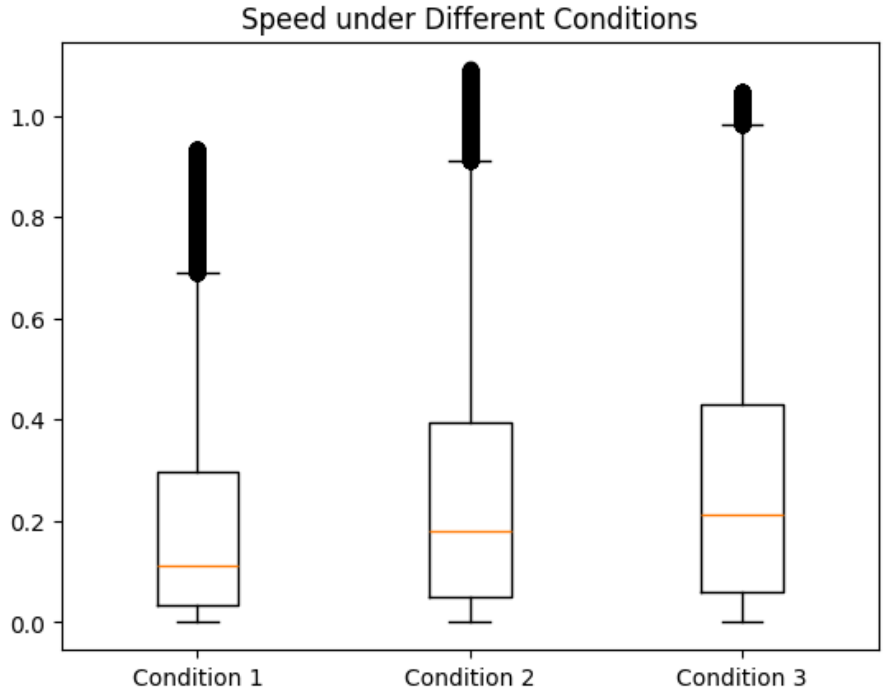


Figure 4.1: The speed under different conditions

Since the data was not normally distributed, it was inappropriate to use analysis of variance (ANOVA) or t-tests to compare differences. Instead, the Kruskal-Wallis test was employed to compare all conditions simultaneously, including Dune's post hoc test to indicate specific differences between the conditions, and the Mann-Whitney test was used to compare differences between *Solid Blending* and *Transparent Blending*, as well as between *Solid Blending* and *Partial Blending*. The Kruskal-Wallis test indicated significant differences among all conditions ($p < 0.001$). The hypothesis, which assumes that speeds during the *Solid Blending* condition will be significantly different with those in the *Transparent Blending* and *Partial Blending* conditions, is not rejected by paired-sample Dune's post hoc tests shown in Table 4.2. In fact, the speed data for the *Solid Blending* condition are significantly greater than those for the *Transparent Blending* and *Partial Blending* conditions ($p < .001, p < 0.001$ respectively) as they were tested with pairwise one-tail Mann-Whitney tests.

The analysis also required consideration of the avatar's head rotation speed. Rotation speed was divided into three components: rotation around the front-to-back axis (*roll*), rotation around the left-to-right axis (*pitch*), and rotation around the vertical axis (*yaw*). Normality was tested using the same method as for avatar travel speed, and it was found that none of the rotation speeds for any condition were normally distributed, as all p-values were less than 0.05. Similar to the travel speeds, the null hypothesis—which posited that data across different conditions would exhibit the same tendencies—was rejected by the Kruskal-Wallis test. As shown in Table **??**, the *Solid Blending* condition demonstrated significant differences in roll speeds ($p < .001$). The analysis also indicates that the null hypothesis cannot be rejected for pitch data, as there are no significant differences across the conditions. In contrast to roll, there is no evidence to reject the null hypothesis for yaw speeds between the *Solid Blending* condition and the other two conditions. However, a statistically significant difference was found in yaw speeds between the *Transparent Blending* and *Partial Blending* conditions ($p = .029$).

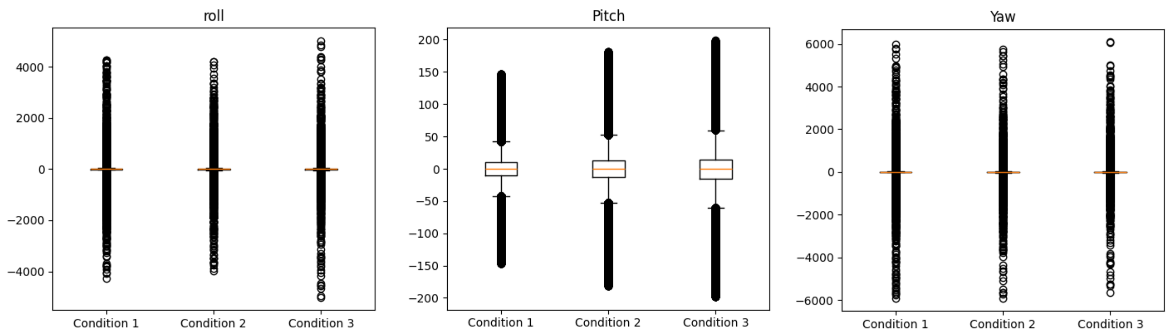


Figure 4.2: The rotation speed under different conditions

First condition	Second condition	Avatar Speed	Roll	Pitch	Yaw
Solid Blending	Transparent Blending	$< .001$	$< .001$	.999	.537
Solid Blending	Partial Blending	$< .001$	$< .001$	.999	.191
Transparent Blending	Partial Blending	$< .001$	.061	.999	.029

Table 4.2: The pairwise p-values of the Dunn's post hoc test with Bonferroni correction comparing first condition and second condition.

4.1.2 Activity Level

This section explains the difference of participants' activity, which was related to the interaction during tasks, under different conditions. It covers analysis on distance traveled, number of spoken words, completion time, spatial frequency, and navigation path, and the data was visualized as descriptive statistics and map graphs.

Distance traveled. The distance traveled by participants during the trials reflected their activity levels as they moved around the rooms searching for matching cubes. The results showed that participants (N=8) walked the most in the *Solid Blending* condition, with an average distance of 164.68 meters, followed by the *Transparent Blending* condition, with an average of 95.72 meters. The *Partial Blending* condition had the shortest median distance traveled, at 54.94 meters. As shown in Table 4.3, these findings support the hypothesis that participants would cover different distance in the *Solid Blending* condition compared to the *Partial Blending* condition.

Number of Spoken Words. The spoken words were transcribed from the video recordings using Clipto [54], with non-experiment-related noises and irrelevant words manually removed. Due to the poor audio quality of the videos, some parts of the conversation were not captured in the transcription, potentially leading to errors in the analysis. In the *Solid Blending* condition, participants spoke the most, averaging 596 words. In the *Transparent Blending* condition, the average word count was 208.67 words, while in the

Partial Blending condition, participants spoke the least, with an average of 94 words. As previously noted, the number of spoken words can provide insight into talkativeness. Based on Table 4.1 and Table 4.3, there is a statistically significant difference among the conditions. These finding support the hypothesis: that participants will have same talkativeness scores in the *Transparent Blending* condition than in the *Solid Blending* condition. However, they contradict another hypothesis stating that participants will have significant difference on talkativeness scores between *Partial Blending* condition and *Solid Blending* condition.

Completion Time. The results showed an inverse pattern compared to the distance traveled. All four dyads completed the trial more quickly in the *Partial Blending* condition than in the other conditions, with an average completion time of 6 minutes and 57 seconds. In contrast, the tasks in the other conditions took longer to complete, with an average time of 3 minutes and 20 seconds in the *Solid Blending* condition and 1 minute 55 seconds in the *Transparent Blending* condition. The p-values from Table 4.1 ($p = .224$) and Table 4.3 ($p = .275$) shows that there is not strong evidence of a difference between *Solid Blending* and *Partial Blending* condition conditions, which contradicts the hypothesis that participants would complete the trial faster in *Partial Blending* condition than in *Solid Blending* condition.

First condition	Second condition	Distance traveled	Spoken Words	Completion Time
Solid Blending	Transparent Blending	.077	.370	.323
Solid Blending	Partial Blending	.004	.246	.174
Transparent Blending	Partial Blending	.062	.453	.275

Table 4.3: The pairwise p-values of one-tail t-test on the difference between first condition and second condition.

4.1.3 Analysis on Questionnaire

The questionnaire were asked after completing all tasks, and the questions were divided into 3 sections, talkativeness, presence and cooperativeness. The purpose of these questions was to observe an insight of user experience during the study thoroughly, and these questions were referenced from Numan et al. [10]; However, this study focused on collaborating on equal stance, so the questionnaire removed the part related to leadership.

Talkativeness. This section includes questions about talkativeness and activity level, where participants rated both their own and their partner's level of activity and communication on a scale from 1 to 5. The final score for each participant per condition are calculated by averaging the self-rated scores and partners' rated scores. This section is also related to two hypotheses mentioned in Section 3.1: "participants will have the same participant-rated talkativeness score in both *Solid Blending* and *Transparent Blending* conditions," and "participants will have a higher participant-rated talkativeness score in the *Partial Blending* condition than in the *Solid Blending* condition."

After conducting a two-tail t-test, the results shows that the *Solid Blending* and *Transparent Blending* conditions had a p-value of 0.186 ($p = .186$), which is greater than 0.05. This indicates that the variables in these conditions are not evidently different. However, when comparing the *Solid Blending* and *Partial Blending* conditions, the low p-value ($p = .008$) and the t-statistic ($t\text{-}stat = -4.847$) exceeding the critical value of 1.6698 suggested that *Partial Blending* condition has a significant difference of talkativeness with *Solid Blending* condition.

Presence and Co-presence. The presence and co-presence variables were tied to the hypothesis for the first research question, which posited that "Participants will have the same participant-rated presence score in both *Solid Blending* and *Transparent Blending* conditions." Participants answered eight related questions concerning the accessibility of their assigned room, their partner's room, the presence of their partners, and their ability

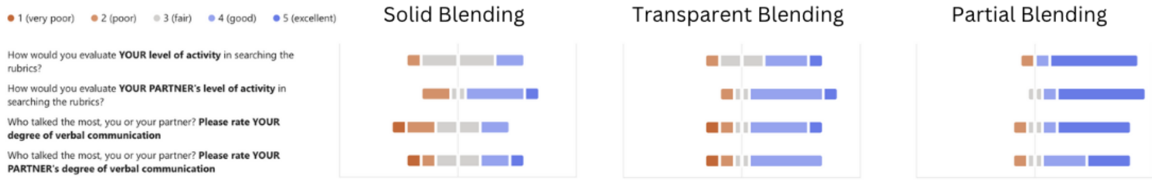


Figure 4.3: The box plot of questions related to talkativeness

to distinguish between virtual and physical objects.

The t-test results for the *Solid Blending* and *Transparent Blending* conditions showed a p-value of 0.353 for the one-tailed test and 0.706 for the two-tailed test, both of which are greater than 0.05. This indicates no statistically significant difference between the means of the two conditions. Additionally, the t-statistic ($t\text{-}stat = 0.378$) was significantly lower than the critical values ($t\text{-}critical = 1.658$ and *t Critical two-tail:* 1.98), leading to the failure to reject the null hypothesis that the variables in both conditions were equal. A similar t-test between the *Solid Blending* and *Partial Blending* conditions yielded the same conclusion, as the p-value exceeded 0.05 ($p = .689$), indicating that both conditions exhibited the same levels of presence and co-presence.

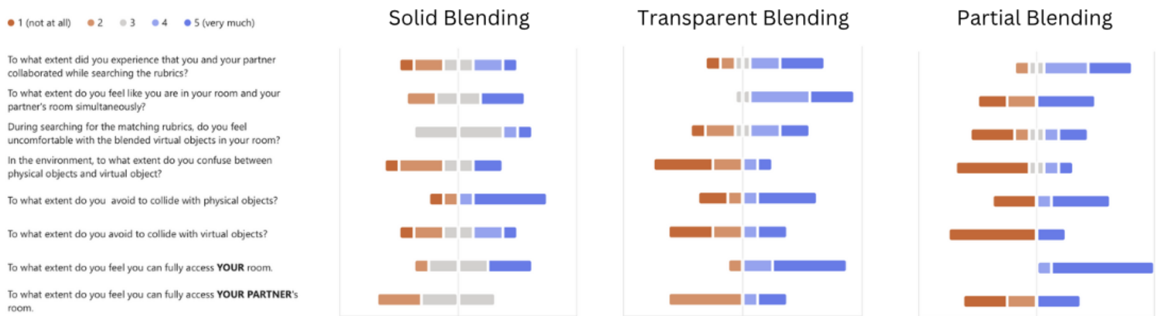


Figure 4.4: The boxplot of questions related to presence and co-presence

Cooperativeness. The final section evaluated the level of collaboration between the dyads, focusing on differences in interaction and collaboration across conditions. The assessment used the same method as the evaluation of presence and co-presence, employing a paired two-tailed t-test to test two hypotheses. The first hypothesis stated, "there will be no difference between *Solid Blending* and *Transparent Blending* conditions," while the second hypothesis was, "there will be no difference between *Solid Blending* and *Partial Blending* conditions."

The analysis using the two-tailed t-test fails to reject the null hypothesis on the difference between *Solid Blending* and *Transparent Blending* conditions, as the p-value ($p = .884$) exceeded 0.05, However, the second t-test between *Solid Blending* and *Partial Blending* conditions results in a p-value ($p = .006$) less than 0.05, which suggests there is a statistically significant difference between two conditions.

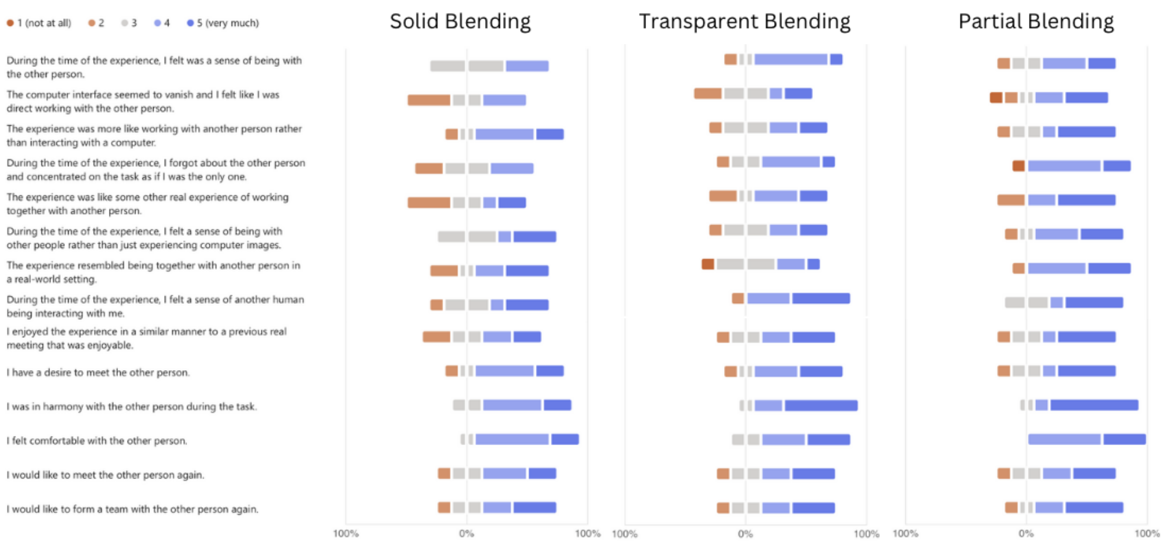


Figure 4.5: The boxplot of questions related to cooperativeness

First condition	Second condition	Talkativeness	Presence	Cooperation
Solid Blending	Transparent Blending	.369	.706	.884
Solid Blending	Partial Blending	.008	.689	.006
Transparent Blending	Partial Blending	.056	.496	.470

Table 4.4: The pairwise p-values of two-tail t-test on questionnaire stating that first condition and second condition are significantly different.

4.2 Spatial Behavior

The avatars' movement data were captured through network logs, while activity and conversation were recorded via screen recording in Unity Editor. During the study, all participant pairs referred to the task cubes using the numbers displayed on the signs above the cubes. It was observed that several pairs, after identifying the matching cubes, confirmed to each other if they were seeing the same cubes in their respective rooms. Based on the room layout (Fig.3.3) and the positions of the Rubik's cubes (Fig.3.4), there were two cubes placed on a black table in *Room 1* and another two cubes on a black table in *Room 2*, with both pairs being at similar distances relative to the size of the sign numbers. The primary strategy participants used to reference the matching cubes during the trial involved spatial terms such as "inside," "outside," "left," and "right." For example, U1 is user in *Room 1* and U2 is user in *Room 2*:

U2 : *"I think number one is this cube. This one?"*

U1 : *"Yeah, that one. The one on the right."*

or

U1 : *"I think maybe this one which one for me it's like the closer one to you, closer one to me, this one or this one, like closer one left one."*

Another strategy involved referencing elements within the rooms, which was employed during the *Solid Blending* and *Transparent Blending* conditions when participants could see the surroundings of both rooms. This aligns with findings by Muller et al. [55]. For example:

U2 : *"The first one is the back computer. And it's near the bookshelf. And the third one is in the cupboard here."*

However, this strategy was not used during the *Partial Blending* condition, as participants couldn't see the same objects in the remote user's room, as shown below.

U1 : *"Near the computer."*

U2 : *"Where is the computer?"*

U1 : *"The computer is at the back. Turn around. Turn around."*

U2 : *"Yeah, this one?"*

U1 : *"No, no, computer. Can you see the computer? Oh, no, We have different items in the room. So it's, I don't know how to describe."*

Overall, participants typically employed a *Divide and Conquer* strategy, where they searched for the Rubik's cubes independently and then confirmed their findings afterward, as observed in the videos. They validated the results by having one person check the colors on the task cubes while the other verified them against the answer cubes. In some cases, dyads memorized different sides of the task cubes to conduct a more thorough search for the matching cubes.

In addition to measuring total distance traveled and median avatar head rotation, the head avatar positions were visualized using heat maps and navigation paths to identify patterns or common movement behaviors among participants. The maps were created by overlaying processed data on a top-down view of the rooms. Since the positions were recorded from the HMD, the paths appeared to intersect with room elements when participants bent their heads over objects. This makes it difficult to conclude whether participants intentionally avoided or walked through virtual objects based solely on the navigation path plots (Fig.4.6 and Fig.4.7). However, the heat maps provided insights into how often participants visited specific areas, showing that most tended to navigate in clear spaces and generally avoided only physical objects (Fig.4.8 and Fig.4.9). Additionally, video recordings revealed that participants often dodged each other during the trials, despite interacting only as virtual avatars without physical collision detection.

4.3 Relationships Between Measurement

In this section, it reports on notable relations between variables that were analyzed by a Spearman's rank-order correlation test per conditions. Its purpose is to observe the connection based on the prior prediction, and evaluate final results. The graph from Fig.4.11 presents the correlation coefficient ranging from -1 to 1. A value of +1 indicates a perfect correlation of ranks, and a value of +1 indicates a perfect correlation of rank. On the other hands, a value of 0 means there is no correlation between ranks. The values of talkativeness are averaged with the survey responses related participants' and partners' scores, while other variables are averaged by participants' own scores.

4.3.1 Avatar Speed and Yaw Speed

Prior to the study, the null hypothesis posited that there would be a positive correlation between the two variables, based on the expectation that participants would have the same change on speed and vertical orientation. The results indicate a significant, strong positive correlation across all conditions. For the *Solid Blending* condition, a notable correlation was observed ($r_s(2) = .99$, $p = .99$). Strong evidence of a high positive correlation was also found in the *Transparent Blending* condition ($r_s(2) = .99$, $p = .99$). Similarly, the *Partial Blending* condition demonstrated a high correlation according to the results ($r_s(2) = .99$, $p = .99$).

4.3.2 Talkativeness and Number of Spoken Words

Prior to the study, it was assumed, as null hypothesis, that these two variables would have a positive correlation, since they are relevant to mutual communication. The results indicate a significant, strong positive correlation across all conditions. For the *Solid Blending* condition, a notable correlation was observed ($r_s(4) = .739$, $p = .094$). Strong evidence of a high positive correlation was also found in the *Transparent Blending* condition ($r_s(4) = .728$, $p = .101$). Similarly, the *Partial Blending* condition demonstrated a high correlation according to the results ($r_s(4) = .788$, $p = .063$).

4.3.3 Distance Traveled and Completion Time

Prior to the study, the null hypothesis posited that there would be a positive correlation between the two variables, based on the expectation that participants would spend time mostly navigating within the collaborative space. The results indicate a significant, strong positive correlation in *Solid Blending* condition ($r_s(4) = .956$, $p = .003$) and *Transparent Blending* condition ($r_s(4) = .837$, $p = .038$); However, it did not apply for *Partial Blending* condition, which there is no statistically significant support the strong correlation ($r_s(4) = .359$, $p = .485$), which might be effected from less virtual objects.

4.3.4 Cooperation and Number of Spoken Words

Prior to the study, it was assumed, as null hypothesis, that these two variables would have a positive correlation, since they are relevant to mutual interaction. The results indicate a significant, strong positive correlation across all conditions. For the *Solid Blending* condition, a notable correlation was observed ($r_s(4) = .598$, $p = .210$). Strong evidence of a high positive correlation was also found in the *Transparent Blending* condition ($r_s(4) =$

.606, $p = .202$). Similarly, the *Partial Blending* condition demonstrated a high correlation according to the results ($r_s(4) = .956$, $p = .003$).

4.3.5 Cooperation and Avatar Speed

The results indicate a significant, strong negative correlation across all conditions. For the *Solid Blending* condition, a notable correlation was observed ($r_s(6) = .500$, $p = .207$). Strong evidence of a high negative correlation was also found in the *Transparent Blending* condition ($r_s(6) = .503$, $p = .204$). Similarly, the *Partial Blending* condition demonstrated a high correlation according to the results ($r_s(6) = .643$, $p = .086$). This statement contrasts to null hypothesis that there will be a strong positive correlation since users would travel faster when they had more cooperative score.

4.3.6 Avatar Speed and Completion Time

Prior to the study, the null hypothesis posited that there would be a negative correlation between the two variables, based on the expectation that participants would complete the tasks faster when they navigated faster. The results indicate a significant, moderate negative correlation across all conditions. For the *Solid Blending* condition, a notable correlation was observed ($r_s(4) = .598$, $p = .210$). Strong evidence of a moderate negative correlation was also found in the *Transparent Blending* condition ($r_s(4) = .598$, $p = .210$). Similarly, the *Partial Blending* condition demonstrated a medium correlation according to the results ($r_s(4) = .478$, $p = .338$).

4.3.7 Talkativeness and Completion Time

The results indicate a significant, strong positive correlation across all conditions. For the *Solid Blending* condition, a notable correlation was observed ($r_s(4) = .739$, $p = .094$). Strong evidence of a high positive correlation was also found in the *Transparent*

Blending condition ($r_s(4) = .728$, $p = .101$). Similarly, the *Partial Blending* condition demonstrated a high correlation according to the results ($r_s(4) = .788$, $p = .063$). This statement contrasts to null hypothesis that there will be a negative positive correlation across all conditions since users would complete the tasks faster when they have more talkativeness score.

4.3.8 Yaw Speed and Distance Traveled

The results indicate a significant, strong positive correlation across all conditions. For the *Solid Blending* condition, a notable correlation was observed ($r_s(2) = .8$, $p = .2$). Strong evidence of a high positive correlation was also found in the *Partial Blending* condition ($r_s(2) = .8$, $p = .2$). The *Transparent Blending* condition demonstrated a high correlation according to the results ($r_s(2) = .6$, $p = .4$). It agrees with null hypothesis based on prediction that participant would turn their head multiple times if they walked for a long distance.

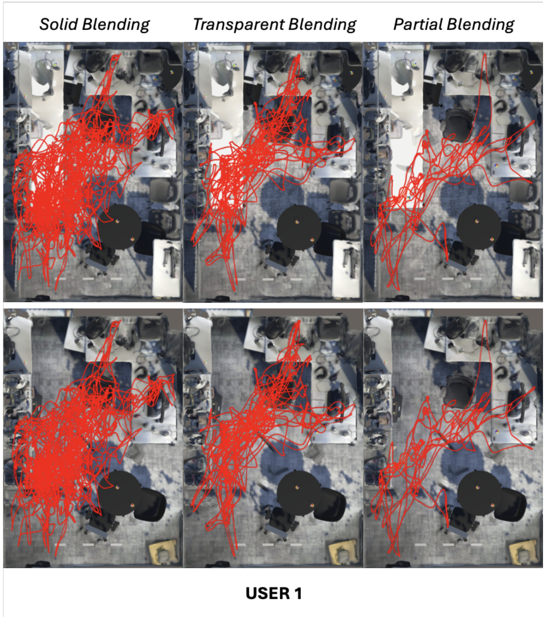


Figure 4.6: Navigation path of user 1.

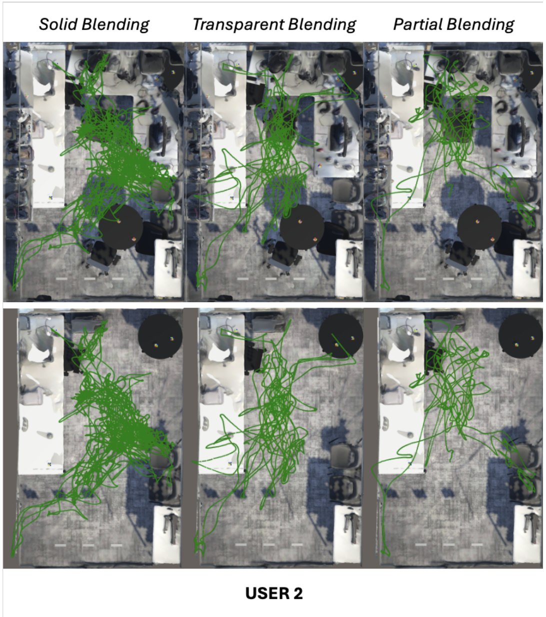


Figure 4.7: Navigation path of user 2.

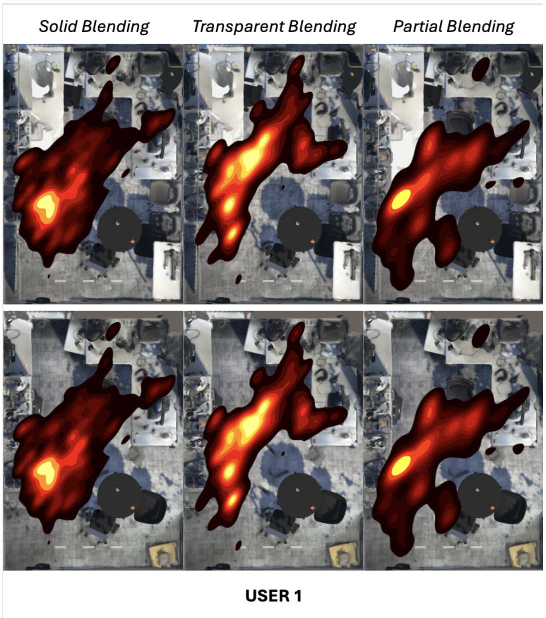


Figure 4.8: Heat map of user 1.

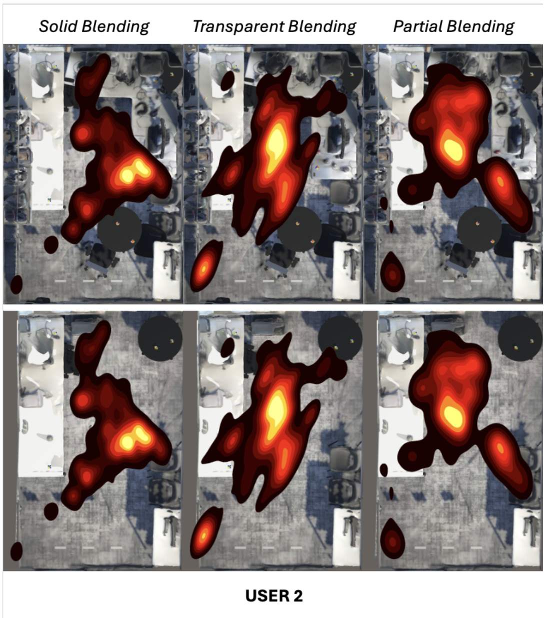


Figure 4.9: Heat map of user 2.

Figure 4.10: Top row: Navigation paths of head position of participants in the fourth trial. The left image is user 1, and the right image is user 2. Bottom row: Heat maps of head position of participants in the fourth trial. The left image is user 1, and the right image is user 2.

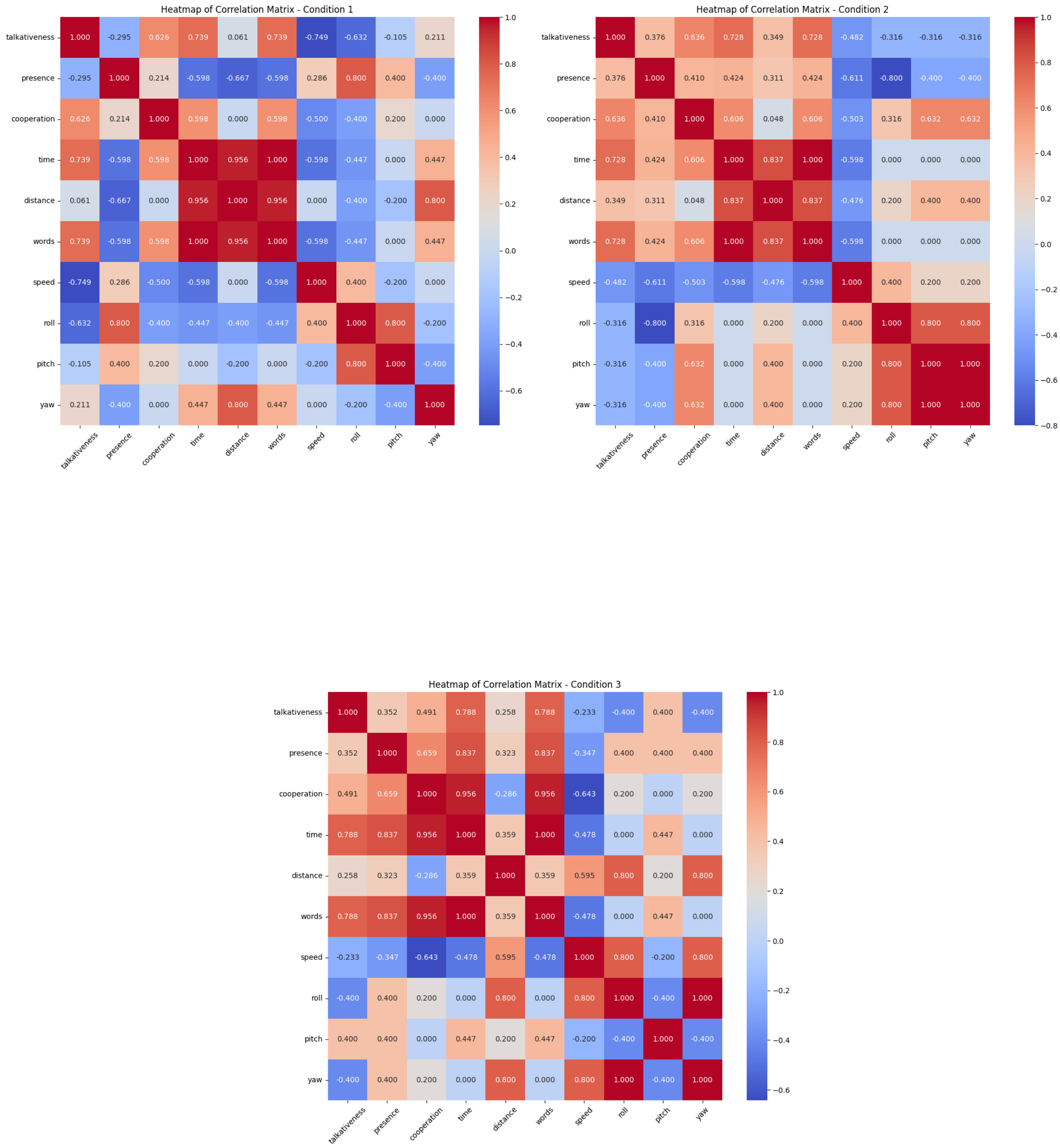


Figure 4.11: The heat map of correlation coefficient visualizing the relationship between measurement

5 | Discussion

The purpose of this chapter is to discuss the key findings of the study linking the results to theoretical frameworks and discussing the study's significance in the field. The feedback received from the participants are considered to evaluate the discussion. Additionally, limitations and suggestions for future research are also highlighted.

5.1 Limitations

In the experiment, participants were not allowed to move either physical or virtual objects, except for the number signs. This limitation reduced the level of interaction in the MR collaborative scene, as users were unable to engage with the remote environment virtually. This constraint arose from the absence of an object tracking algorithm capable of determining the real-time positions of objects, which led to the need for a static virtual environment throughout the study. Additionally, the mesh used in this experiment was generated from pre-scanning and could not be updated in real time. The experimental rooms were also used by multiple individuals, and the study took place across different days. Consequently, the placement of physical objects, excluding the Rubik's cubes, varied between trials. The Rubik's cubes were positioned based on approximate locations in the virtual environment, but without accurate object tracking, the physical positioning could differ from the intended virtual coordinates.

The analysis in Section 4 revealed that participants (N=8) traveled the least distance, completed the task in the shortest time, and spoke the fewest words in the *Partial Condition* compared to the other conditions. This outcome may have been influenced by a learning bias, as participants became more familiar with the task as the trials progressed. The study design had participants start with the *Solid Blending* condition, followed by *Transparent Blending*, and finishing with *Partial Blending*. Initially, participants were unfamiliar with the task, which slowed their progress. However, after completing the first trial, they understood the task better and were able to perform it more quickly. This learning effect significantly impacted the numerical data collected from the HMD and video recordings, introducing potential bias into the analysis.

5.2 Participants' Feedback

After completing all trials, an unstructured interviews were conducted to perceive the participants' interaction more extensively, and review their feedback. To sum up, in *Solid Blending* condition, participants could not approach to some physical Rubik's cubes due to occlusion with virtual objects. Also, some participants confused with virtual objects based from the interview and questions in presence questionnaire (Fig.4.4). Participants during *Transparent Blending* condition had less confusion between virtual and physical objects; however, due to rendering problem, participants perceived that the virtual objects was not transparent in some angles, which was similar to *Solid Blending*.In the *Partial Blending* condition, participants felt they performed the task more effectively compared to other conditions because the cubes were easier to locate; however, they lacked any connection to another room.

The results indicate that all participants completed the tasks most quickly in the *Partial Blending* condition, followed by the *Transparent Blending* condition, with the longest completion times occurring in the *Solid Blending* condition. This same sequence was also observed in the distances traveled. Based on the interview, seven out of eight participants preferred to perform the task in *Partial Blending* condition, since it had more walking path, and they felt more comfortable to navigate. Based on these statements, it can be assumed that the most effective condition is *Partial Blending* condition.

5.3 Application

The purpose of this study is to find the most effective blended environment for MR remote collaboration. This experiment is benefit to the task where they require two users need to fully access remote environment, for instance, maintenance, assembly, education.

6 | Conclusion

The need of remote collaboration has been increasing recently, and mixed reality has been applied to increase immersive interaction. Previous studies focused on improving user interface and user experience of mixed reality remote collaboration, including a survey comparing functionality domains.

This study is inspired by the works of Fidalgo [9] and Numan [10]. In this experiment, it focuses on merging environments of two users to observe the most efficient method of blending or the approach to access remote user's room, which consists of *Solid Blending* which virtual objects are rendered similar to the physical objects, *Transparent Blending* which virtual objects are rendered to be transparent, and *Partial Blending* which merges the virtual objects relevant to tasks only. Participants were asked to perform tasks by finding matching Rubik's cubes, and the data were recorded by HMD messaging position and orientation through Ubiq, and screen recording.

The null hypotheses state that there is no difference in user interaction and communication across the conditions. Interaction encompasses talkativeness, presence, cooperativeness, number of spoken words, completion time, distance traveled, avatar movement speed, and avatar rotation speed. According to the results in Table 4.1, there is no significant evidence to reject the null hypotheses for pitch speeds, presence, number of spoken words, and completion time. On the other hands, there are statistically significant differences in cooperativeness, avatars' movement speed, distance traveled and roll speed between conditions. There is also evidence supporting significant differences in participants' perceptions of talkativeness and cooperation between the *Solid Blending* and *Partial Blending* conditions. Additionally, the evidence does not support rejecting the null hypotheses on talkativeness, presence, and cooperation between *Solid Blending* and *Transparent Blending*.

To summarize, across eight aspects of interaction, significant evidence exists to reject the null hypotheses between the *Solid Blending* and *Transparent Blending* conditions, specifically for avatar movement speed and roll speed. For the *Transparent Blending* and *Partial Blending* conditions, significant differences have been confirmed in avatar movement speed, yaw speed, and talkativeness, aligning closely with the initial findings. Lastly, the null hypotheses concerning the *Solid Blending* and *Partial Blending* conditions are predominantly rejected based on statistical tests, which include avatar movement speed, roll speed, distance traveled, talkativeness, and cooperation. Based on task performance, questionnaire, and interview, it can be assumed that the most effective condition for participants is *Partial Blending* condition.

6.1 Future Works

The results and analysis were derived from 4 dyads, or 8 participants, which is insufficient to draw definitive conclusions. To accurately analyze behavior, a sample size of 10-20 participants would be more appropriate. This section provides recommendations for future studies that can build upon the current findings.

From the previous section (Section 5.1), it mentions that this study cannot involve with changing the positions of objects inside experimental space because it was not implemented with real-time object tracking system. Previous works applied the systems with panorama camera, or 360 video cameras to present the environment of remote users in real time. [2, 22, 23, 24, 25, 27, 28, 33, 56] By applying these methods to the experiment, it could extend the interaction with the virtual environment, and increase accessibility of users to remote environment.

Slater et al. [8] investigated user behavior under different rendering techniques across three conditions. The conventional condition presented virtual meshes closely resembling reality. The stylized condition emphasized edge visibility, while the virtualized condition lightened edges and removed mesh colors. This study draws inspiration from their work to compare interactions between Solid Blending and Transparent Blending. Currently, three conditions are being used to examine user behavior in MR remote collaboration. In the future, additional conditions could be introduced, such as glowing edges on virtual meshes or an added color layer inspired by holographic effects seen in sci-fi movies.

The post-study questionnaire should be administered during each condition change to capture participants' real-time feelings, as suggested in the feedback. Participants noted that recalling their experiences and interactions accurately was difficult after completing all trials. Initially, the questionnaire was scheduled to be given only after all trials were completed. This design choice was intentional, as the study primarily focused on analyzing the spatial frequency of the user's head avatar, as well as measuring distance traveled and average speed, estimated from the sensors on the MR HMD. The questionnaire was created using Microsoft Forms [57], which participants had to access via a PC or mobile device. This process required them to remove the HMD, potentially impacting the accuracy of the measurements. To address this issue, it is recommended to integrate the questionnaire directly into the MR environment, displaying the user interface within the collaboration space immediately after each trial. This would allow participants to express their feelings simultaneously without interrupting the flow of the study or affecting the measurements.

Lastly, to better examine talkativeness and mutual interaction between users, the communication task should be modified. Initially, both users could see the virtual cubes (task cubes). During the trials, some dyads strategized by searching for matching cubes independently, which reduced communication between participants. It is recommended that the task cubes be visible to only one participant, encouraging more detailed information exchange and increasing talkativeness.

References

[1] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE TRANSACTIONS on Information and Systems*, vol. E77-D, no. 12, pp. 1321–1329, 1994.

[2] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock, "Mixed reality remote collaboration combining 360 video and 3d reconstruction," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, (New York, NY, USA), p. 1–14, Association for Computing Machinery, 2019.

[3] R. Cools, A. Esteves, and A. L. Simeone, "Blending spaces: Cross-reality interaction techniques for object transitions between distinct virtual and augmented realities," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 528–537, 2022.

[4] "Meta quest 3: New mixed reality vr headset." Meta. Retrieved July 27, 2024 from https://www.meta.com/gb/quest/quest-3/.

[5] C. Gutwin and S. Greenberg, "Workspace awareness for groupware," in *Conference Companion on Human Factors in Computing Systems*, CHI '96, (New York, NY, USA), p. 208–209, Association for Computing Machinery, 1996.

[6] J. E. S. Grønbæk, K. Pfeuffer, E. Velloso, M. Astrup, M. I. S. Pedersen, M. Kjær, G. Leiva, and H. Gellersen, "Partially blended realities: Aligning dissimilar spaces for distributed mixed reality meetings," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, (New York, NY, USA), Association for Computing Machinery, 2023.

[7] J. Herskovitz, Y. F. Cheng, A. Guo, A. P. Sample, and M. Nebeling, "Xspace: An augmented reality toolkit for enabling spatially-aware distributed collaboration,"

Proc. ACM Hum.-Comput. Interact., vol. 6, nov 2022.

[8] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder, "Small-Group Behavior in a Virtual and Real Environment: A Comparative Study," *Presence: Teleoperators and Virtual Environments*, vol. 9, pp. 37–51, 02 2000.

[9] C. G. Fidalgo, Y. Yan, H. Cho, M. Sousa, D. Lindlbauer, and J. Jorge, "A survey on remote assistance and training in mixed reality environments," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2291–2303, 2023.

[10] N. Numan and A. Steed, "Exploring user behaviour in asymmetric collaborative mixed reality," in *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, VRST '22, (New York, NY, USA), Association for Computing Machinery, 2022.

[11] C. Pidel and P. Ackermann, "Collaboration in virtual and augmented reality: A systematic overview," in *Augmented Reality, Virtual Reality, and Computer Graphics* (L. T. De Paolis and P. Bourdot, eds.), (Cham), pp. 141–156, Springer International Publishing, 2020.

[12] I. S. MacKenzie, "Human-computer interaction: An empirical research perspective," 2024.

[13] Microsoft, "What is mixed reality?." https://learn.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality, January 2023.

[14] R. Druta, C. Druta, P. Negirla, and I. Silea, "A review on methods and systems for remote collaboration," *Applied Sciences*, vol. 11, no. 21, 2021.

[15] P. Wang, X. Bai, M. Billinghurst, S. Zhang, X. Zhang, S. Wang, W. He, Y. Yan, and H. Ji, "Ar/mr remote collaboration on physical tasks: A review," *Robotics and Computer-Integrated Manufacturing*, vol. 72, p. 102071, 2021.

[16] A. Schäfer, G. Reis, and D. Stricker, "A survey on synchronous augmented, virtual, and mixed reality remote collaboration systems," *ACM Comput. Surv.*, vol. 55, dec 2022.

[17] T. Piumsomboon, A. Day, B. Ens, Y. Lee, G. Lee, and M. Billinghurst, "Exploring enhancements for remote mixed reality collaboration," in *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, SA '17, (New York, NY, USA), Association for Computing Machinery, 2017.

[18] MeetinVR ApS., "Meetinvr," 2021. Available at: https://www.meetinvr.com/.

[19] VRChat Inc., "Vrchat," 2014. Available at: https://hello.vrchat.com.

[20] Spatial Systems, Inc., "Spatial," 2022. Available at: https://www.spatial.io.

[21] D. I. Fink, J. Zagermann, H. Reiterer, and H.-C. Jetter, "Re-locations: Augmenting personal and shared workspaces to support remote collaboration in incongruent spaces," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, nov 2022.

[22] M. Adcock, S. Anderson, and B. Thomas, "Remotefusion: real time depth camera fusion for remote collaboration on physical tasks," in *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '13, (New York, NY, USA), p. 235–242, Association for Computing Machinery, 2013.

[23] A. Dai, M. Nießner, M. Zollhöfer, S. Izadi, and C. Theobalt, "Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration," *ACM Trans. Graph.*, vol. 36, jul 2017.

[24] S. Izadi, R. A. Newcombe, D. Kim, O. Hilliges, D. Molyneaux, S. Hodges, P. Kohli, J. Shotton, A. J. Davison, and A. Fitzgibbon, "Kinectfusion: real-time dynamic 3d surface reconstruction and interaction," in *ACM SIGGRAPH 2011 Talks*, SIGGRAPH '11, (New York, NY, USA), Association for Computing Machinery, 2011.

[25] M. Joachimczak, J. Liu, and H. Ando, "Real-time mixed-reality telepresence via 3d reconstruction with hololens and commodity depth sensors," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ICMI '17, (New York, NY, USA), p. 514–515, Association for Computing Machinery, 2017.

[26] B. Nuernberger, M. Turk, and T. Höllerer, "Evaluating snapping-to-photos virtual travel interfaces for 3d reconstructed visual reality," in *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, (New York, NY, USA), Association for Computing Machinery, 2017.

[27] H. Tian, G. A. Lee, H. Bai, and M. Billinghurst, "Using virtual replicas to improve mixed reality remote collaboration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2785–2795, 2023.

[28] T. Teo, G. A. Lee, M. Billinghurst, and M. Adcock, "Merging live and static 360 panoramas inside a 3d scene for mixed reality remote collaboration," in *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 22–25, 2019.

[29] W. Steptoe, S. Julier, and A. Steed, "Presence and discernability in conventional and non-photorealistic immersive augmented reality," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 213–218, 2014.

[30] S. BENFORD, J. BOWERS, L. E. FAHLÉN, and C. GREENHALGH, *MANAGING MUTUAL AWARENESS IN COLLABORATIVE VIRTUAL ENVIRONMENTS*, pp. 223–236.

[31] P. Dourish and V. Bellotti, "Awareness and coordination in shared workspaces," in *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*, CSCW '92, (New York, NY, USA), p. 107–114, Association for Computing Machinery, 1992.

[32] A.-S. Axelsson, r. Abelin, I. Heldal, R. Schroeder, and J. Wideström, "Cubes in the cube: A comparison of a puzzle-solving task in a virtual and a real environment," *CyberPsychology & Behavior*, vol. 4, no. 2, pp. 279–286, 2001. PMID: 11710253.

[33] T. Teo, A. F. Hayati, G. A. Lee, M. Billinghurst, and M. Adcock, "A technique for mixed reality remote collaboration using 360 panoramas in 3d reconstructed scenes," in *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, VRST '19, (New York, NY, USA), Association for Computing Machinery, 2019.

[34] A. Steed, M. Slater, A. Sadagic, A. Bullock, and J. Tromp, "Leadership and collaboration in shared virtual environments," in *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pp. 112–115, 1999.

[35] "Meta quest 2: Immersive all-in-one vr headset." Meta. Retrieved July 27, 2024 from https://www.meta.com/gb/quest/products/quest-2/.

[36] S. Perera, D. Barnes, and D. Zelinsky, *Exploration: Simultaneous Localization and Mapping (SLAM)*, pp. 268–275. Boston, MA: Springer US, 2014.

[37] Unity Technologies, "Unity real-time development platform | 3d, 2d, vr & ar engine," 2005. Available at: https://unity.com.

[38] Oculus, "Meta mr utility kit." https://assetstore.unity.com/packages/tools/integration/meta-mr-utility-kit-272450. Accessed: August 1, 2024.

[39] S. J. Friston, B. J. Congdon, D. Swapp, L. Izzouzi, K. Brandstätter, D. Archer, O. Olkkonen, F. J. Thiel, and A. Steed, "Ubiq: A system to build flexible social virtual reality experiences," in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, VRST '21, (New York, NY, USA), Association for Computing Machinery, 2021.

[40] Unity Technologies Inc., “Xr interaction toolkit.” https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@3.0/manual/index.html. Accessed: 2024-08-02.

[41] Meta, “Mixed reality health and safety guidelines.” https://developer.oculus.com/resources/mr-health-safety-guideline/. Accessed: 2024-08-02.

[42] Oculus Samples, “Unity depthapi.” https://github.com/oculus-samples/Unity-DepthAPI. Accessed: 2024-08-01.

[43] Meta, “Spatial anchors overview.” https://developer.oculus.com/documentation/unity/unity-spatial-anchors-overview/?locale=en_GB. Accessed: 2024-08-02.

[44] Black Whale Studio - XR Tutorials, “Meta quest spatial anchors | create, save, erase & load spatial anchors,” February 2024. Available at: https://www.youtube.com/watch?v=vaAQvdGYeak.

[45] “Map controllers.” Meta. Retrieved Aug 10, 2024 from https://developer.oculus.com/documentation/unity/unity-ovrinput/?locale=en_GB.

[46] W. Armstrong, “Guid based reference.” https://github.com/Unity-Technologies/guid-based-reference, 2018.

[47] “Playerprefs.” Unity Technologies. Retrieved Aug 10, 2024 from https://docs.unity3d.com/ScriptReference/PlayerPrefs.html.

[48] Niantic, Inc., “Scaniverse - 3d scanner,” 2021. Available at: http://www.blender.org.

[49] Blender Online Community, “Blender - a 3d modelling and rendering package,” 2021. Available at: http://www.blender.org.

[50] CGTrader, "Armchair beso free 3d model," 2021. Available at: https://www.cgtrader.com/products/armchair-beso.

[51] CGTrader, "Tanya - office chair with armrests with castors - free 3d model," 2021. Available at: https://www.cgtrader.com/products/tanya-02be798f-a013-48c1-9130-9934ecdba5a8.

[52] L. Bailey, "Open broadcaster software," 2012. Available at: https://obsproject.com.

[53] Y. Pan, D. Sinclair, and K. Mitchell, "Empowerment and embodiment for collaborative mixed reality systems," *Computer Animation and Virtual Worlds*, vol. 29, no. 3-4, p. e1838, 2018. e1838 cav.1838.

[54] Clipto Inc., "Clipto," 2024. Available at: https://www.clipto.com/.

[55] J. Müller, R. Rädle, and H. Reiterer, "Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, (New York, NY, USA), p. 6481–6486, Association for Computing Machinery, 2017.

[56] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "Mixed reality collaboration through sharing a live panorama," in *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, SA '17, (New York, NY, USA), Association for Computing Machinery, 2017.

[57] "Microsoft forms | surveys, polls, and quizzes." Microsoft. Retrieved Aug 26, 2024 from https://www.microsoft.com/en-us/microsoft-365/online-surveys-polls-quizzes.

A | Source Code

Source code for all of the methods implemented in Chap. 3 for the project can be found in the GitHub repository:

`https://shorturl.at/KjdSA`.