CARL- COLLABORATIVE AUGMENTED REALITY FOR LABLINKING

A Thesis
Presented to
The Academic Faculty

By

Asmus Eilks

In partial fulfillment
of the requirements for the Degree
Master of Science in the
School of Computer Science
Cognitive Systems Lab

University of Bremen

November 2022

CARL- COLLABORATIVE AUGMENTED REALITY FOR LABLINKING

Thesis committee:

Dr. Felix Putze
Cognitive Systems Lab
University of Bremen

Prof. Dr. Tanja Schultz
Cognitive Systems Lab
University of Bremen

Date approved:

I would like the world to remember me as the guy who really enjoyed playing games and sharing his knowledge and his fun pastimes with everybody else.

Gary Gygax

ACKNOWLEDGMENTS

I would like to thank the following people, without whom this Thesis would not have been possible.

My supervisor Felix Putze for his scientific and technical guidance

My girlfriend Lily Meister for helping me test the application and proofreading my thesis, as well as always being there to lend an ear to my troubles.

My parents, for their infinite patience and support.

The University of Bremen and specifically the Cognitive Systems Lab for providing the technical and research infrastructure required for this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

AR Augmented Reality,A virtual reality environment which is overlaid onto and incorporates the real environment. vi, ix, 5, 8–10, 19, 29, 30

EASE Everyday Activity Science and Engineering, see section 2.1. 2, 4, 12, 39

EEG Electroencephalography, A way of measuring brain activity through electrical sensors on the skin of the scalp. 11

HL Shorthand for HoloLens, a head mounted display for augmented reality applications. 26, 27

HMD Head Mounted Display, A device worn on the users head to display virtual objects to the user. 9

LLL LabLinkingLevel, See subsection 2.1.1. 4, 5, 38

OT Shorthand for Optitrack, a motion capture system. 26, 27

UDP User Datagram Protocol, a connectionless protocol to send data across the Internet. 14

UI User Interface, a digital interface to interact with a program.. vi, 15, 19

VR Virtual Reality, In this Thesis, the use of a Head Mounted Display and 3D-spatially tracked controllers to control an application. vi, ix, 5, 6, 10, 29, 30, 39

GLOSSARY

Digital Twin A close digital representation or copy of a real object. 12

Unity A game engine used to develop the application in this thesis. vi, 13, 14, 19–21, 38

CHAPTER 1

INTRODUCTION

1.1 Motivation

Much research is dedicated to teaching robots how to do everyday activity tasks, such as setting a kitchen table, cooking, or cleaning. One of the primary approaches to accomplish this is the use of machine learning through neural networks trained on human activity data. To successfully do so, large amounts of precise human behavioral data are needed, for example, Hand-Tracking or Motion-Capture recordings [22]. As the robots will often have to interact with changing environments or other human agents in the vicinity, it is not enough to simply record one human doing a task, as the data collected would not be robust against interference from other agents. Instead, data-collection experiments with several participants collaborating on the same task and interacting in the same environment are necessary. However setting up such experiments faces several problems, such as locally available equipment, travel costs and time, or ongoing covid restrictions. These may be circumvented effectively by instead conducting the experiment in a shared virtual space, but it is unclear whether the interaction options offered by virtual objects are sufficient to gather useful data.

In addition, using a shared virtual space allows researchers to simulate different sensory challenges or differences in perception. Users may or may not be able to hear each other, or have only certain parts of their body projected into the virtual space. By adjusting and studying these parameters, models for different types of robots (for example, robots unable to output speech for any reason, or robots with only limited visual perception), and their efficacy, can be researched.

One example of a research project in this field is EASE-CRC (Everyday Activity Science and Engineering Collaborative Research Center), an interdisciplinary research center at the University of Bremen. Its stated primary goal is ”to advance our understanding of how human-scale manipulation tasks can be mastered by robotic agents” [8], in other words, to develop and program robots that can perform complex human tasks of everyday life. Currently, this project lacks the capability to conduct said experiments using an augmented reality virtual space, which this thesis will aim to provide.

1.2 Goal and Approach

The goal of this Thesis is therefore to develop and test an application that allows conducting connected EASE-Experiments through the use of a shared virtual space in augmented reality. This application will be named CARL (Collaborative Augmented Reality for LabLinking). CARL should therefore provide the following features:

- Allow the conducting of various EASE-Experiments (import of custom 3D-Models and development and integration of additional functionality should be easy for the researchers)
- Synchronize virtual and real objects across the connected devices
- Allow for recording and other forms of data collection to the lab streaming layer used in the EASE project[7].

To validate the application's usefulness, a small study will be conducted. This study sees participants perform a table-setting task very similar to the one used in the collection of motion data for machine learning, and will investigate the question of whether pings are a useful tool for communication in collaborative augmented reality.

1.3 Structure of this Thesis

Chapter 2 explains the key concepts for this thesis,

Chapter 3 highlights some of the related work about collaboration in augmented reality,

Chapter 4 details the design considerations for the system, followed by Chapter 5 which explains the implementation of the system, first from an interaction and then a technical perspective.

Chapter 6 elaborates the procedure of the validation study

Chapter 7 lists and interprets the results of the validation study

Chapter 8 summarizes the work and outlines directions for future work and development with this project.

CHAPTER 2
KEY CONCEPTS AND BACKGROUND

2.1 EASE-CRC & LabLinking

EASE-CRC (Everyday Activity Science and Engineering Collaborative Research Center) is an interdisciplinary research center at the University of Bremen. Its stated primary goal is ”to advance our understanding of how human-scale manipulation tasks can be mastered by robotic agents” [8], in other words, to develop and program robots that can perform complex human tasks of everyday live. One of the primary approaches in this effort is the use of machine learning from human motion data. To gather this data, augmented reality devices have proven useful, as they can simultaneously record the participants’ behaviour&vision, track the location of specific objects which are being manipulated, and potentially simulate such objects if they arent available at the experiment’s location.

2.1.1 LabLinking

Another important aspect for this thesis is the concept of Lab-Linking as described by Schultz et al [32]. Lab-Linking herein describes an experimental approach, in which participants and researchers may conduct joint experiments while being spatially distributed. Lab-Linking is split into five Lab-Linking-Levels(LLL), with increasing interconnectedness as the level increases:

- **LLL-1: Coordinated Studies:** Labs share experience, data formats, and other types of information, establishing a common basis for information exchange. However, they still conduct experiments separately.
- **LLL-2: Asynchronous Data Coupling:** Labs conduct the same experiment using different conditions, for example, one lab might use different recording equipment

or a different perspective on the experiment than another. The recorded data is then synchronized afterward.

- **LLL-3:Synchronous bi-directional data coupling:** Recorded data is synchronized in near real-time, and spatially distributed participants conduct the experiment simultaneously. This allows a degree of interaction between participants, for example using voice communication to discuss a task at hand.
- **LLL-4: Immersive synchronous interaction and collaboration:** Technology is used to simulate a co-located experiment to spatially-distributed participants, for example using VR or AR devices, including motion tracking. This allows complex interactions between participants.
- **LLL-5: LabLinking based on future emerging technologies:** With new technologies developing, more in-depth immersion could be possible, for example by haptic simulation.

So far, LLL-4 is still a growing field and not widely established, but it is expected that the higher fidelity of visuals and tracking offered by VR and AR devices will yield to higher immersion and realism when conducting experiments. This thesis aims to develop a system to allow for experiments on LLL-4, through the use of augmented reality glasses and the Internet.

2.2 Consumer Augmented Reality

Augmented Reality describes an environment in which users perceive and/or can interact with real as well as virtual objects. Augmented reality can be achieved by using various different devices, the most common of which is likely smartphones, where real-time camera filters are used with object recognition software to create fun videos, or games like *Pokemon Go* [24] that use GPS-Locations and landmarks as part of their game mechanics. Augmented reality on smartphones has also led to applications for more serious purposes such as education, for example, the app Starwalk 2, which can highlight celestial objects viewed through the smartphone's camera [35]. Another frequently used device is augmented reality glasses, such as the Google-Glass [12] or the Microsoft HoloLens[23]. They are far less prevalent at the time of writing, most likely due to their steep price tag. Glasses-like devices however come with the advantage of leaving the user's hands free for three-dimensional interaction with the virtual and real environments and can cover a wider visual area than the usually small smartphone screens.

2.3 Collaborative Communication

As the primary feature of CARL is remote collaboration, it was important to develop an understanding of the requirements needed to allow for collaboration through software, specifically in augmented reality. An often cited important requirement for effective collaboration is communication. While the most common ways of communication in online games and software are voice and text chats, these have some significant drawbacks, such as requiring a shared language and potentially breaking immersion. Text chats are also often comparatively slow, and require the use of the user's hands, blocking them from interacting with the software at the same time. Innocent & Haynes propose *Symbolchat* as an alternative, a chat window consisting of easy-to-understand contextual symbols, which can also be placed in the virtual world [16].

2.3.1 Gestures

Especially in full-body mirroring environments, another way of communicating is gestures. These are most prevalent in VR-games & Chatrooms, where they are commonly used together with, but sometimes even instead of, voice chat [21]. Some environments have also successfully used natural gestures between users as part of the applications controls, such as fistbumping someone in *RecRoom* [29] to form a virtual group [21]. However, they also note the danger of gestures and other non-verbal behaviors as a potential avenue for harassment.

Even in Multiplayer-Desktop-Games it is not uncommon to have a way to express emotions or feelings through gestures, albeit in the limited form of Emotes, which make the players avatar act out an emotion in a predetermined animation (such as dancing, laughing, or crying). While it's unlikely that emotes are as effective as full facial and motion interaction, they have been shown to have similar effects to communication involving facial-motion capture [25].

2.3.2 Pings

Another non-verbal communication tool that has seen a lot of use, especially in video games, are pings. A ping is a temporary audiovisual notification placed by a player in the virtual world [20]. Pings are used in many popular multiplayer video games, for example *League of Legends* [31, 19], *Deep Rock Galactic* [11] and *Apex Legends* [30].

Apex Legends' ping system has received high praise from media outlets, being called a 'tiny miracle'[18] and 'revolutionary' [3, 13], these articles often highlight the utility of contextual pings as a way of communicating quickly and efficiently, without the need of a voice chat.

Levitt et al. investigated the relation between pings and team performance in League of Legends and found a concave correlation between the number of pings and a team's performance in the game [20]. They conclude that this likely means pings are an effective tool for communicating, except if they are used maliciously, for example by spamming them.

CHAPTER 3

RELATED WORK

3.1 Taxonomies of Collaboration in Augmented Reality

There are various approaches and definitions of collaboration in augmented reality environments, which can differ strongly depending on context. Pidel and Ackermann for example break up collaboration-systems along two axes, *Time* (Synchronous vs Asynchronous) and *Space* (On-Site vs Remote)[26]. The application developed for this thesis will fall into the Synchronous/Remote area of their differentiation. Another differentiation can be found in the work of Wang & Dunston, who categorize AR-systems in general. Their categorization splits collaborative AR systems between Mobile/Stationary and Collocated/Distributed [37]. In this dichotomy, this thesis would be Stationary/Distributed.

3.1.1 Co-Located

One of the earliest collaborative AR-Systems [5], the ”Studierstube” project [33] by Szalavàri et. al., lists six key aspects for their system, used to collaboratively visualize and discuss scientific data in various ways:

1. *Virtuality:* Virtual Objects are visible in the real world and can be interacted with naturally

2. *Augmentation:* Real-world objects can be annotated with virtual information

3. *Multi-User Support:* Users can work together seamlessly. The Studierstube's group argues that augmented reality has a great advantage as a medium here, as normal human interactions can be incorporated into the setup.

4. *Independence:* Each user has an independent viewport and can interact on their own, without requiring the aid of others

5. *Sharing and Individuality:* While the underlying information for the system is shared with all users, the individual representation may differ based on the needs of an individual

6. *Interaction and Interactivity:* Changes to the underlying data in the system are shared and visualized immediately

In 2002 Billinghurst and Kato build upon this, and highlight the related aspect of seamlessness, allowing perception and interaction in a blend of the real and virtual space [5]. They also list several challenges for collaborative AR:

First and foremost, the issue of gaze, namely that the eyes of a user are hidden by the HMDs, which might hinder communication.

Secondly, they mention the visual difference between virtual and real objects, citing limitations in Field of View, Resolution, and color depths of HMDs. While field of view remains a challenge today, color depth and resolution have been vastly improved over time [6].

Thirdly they list the issue of tracking, citing issues of delay and the tracking only working while markers are within the user's viewport. They hypothesize that hybrid-tracking techniques such as the ones used in this thesis may be particularly fruitful in solving this issue.

3.1.2 Asymmetric Remote Collaboration

A common approach for collaboration involving AR is asymmetric collaboration, where one user acts as a remote helper for another that interacts with the real world. While all of these systems struggle with the ”Independence” aspect, since the worker and helper have different strictly defined roles and interaction possibilities, they can still yield valuable insights into other aspects of AR-Collaboration.

One example of this is *In Touch with the Remote World* [10] by Gauglitz et al. in which they explore remote navigation using an AR headset. In their setup, the worker is wearing an AR headset and is being guided by the helper, who is annotating the scene from a 2D-Touchpad. They conclude that this approach works better than previous research using mouse-based interfaces for a similar purpose. Similarly, RemoteFusion [1] renders the 3D-Scene for the remote helper onto a touchscreen, who then makes annotations that are displayed to the worker using a 2D-Projector.

Another approach is to invert the hardware setup of the helper and the worker, as was done for example by Wang et. al. [36] and Tecchia et al. [34], in their systems, the helper is wearing an AR or VR headset, and sees a projection of the worker's space and actions. The helper's hands are captured and their perspective is shown to the worker on a 2D-Screen/via a projector near the workspace. Wang et. al. concluded that this approach was significantly faster and more accurate than pointer-based guidance systems.

Even closer to a truly co-located space is the CoVAR application developed by Piumsomboon et al [28]. They implemented a shared interaction space between a user in augmented and one in virtual reality. Their design featured various novel technologies, such as embodiment cues to show another user's FoV/Gaze during use. In a later study, they verified that these cues could be useful to lessen a user's task load and ease collaboration [27].

3.2 AR and VR for Everyday Activity Research

There are also various other projects using virtual reality to investigate everyday activities, showing that this approach is fruitful. Haidu and Beetz have for example developed AMEvA, a VR kitchen to collect data for their KnowRob system. They use a detailed physics simulation as well as hand-tracking controllers to gather optimal data [15]. They have also shown that with this data, a simulated robotic agent can perform the same task that was conducted in the experiment.

Bates et. al. developed a similar environment to automatically recognize human behaviors and behavior chains when conducting complex tasks [2]. They note some difficulties with the physics and force simulation that result from purely virtual objects but still manage to accurately recognize the intended behavior in 92% of their dataset.

Promising findings have also been made for the recording of biodata in scenes displayed in augmented reality. Krugliak & Clarke [17] investigated the face inversion effect, a well-understood cognitive response, finding similar EEG data in three different tasks: Viewing faces & inverted faces on a computer screen, walking through an environment seeing real photographs of said faces, and walking through an environment with the faces being displayed in augmented reality. They found similar EEG measurements in the second and third task, and conclude that augmented reality devices offer EEG data comparable to real-life visual stimuli, while offering experimenters greater control over what is visible in the scene.

All in all, it becomes apparent that virtual and augmented reality devices have shown significant potential for research about everyday activities, and are also used in some capacity to facilitate cooperation, but no framework has successfully combined and investigated these aspects so far.

CHAPTER 4
DESIGN

4.1 Experimental Design

As CARL's primary design goal was to closely simulate co-located EASE-experiments, the experiment used for validation was taken from the EASE-context as well. In the experiment, two participants are given the task to set a table together in augmented reality. To do so, participants have access to virtual objects, only present in augmented reality, and real objects with a Digital Twin at the same location. This real/virtual object difference could then be analyzed to validate or further research the interaction and interaction differences. To have a measure of accuracy for the task, the table-setting will not be free form, but instead feature goal positions shown to the participants.

The interaction used to do so is aimed to be as close to real object interaction as possible. As such, virtual objects should be grabbed with the user's actual hands as if grabbing a real object and moved by moving the grabbing hand. Real objects can be interacted with by moving the real object, the Digital Twin should follow the motion smoothly.

As the validation study in this thesis focuses on nonverbal communication, no verbal communication should be included, as it would likely overshadow any ongoing nonverbal cues and make them more difficult to analyze. Instead, users will be tasked to communicate solely by using pings, hand gestures, and head motion. To further facilitate communication, each participant will only be shown half of the goals-positions, which they cannot fulfill on their own given the objects available to them. This should lead to natural communication about the placement of the objects.

4.2 Technical requirements

This means that the underlying technical platform needs to be capable of the following:

- Participants need to be able to move and rotate virtual objects.
- The platform needs to track specific real objects and represent their virtual twins.
- The platform needs to display the participants' locations and movements and visualize them to the other participants.
- Participants need to be able to see the goal positions.
- Participants need to be able to ping, and this ping needs to be able to grab the other participants' attention.
- The platform needs to be able to record data and stream it to the lab streaming layer.
- The framework should offer easy extensibility and usability for other researchers who may want to use/improve on the existing system.

4.3 System Architecture

In order to fulfill these requirements, I have chosen to combine three different systems (see fig. 4.1):
HoloLens-2 devices will be used to display the virtual world to the participants. They also feature built-in libraries for Hand-Tracking and moving virtual objects with natural motion, and are able to track their own location in space.
An Optitrack-system will be used to track specific real objects and their movement in space.
A server running on a Desktop-PC in the Unity-Editor will synchronize the data between the HoloLenses and the Optitrack-system, and provide some administrative functions, like spawning objects and a birds-eye-view.

The devices all run Unity-Applications and communicate through Netcode for Unity, a UDP-based framework to implement multiplayer games in Unity.

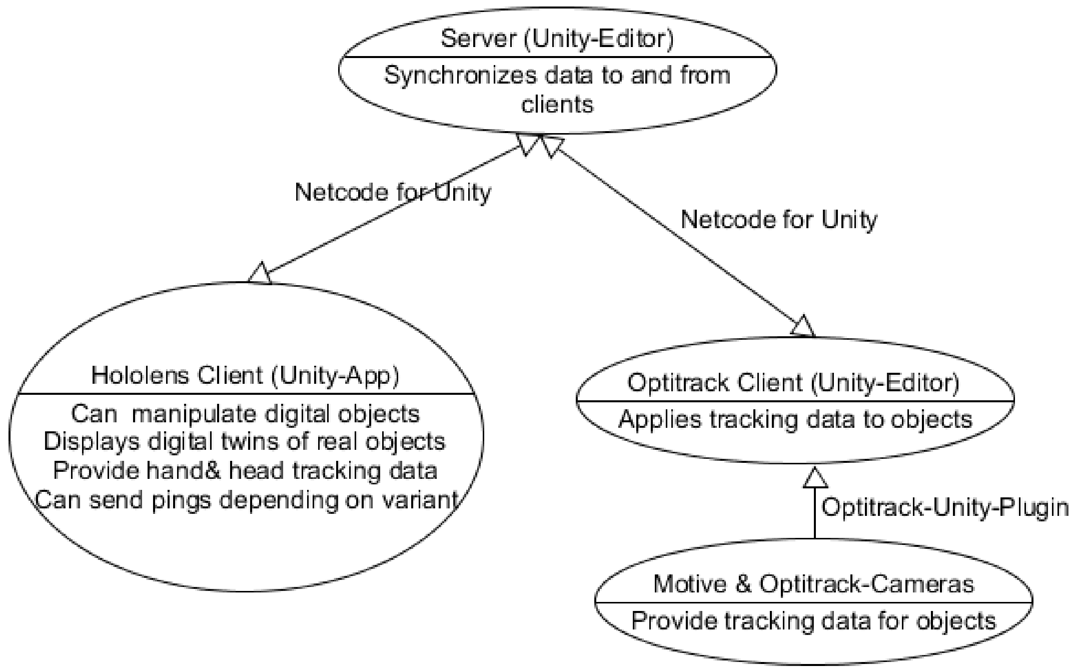

Figure 4.1: An overview of the Network structure/responsibilities

CHAPTER 5
IMPLEMENTATION

5.1 User Interaction and UI in the HoloLens-Application

The interaction for participants and direct users of CARL happens primarily through the Microsoft-HoloLens-2 device, using Microsoft's existing libraries. Objects can be grabbed by making a grabbing or pinching gesture while the hand is near or in the object, or alternatively grabbed from afar by pointing at the object and grabbing or pinching. In order to convey to the user when an object is grabbed, the object is highlighted in red while held (see fig. 5.1).

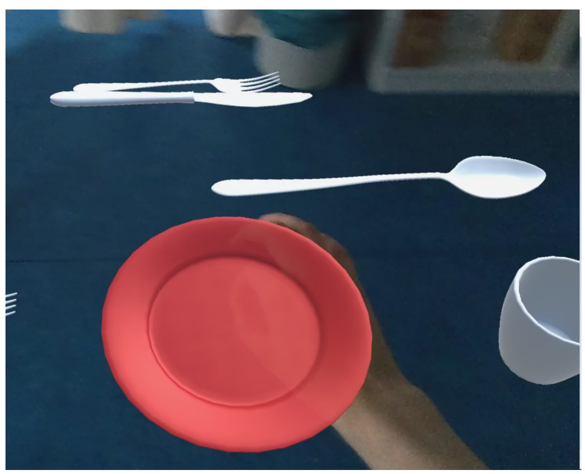

Figure 5.1: An object being held

While holding an object, users can move or rotate it by simply moving/rotating their hand, the object will follow the motion. For more precise rotation, users can grab the object with both hands, in which case it can be rotated by rotating the imaginary axis between the hands. The application uses only one type of interactable UI-Element at the moment: Buttons. To interact with a button, the user simply presses the button with their finger, as if they would with a real button. Users can see each other as avatars represented by a cloud of cubes for the finger joints and a blue head, allowing for gesture communication (see fig. 5.2).

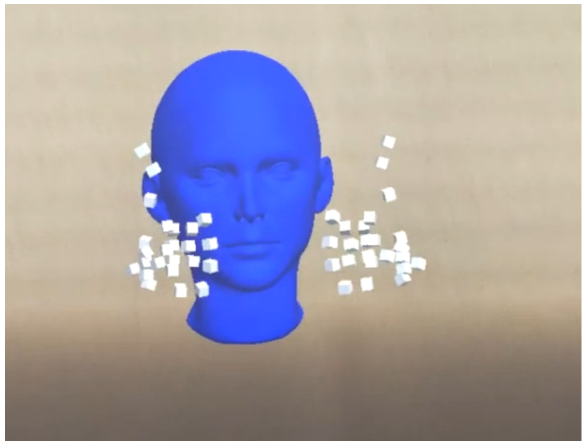

Figure 5.2: An avatar making a thumbs-up-gesture with both hands

5.1.1 Pings

Users that participated in the ping group of the study were also able to ping objects by pointing with both virtual pointers at the same spot for a short time. When the pointers get close to one another, a small loading circle will show up, and once the circle was filled, a ping signal is sent to the other participant, accompanied by a sonar-like sound effect (see fig. 5.3).

Figure 5.3: A ping

The ping utility is then blocked for a few seconds, so users don't accidentally send dozens of pings in a short time.

5.2 Server-Side Administrative Objects

The Server starts up automatically when the Server-Scene is run. The Server-Scene shows all relevant objects, allowing for an overview of the scene at play (see fig. 5.4).

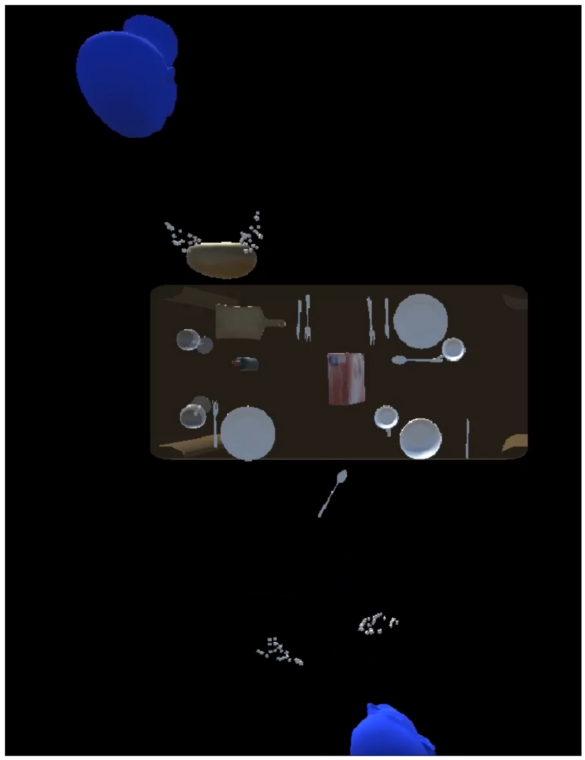

Figure 5.4: Admin view of one user directing another to place bread

It also features a handful of objects for administrative purposes for use in the editor:

5.2.1 Object Manager

The Synchronized Object Manager GameObject displays a list of all registered Synchronized Objects and shows buttons to spawn each of them. When the button is pressed, the corresponding object is spawned across all clients at the 0,0 position. It can then be moved by the experimenter using the Scene-View to wherever it is needed.

Figure 5.5: The object managers spawn Buttons

5.2.2 Device Identifier

The Device Identifier GameObject on the Server features most importantly the ”Client Meta Data Holder” Component. It shows Meta Data (the client Type, device name, and client ID) about each client, which the clients send shortly after connecting. It also allows the configuration of the client connections to the lab streaming layer bridges[9] used for data collection (see fig. 5.6).

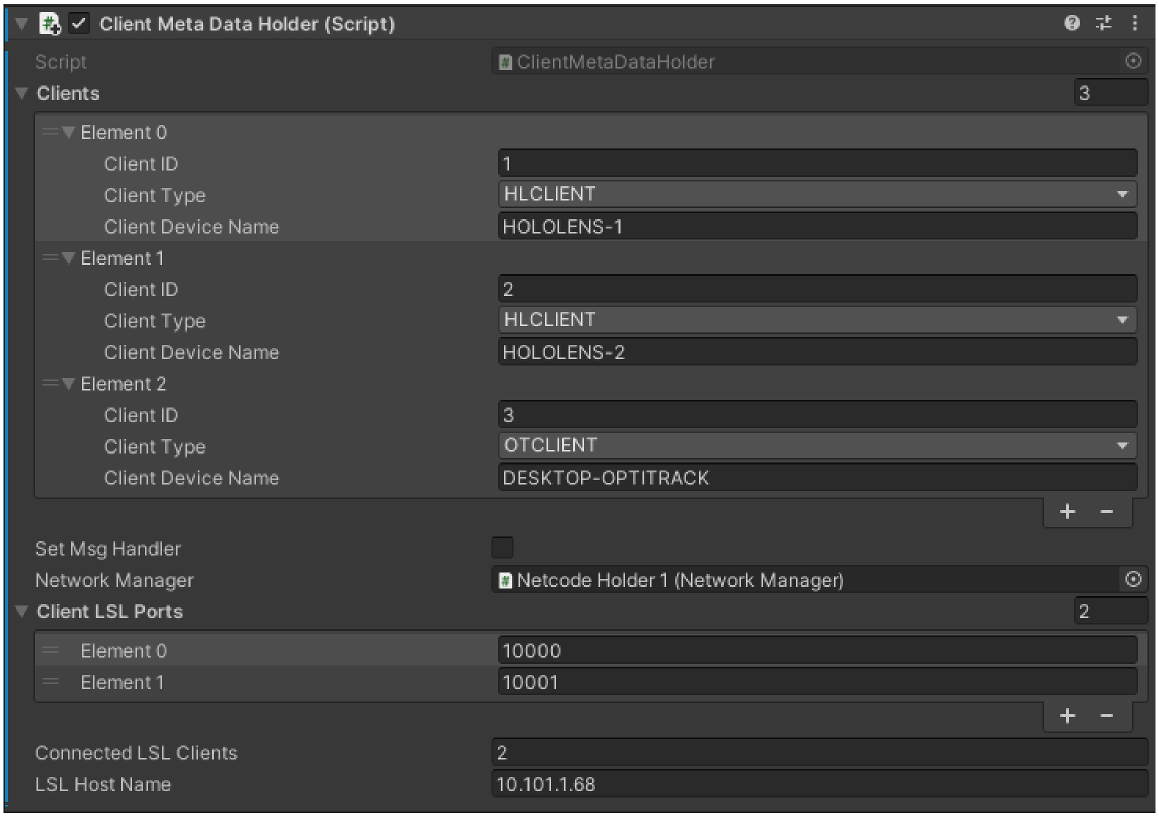

Figure 5.6: The client metadata holder

5.3 Technical Implementation

5.3.1 HoloLens Development with the Unity Engine (MRTK)

The manufacturer-recommended way to build and develop applications for the HoloLens with Unity is by using the Mixed Reality Toolkit (MRTK). After installing a setup tool, it can be automatically integrated into an existing Unity project and then provides Scripts and Prefabs for essential AR interaction in the HoloLens. The Mixed Reality Toolkit provided functionality for the virtual camera, the QR-Code detection, and to interact with objects locally, including UI-Interactions.

5.3.2 Unity Netcode for GameObjects

Unity Netcode for GameObjects is a framework to develop multiplayer applications in the Unity-Engine. It is consists of two accessible communication layers, the Transport and the RPC layer. The RPC layer allows high-level access to sometimes complex Networking functionality, like spawning an object across the Network or establishing & keeping an active connection to the Server.

Since most multiplayer applications require remote function calls, the RPC layer provides a functionality called Unity RPCs. This allows for a function to be declared a ServerRPC(sent to and only executed on the server) or a ClientRPC (sent to and executed on all clients). While RPCs are more than sufficient for most functionalities, they have significant drawbacks for this application:

RPCs are always sent from a Network Object to the clone of that same network Object on the receiver. However, it may be desirable that a different object receives the RPC than sent it, especially since different clients may want to handle the received calls differently.

ClientRPCs also always echo to all clients. This could cause rubber banding (see fig. 5.7) when an object is moved & updated continuously as is required by this application. Of course, this may be avoided by sending an additional parameter and then ignoring the echoed RPCs, but this causes unnecessary bandwidth and performance overhead, as well as boilerplate code.

For these reasons, i have decided to instead use the Messaging system of the lower-level transport layer. The transport layer provides functionality to send named messages and register handlers for said messages. Messages are sent to specific clients, identified by a clientID, meaning that it's easily avoidable to send position updates back to the calling client. This allowed for a flexible and scalable messaging implementation.

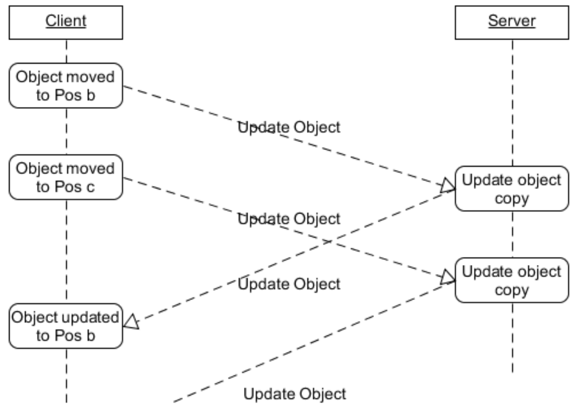

Figure 5.7: Rubberbanding Example

5.3.3 Object Synchronization

Objects are being synchronized spatially in two ways. Their position, rotation, and scale are synchronized relative to a local origin by sending network packets in set intervals. These are based on the Unity-object hierarchy to allow for simple usage for further development: Only the top-level object needs to have SynchronizedObjectParent and NetworkObject scripts attached. The NetworkObject uniquely identifies this object tree across the network with an ID. If the SynchronizedObjectParent is owned by the local client, it checks in a set interval whether any of the child objects poses or scales changed. If so, that object sends out an update package to the server. This package contains information of the following form:

```
[Header]
uint: ObjectID
int: numberOfSubobjects
foreach object in the tree
    string: childObjectName
    float3: positionX, positionY, positionZ
    float4: rotationX, rotationY, rotationZ, rotationW
```

```
    float3: scaleX, scaleY, scaleZ
[End]
```

The server then updates its representation of the object tree accordingly and spreads that message to the other clients.

This approach to synchronization has the advantage that scripts can reference objects locally within their tree, without a need to synchronize these references across the network, as only a whole tree can be spawned at once. Of course, this brings with it the disadvantage objects can not be added locally to the object tree. This needs to be worked around by using object pooling within the trees instead.

In order to synchronize the aforementioned local origin in real space, two technologies are used:

HoloLens clients detect their local origin by detecting a QR-Code, which is then set as the origin coordinate system for the synchronized object. Optitrack clients detect their origin by finding a pre-determined optitrack object and then use that as reference. For a single user using both technologies, a board has been set up to allow for easy synchronization of both systems (see fig. 5.8).

Figure 5.8: The Synchronization Board

Linking of Optitrack Objects

Since only the server can spawn objects, communication about which virtual objects are controlled by an optitrack client (and therefore immovable for HoloLens-users) is required (see fig. 5.9). To establish this, every SynchronizedObject has a state variable if it's OT-Tracked and a trackingID. Directly after an OT-Object is spawned, the server sends a second command to all clients, updating these variables. The Optitrack-Client stores the tracking ID of the last object it asked the server to spawn, therefore being able to link it to the spawned object once the update arrives.

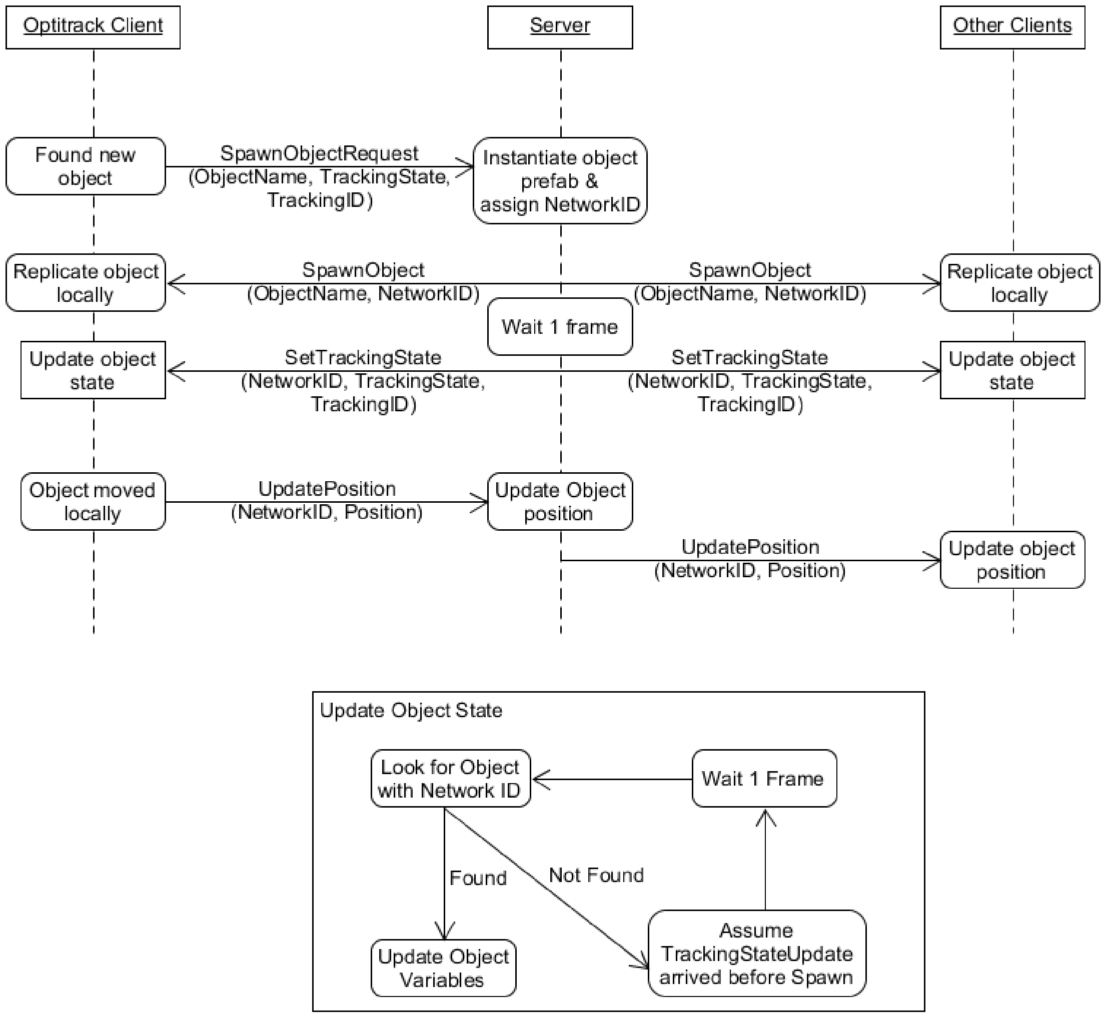

Figure 5.9: Optitrack Spawn Sequence (messages are not instant in reality)

5.4 Cut Features

Some features that were implemented during the preparation of the study were cut after testing or feedback from the pre-study.

5.4.1 Table Snap

This functionality would cause objects released slightly above the table to snap to the table's surface, to increase realism by avoiding floating objects on the table. However, during the pre-study participants noted that this behavior was irritating, and if the alignment of the

QR-Code was not perfect, it could cause the virtual objects to snap to a position inside the real table, making it impossible to interact with them again without recalibrating the table. I judged these drawbacks to outweigh the benefits of this feature, and therefore cut it from the final study.

5.4.2 Arrow-Based highlights

Instead of using pings, an earlier implementation allowed a user to drag an arrow out of an object if they couldn't manipulate said object directly. This arrow would then point from the object to wherever it was dragged, to show the desired change of the object's position to the other user. However, after some testing and consideration, this implementation was dropped in favor of pings, as they are more established and more intuitive to use.

5.5 Extensibility

Easy extensibility was a key focus of the development. As such, all functionality is tied to individual scripts rather than GameObject references, which means that it can be applied to new objects by simply adding the script to said object. Pings, interaction, and synchronization work independently of other scripts or the underlying object structure, only requiring a collider to detect touches/pointers. The system as is can support any number of participants in the same session and can link up any number of optitrack-systems to provide tracking data for objects.

CHAPTER 6

EVALUATION

6.1 Procedure

In order to evaluate the questions posed and validate the system's usability, a small user study was conducted. In it, the participants were split into two groups: Ping and NoPing. In the study, participants were tasked with setting a table together, while being spatially separated. One participant (OT) had access to a real table including some cutlery, the other (HL) only to a digital representation of the same table&objects (see fig. 6.1).

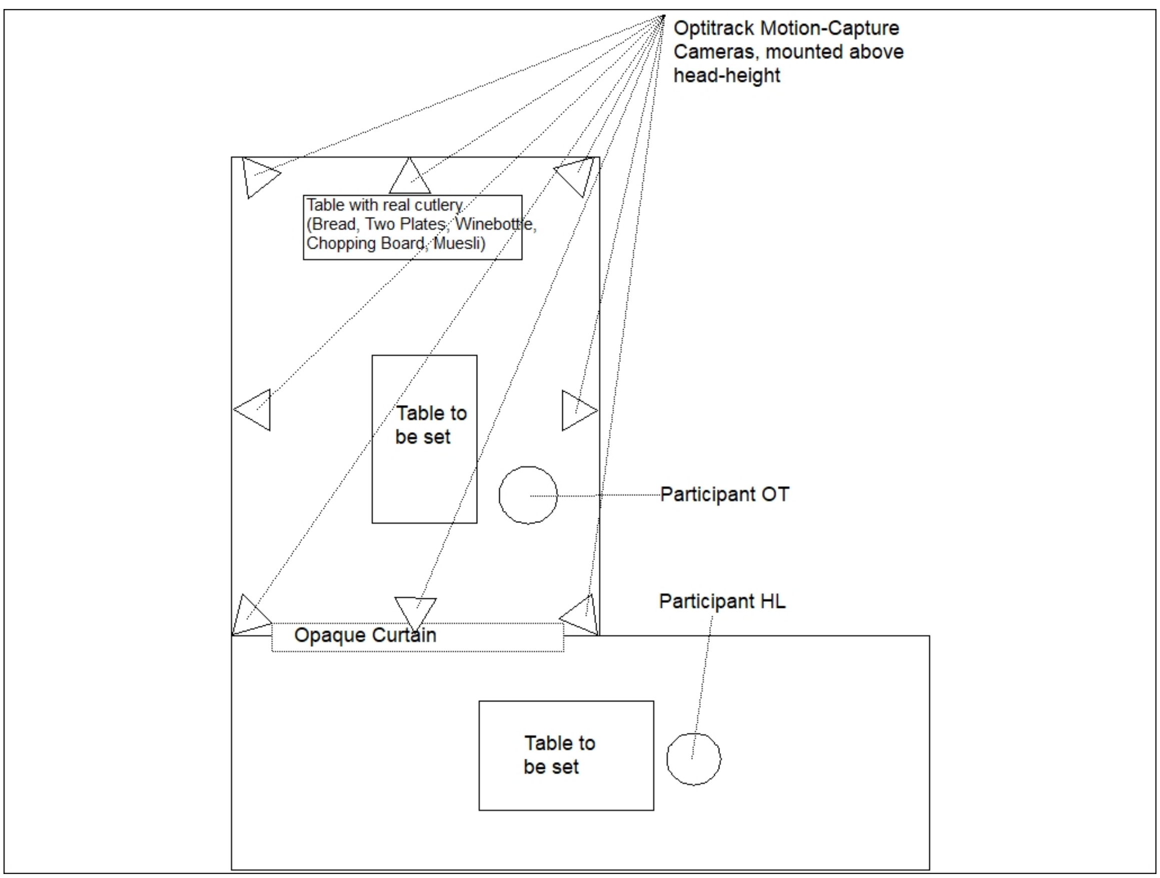

Figure 6.1: Experiment Birdseye View

Both participants shared the digital representation including additional virtual objects, not present in the real world, through a HoloLens.

Both participants could also see target positions for the dishes, in the form of semi-transparent holograms with a blue hue. This representation differed for both participants, half the target locations were only available to participant OT, the other to participant HL. Intentionally, the real objects alone could not fulfill all of the goal locations for the OT participant, and the virtual objects could not fulfill all of the goal locations for the HL participant, forcing both users to communicate about the goals during the exercise.

In total, participants were tasked with placing twenty objects, six of which had real counterparts. The goals formed an asymmetric setup for a four-person breakfast table, as shown in Figure 6.2.

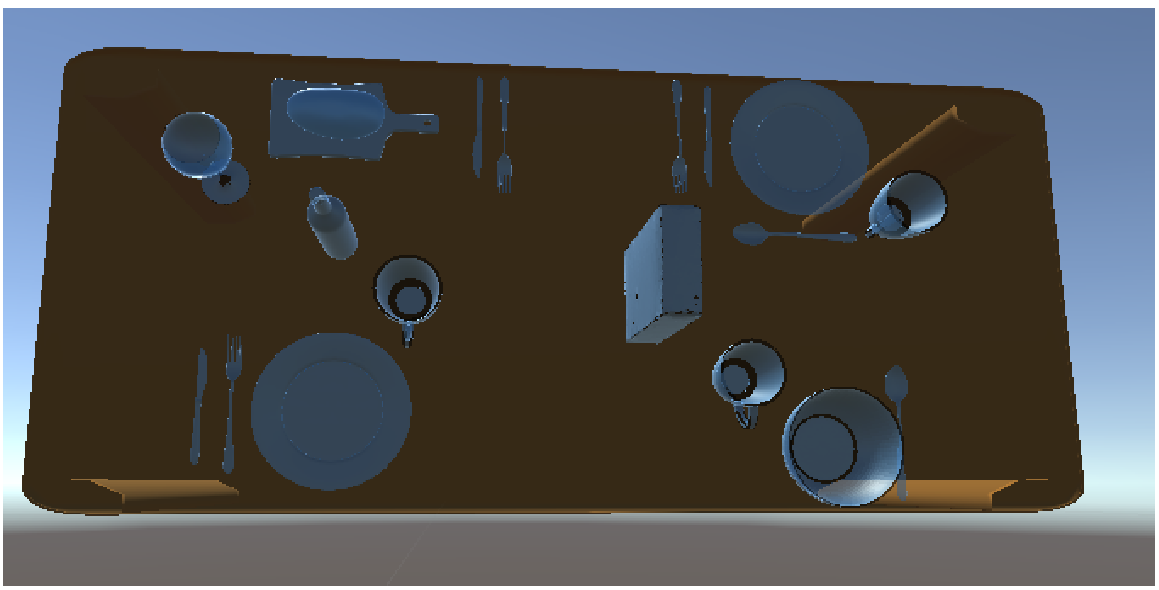

Figure 6.2: Goal-Setup

During the exercise, the following quantitative measures were recorded:

- Hand motion data, as position & rotation of the hand and individual finger joints provided by the HoloLens, at a rate of 10 data points per second.

- Object motion data, as position & rotation of the individual objects tracked via Optitrack.

- Number and time of pings, to see whether a higher ping frequency results in better communication/collaboration.
- At the end of the experiment, the accuracy of the placement was measured for all objects.

After the exercise, participants were also asked to fill out the Shared Workspace Usability Scale, a questionnaire designed to assess how well software supports group work.

6.2 Pre-Study

Before the main study, a small pre-study was conducted with 2 groups of participants. This was done to ensure that the application was running smoothly and that the instructions given in the experiment were understandable to the participants. From a technical view, the pre-study went through successfully, aside from one HoloLens shutting down due to overheating towards the end of an experiment. It is likely that this occurred because the headset was charged immediately before the experiment took place, and the battery was already warm when the experiment started. From then on, I avoided charging the headsets in the 15 minutes before future experiments would start, and reduced strain on the devices by not using the integrated video recording, and the problem did not occur again. Participants in the pre-study had no problem understanding or undertaking the task, but noted minor troubles with the interaction, mostly that it was sometimes difficult to grab objects, especially near the table. The table-snapping feature still present in the pre-study (see section 5.4.1) was noted as a possible source of the problem, and I therefore removed it afterward.

6.3 Shared Workspace Usability Scale

The Shared Workspace Usability Scale (SWUS) was developed in 2018, with the goal of ”assessing the usability of shared workspace groupware applications”. It has been compiled from various other frameworks, and measures seven latent constructs through 22 questions total:

- Grounding: The ability to establish a shared understanding with other participants
- 3CMechanisms: Communication, Coordination, and Cooperation
- Team Integration: Satisfaction with other participants' contribution to the task
- Communication: How easy it is for participants to communicate
- Shared Access: Ease of transferring and accessing data and objects between participants
- Awareness: Knowledge and understanding of other participants, objects, and interactions
- Usability: A subset of more complex usability questionnaires, herein focusing only on satisfaction, effectiveness, and efficiency.

The SWUS was validated in a study with 398 participants, and showed no significant sensitivity to demographic differences, but was sensitive to differences in the software and experience with the software [4]. While the latter is an undesired (albeit possibly unavoidable) aspect, it should not cause a problem for the use of the questionnaire in this thesis, as it is unlikely that participants have large experience differences in the use of augmented reality groupware, since this kind of software is not widely established on the consumer market at the time of writing.

6.4 Demographics

In total, twelve participants joined the study, forming six groups of two. All participants completed the study to the end. Because the study was conducted in Germany, all participants were of Caucasian ethnicity. Participants were primarily young males and had very little prior VR or AR experience (see fig. 6.3).

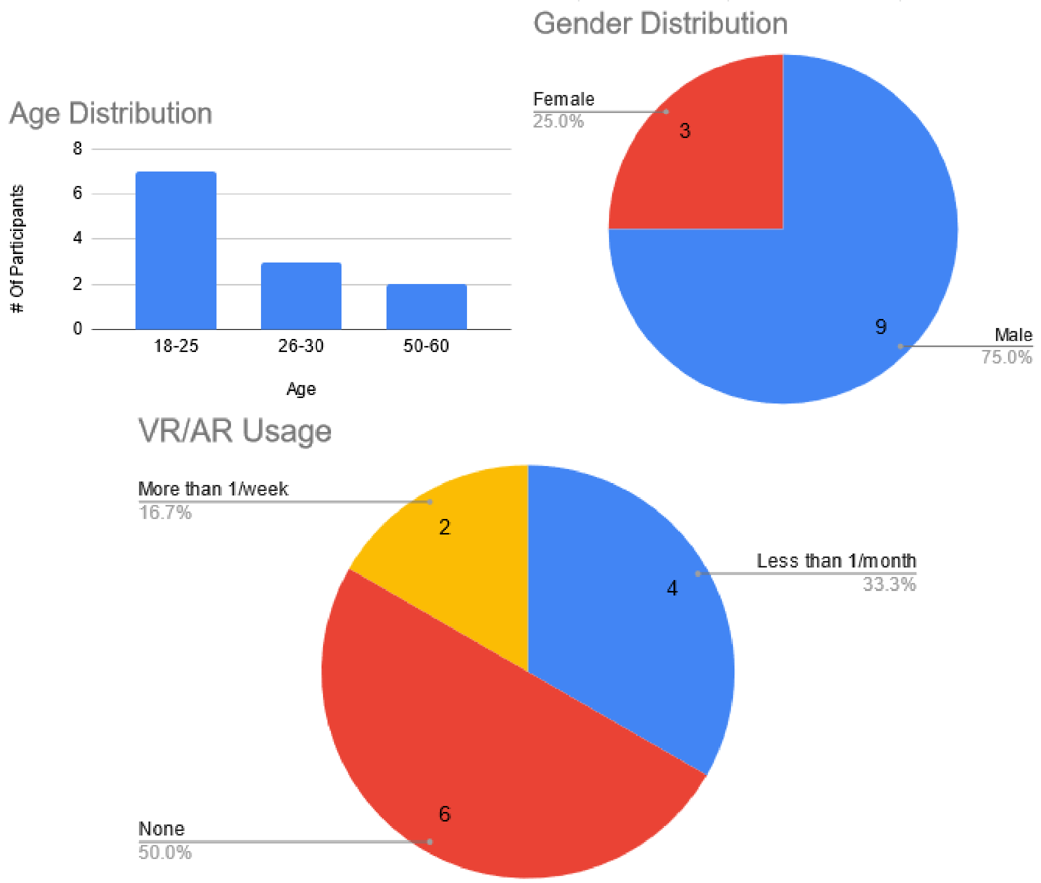

Figure 6.3: Age, Gender and VR/AR-Experience distribution

CHAPTER 7
RESULTS AND DISCUSSION

7.1 Qualitative Evaluation

During the study, I took notes on the participants' comments, behaviors and success. Most (11/12) participants noted that they had fun using the software, and laughter was a common occurrence during the experiments. This however also exposed a small flaw in the setup, as the laughter could be heard by the other participant, which would sometimes be interpreted as positive acknowledgment of an action. In other words, while likely not consciously intended that way by participants, it sometimes circumvented the restriction to nonverbal communication. All participants managed to set the table according to the goals given.

During the study, participants frequently (8/12) noted trouble grabbing smaller objects, like forks & knives, often this was helped by using a pinch gesture instead of a "fist-grip" to interact with the object. Another common issue related to this was the poor tracking of the hands near the table, making it hard to grab and correct the position of an object after it was placed on the table.

Users also noted troubles finding each other, and afterward often said that they weren't aware of the other participant's location in the room most of the time (7/12), the relatively small projection Window of the HoloLens was often stated to be the most likely cause for the problem. Some participants (4/12) also noted issues differentiating between the placeable objects and the transparent goals, stating that they weren't sure if an object had already been placed on the goal or if it still was just the goal. This was especially problematic in the case of the wine glasses which were also somewhat translucent due to their texture.

Nonverbal communication went well in both conditions once established. Users at first always had trouble finding one another or getting the other's attention, but once they did, they were able to show their intent after a few attempts by either pinging or pointing at objects/locations. Subjectively, i did not notice a stark difference in how successful the communication was between the study conditions.

7.2 SWUS-Results

The SWUS-Questionnaire measures 7 subconstructs to combine into a complete collaboration assessment. These subconstructs are *Grounding*, *3C-Mechanisms*, *Usability*, *Team Integration*, *Shared Access*, *Communication*, and *Awareness*. While some of these show statistically significant differences, the small sample size, and subsequent low power scores do not allow for conclusive results (see Table 7.1 and Figure 7.1).

Subscale	Ping Mean	Ping SD	NoPings Mean	NoPings SD	p-Value	Power
Grounding	4.38	1.28	5.66	0.78	0.065	0.46
3C-Mechanisms	4.16	1.06	5.38	1.32	0.109	0.35
Usability	4.0	1.54	5.83	0.51	0.020	0.69
Team Integration	5.33	0.62	6.16	0.30	0.014	0.75
Shared Access	4.61	1.04	5.11	0.77	0.037	0.13
Communication	4.05	1.18	5.5	0.54	0.021	0.68
Awareness	4.88	1.29	5.83	1.20	0.216	0.21

Table 7.1: SWUS Subscales

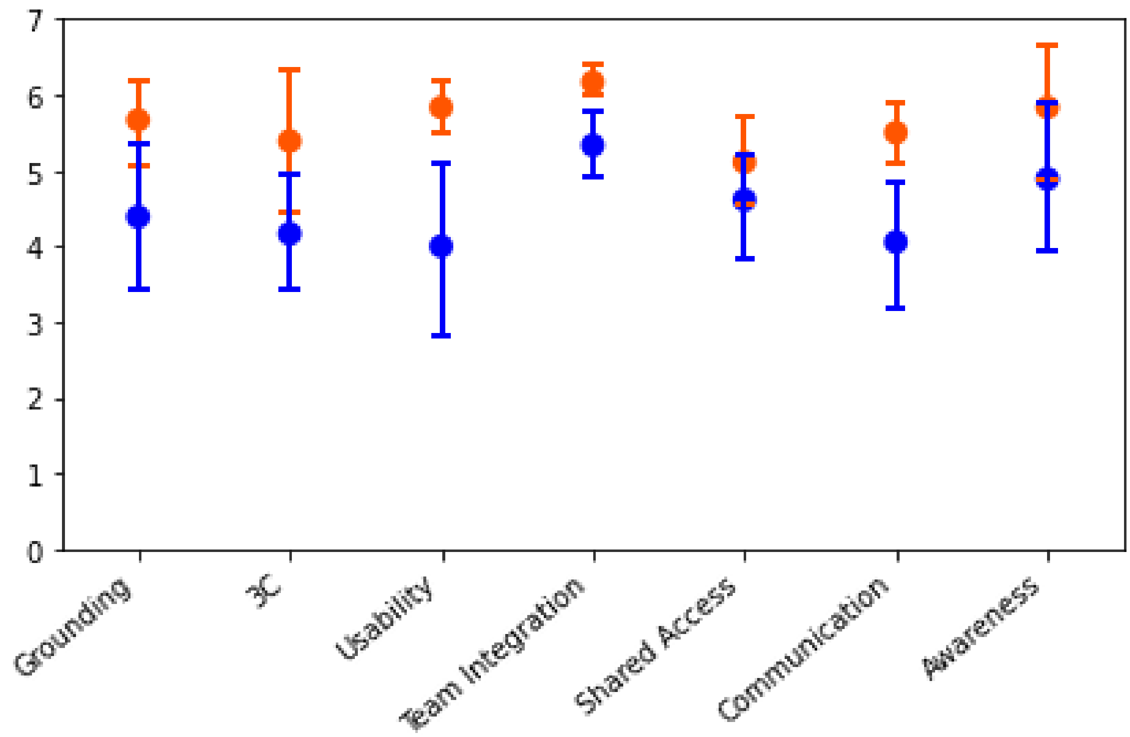

Figure 7.1: SWUS Subscales, Blue: Ping-Condition, Orange: NoPing Condition

Combining all of the subscales results in a collaboration mean value of 4.49 for the Ping condition and 5.64 for the NoPing condition, with a p-Value of 0.019 and a power of 0.70.

7.3 Handtracking

During the study, the users' Hand-Movements were recorded. Figure 7.2 shows the average movement of the participants' left&right hands during the study. While the figure shows that there is a difference in the average Left-Hand movements, this difference is not statistically significant (p=0.098).

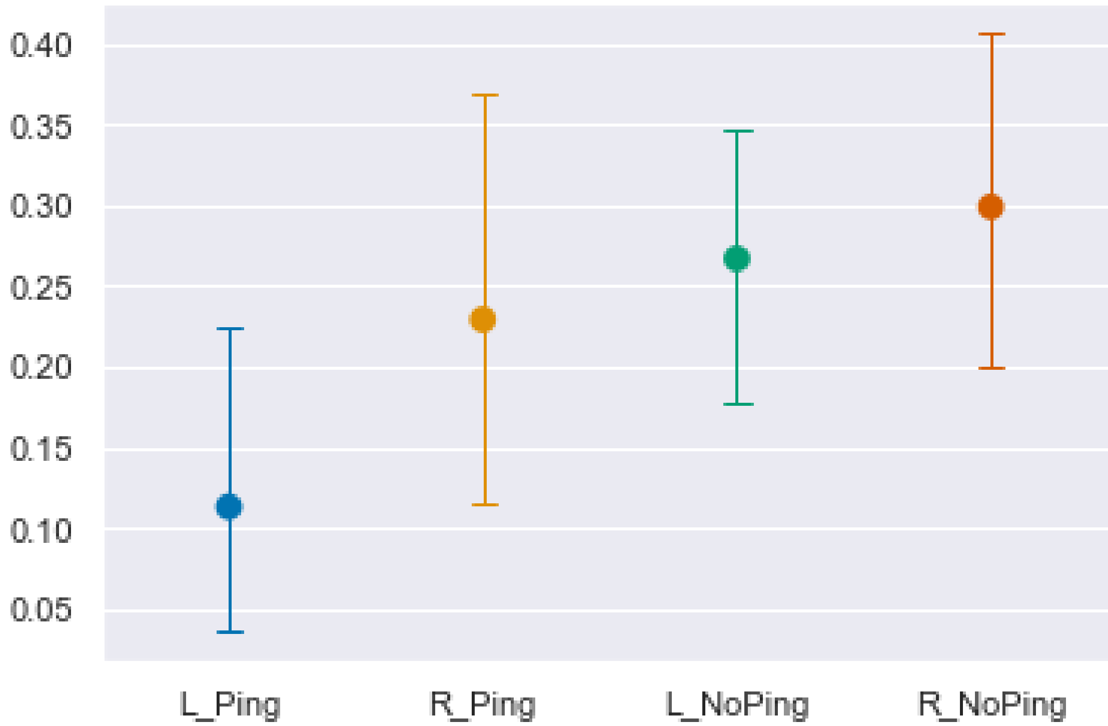

Figure 7.2: Hand Movements (in m/s)

7.4 Pings

7.4.1 Pings and Object Movement

Figure 7.3 shows the average real object movement for each second around a ping. It is apparent that movement spikes after a ping, which aligns with the observation that pings were used to indicate objects and where they should be placed. The smaller spike around the -15 seconds mark likely corresponds to picking up an object after the previous ping.

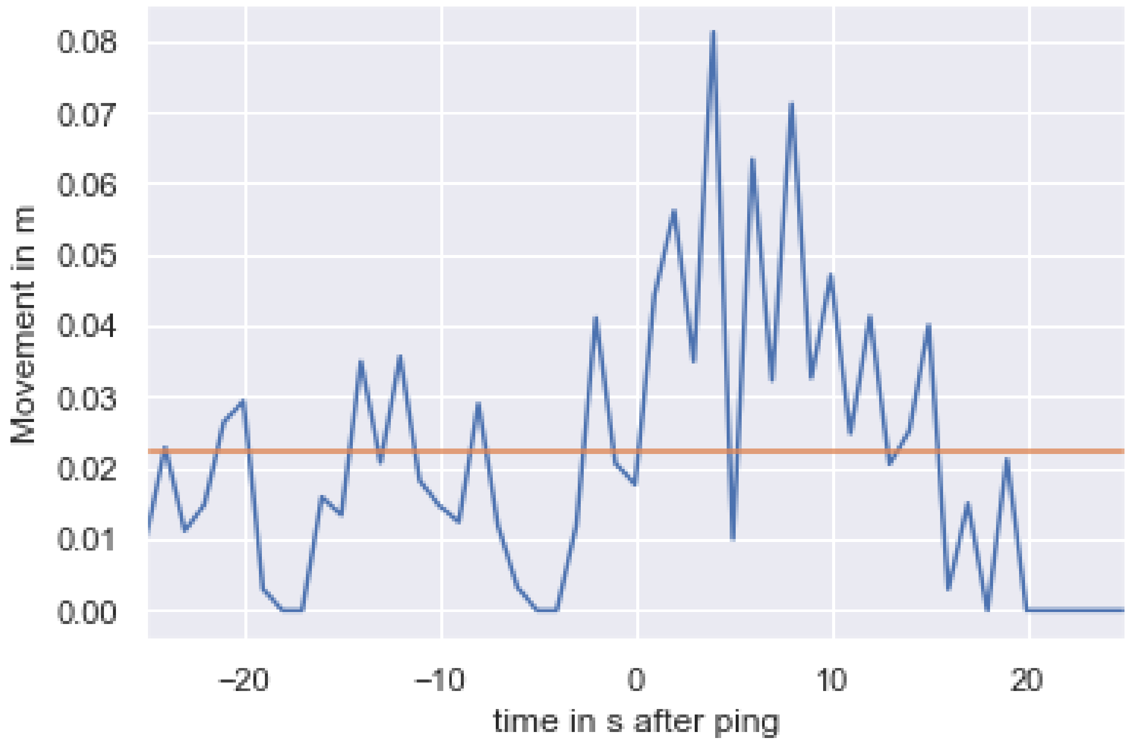

Figure 7.3: Object Movement around Pings, Orange: Average overall object movement

7.4.2 Pings and Handmovement

While Figure 7.3 shows that Pings correlate to spikes of real-object movements, the same can not be said about the movement of the users' hands, where no significant difference could be found in relation to a ping (see fig. 7.4).

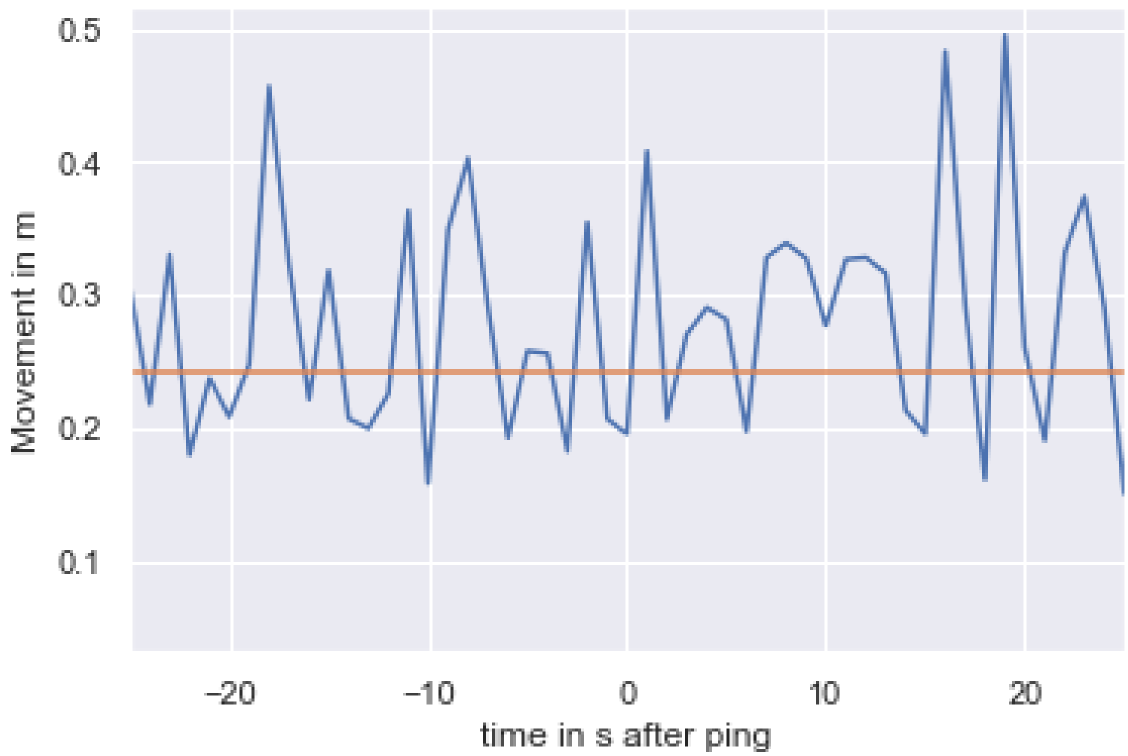

Figure 7.4: Hand&finger-movements around Pings, Orange: Average overall hand&finger-movements excluding resting periods

7.5 Accuracy

Due to a technical error, the data collected on accuracy was not useful for analysis, as an incorrect transform was used to measure the distance. This resulted in most measurements showing a positioning error of > 1 meter. Therefore, a comparison of the accuracy of placements in both conditions or in regard to real/virtual objects has to be excluded from this thesis.

7.6 Discussion

Overall, the data analyzed shows a conflicting picture: While the SWUS results indicate that participants found it easier to collaborate without using Pings, especially in terms of Communication, Usability and Team Integration, the motion data collected around Pings reinforces that Pings are a useful tool for communication, and have been used successfully by the participants to communicate about their intents and actions.

7.7 Technical Performance

The application runs on current HoloLens 2 Devices at a stable 60FPS (limited by the device's refresh rate). Of course, this depends on the objects and object meshes used in the application. Currently, most objects had a relatively high polygon count due to many round surfaces, totaling about 100k triangles if every object would be visible in the scene. Microsoft suggests that performance issues are to be expected at around 150k polygons, but a test by Fologram [14] concluded that the HoloLens 2 can render about 2 million polygons without dropping below 30FPS. It should however be noted that in a typical experiment setup only a subset of all objects will be inside the camera's viewport and therefore rendered at any given time.

At a network update rate of 30 Hz, a constantly moving object requires about 3kB/s to stay synchronized. This scales linearly with the number of objects. As such, even slow connections of 500kB/s can easily handle dozens of objects in synchronized motion. In reality, it is very rare that multiple objects need to be updated every update tick, and as such hundreds of objects can feasibly be kept in sync.

CHAPTER 8
CONCLUSION AND FUTURE WORK

8.1 Conclusion

In this thesis, I researched the potential usefulness of an augmented reality environment to conduct collaborative everyday activity experiments. Based on that background research, I laid out design requirements for an application to provide such an environment and implemented said design in the Unity Game Engine. I then conducted a short prestudy, where I found a few flaws in the initial implementation. After reworking these features, a larger study was conducted, investigating the use of pings for nonverbal communication in augmented reality through a table-setting experiment. This study found indications that pings may not be a good tool to communicate in augmented reality, but these indications are inconclusive due to the small number of participants. All participants in the study did manage to set the table together, without major complications, and several participants noted that using the system was fun.
In conclusion, I believe that the developed application offers new opportunities for research in the fields of everyday activities and augmented reality collaboration, as well as a platform for experiments on LLL-4 through the web.

8.2 Future Work

Future work should include a larger study to validate or refute the indications regarding communication through pings in augmented reality. Further studies in the field of augmented reality collaboration using this system could also yield interesting insights on how to improve the system, for example by investigating further communication channels, simulating impairments, or adding new client types, such as smartphone or VR clients.

As participants have often cited awareness issues, it might be a good idea to add the awareness cues described by Piumsomboon et al [28] to the application. In addition, the user avatars could be improved by constructing hand-meshes from the given positions instead of the currently used cube-cloud, and by potentially adding more detailed avatars than the currently used floating heads & hands.

In order for the application to be more useful for data collection in the EASE-project, additional data-logging channels such as eye-tracking could be added. Another technical inconvenience is the current necessity of bridge programs to stream data to the lab streaming layer. As these were subject to some crashes during the study, it would be a significant improvement to implement native lab streaming layer connectivity.

REFERENCES

[1] 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*, Hong Kong, Hong Kong: ACM Press, 2013, pp. 235–242, ISBN: 978-1-4503-2590-5. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2534329.2534331.

[2] T. Bates, K. Ramirez-Amaro, T. Inamura, and G. Cheng, "On-line simultaneous learning and recognition of everyday activities from virtual reality performances," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, BC: IEEE, Sep. 2017, pp. 3510–3515, ISBN: 978-1-5386-2682-5. [Online]. Available: http://ieeexplore.ieee.org/document/8206193/.

[3] C. Bennett, *EA patents Apex Legends ping system but want other games to use it*, Aug. 2021. [Online]. Available: https://www.dexerto.com/apex-legends/ea-patents-apex-legends-ping-system-but-want-other-games-to-use-it-1637929/.

[4] M. İ. Berkman, D. Karahoca, and A. Karahoca, "Shared Workspace Usability Scale," American Psychological Association, Tech. Rep., Sep. 2018, Type: dataset. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/t68870-000.

[5] M. Billinghurst and H. Kato, "Collaborative augmented reality," *Communications of the ACM*, vol. 45, no. 7, pp. 64–70, Jul. 2002. [Online]. Available: https://dl.acm.org/doi/10.1145/514236.514265.

[6] Y. Chen, Q. Wang, H. Chen, X. Song, H. Tang, and M. Tian, "An overview of augmented reality technology," *Journal of Physics: Conference Series*, vol. 1237, no. 2, p. 022 082, Jun. 2019. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1237/2/022082.

[7] EASE, *Lab Streaming Layer*, Aug. 2022. [Online]. Available: https://github.com/sccn/labstreaminglayer.

[8] *EASE Collaborative Research Center*. [Online]. Available: https://ease-crc.org/.

[9] Felix Kroll, *LSLHoloBridge*, Aug. 2022. [Online]. Available: https://gitlab.csl.uni-bremen.de/fkroll/LSLHoloBridge.

[10] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, "In touch with the remote world: Remote collaboration with augmented reality drawings and virtual navigation," in *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST '14*, Edinburgh, Scotland: ACM Press, 2014, pp. 197–205, ISBN:

978-1-4503-3253-8. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2671015.2671016.

[11] Ghost Ship Games, *Deep Rock Galactic*, Feb. 2018.

[12] *Glass*. [Online]. Available: https://www.google.com/glass/start/.

[13] B. Grothaus, *Das Ping-System von Apex Legends ist so gut, dass andere Shooter es kopieren wollen*, Mar. 2019. [Online]. Available: https://mein-mmo.de/rainbow-six-will-pings-von-apex-legends/.

[14] Gwyll/Fologram, *HoloLens 2 Polygon Count and Frame Rate - Tech Support / Models*, Jan. 2020. [Online]. Available: https://community.fologram.com/t/hololens-2-polygon-count-and-frame-rate/49.

[15] A. Haidu and M. Beetz, "Automated Models of Human Everyday Activity based on Game and Virtual Reality Technology," in *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada: IEEE, May 2019, pp. 2606–2612, ISBN: 978-1-5386-6027-0. [Online]. Available: https://ieeexplore.ieee.org/document/8793859/.

[16] T. Innocent and S. Haines, "Nonverbal Communication in Multiplayer Game Worlds," in *Proceedings of the 4th Australasian Conference on Interactive Entertainment*, ser. IE '07, event-place: Melbourne, Australia, Melbourne, AUS: RMIT University, 2007, ISBN: 978-1-921166-87-7.

[17] A. Krugliak and A. Clarke, "Towards real-world neuroscience using mobile EEG and augmented reality," *Scientific Reports*, vol. 12, no. 1, p. 2291, Dec. 2022. [Online]. Available: https://www.nature.com/articles/s41598-022-06296-3.

[18] E. Lathi, "Apex Legends' ping system is a tiny miracle for FPS teamwork and communication," *PC Gamer*, Feb. 2019. [Online]. Available: https://www.pcgamer.com/apex-legends-ping-system-is-a-tiny-miracle-for-fps-teamwork-and-communication/.

[19] *League of Legends - Smart Ping*. [Online]. Available: https://support-leagueoflegends.riotgames.com/hc/en-us/articles/201752974-Smart-Ping.

[20] A. Leavitt, B. C. Keegan, and J. Clark, "Ping to Win?: Non-Verbal Communication and Team Performance in Competitive Online Multiplayer Games," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA: ACM, May 2016, pp. 4337–4350, ISBN: 978-1-4503-3362-7. [Online]. Available: https://dl.acm.org/doi/10.1145/2858036.2858132.

[21] D. Maloney, G. Freeman, and D. Y. Wohn, “”Talking without a Voice”: Understanding Non-verbal Communication in Social Virtual Reality,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW2, pp. 1–25, Oct. 2020. [Online]. Available: https://dl.acm.org/doi/10.1145/3415246.

[22] C. Mason, K. Gadzicki, M. Meier, *et al.*, “From Human to Robot Everyday Activity,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 8997–9004, ISBN: 978-1-72816-212-6. [Online]. Available: https://ieeexplore.ieee.org/document/9340706/.

[23] *Microsoft HoloLens — Mixed Reality Technology for Business*. [Online]. Available: https://www.microsoft.com/en-us/hololens.

[24] Niantic, The Pokemon Company, and Nintendo, *Pokemon Go*, Jul. 2016.

[25] E. Pettersson and V. Sundstedt, “A perceptual evaluation of social interaction with emotes and real-time facial motion capture,” in *Proceedings of the Tenth International Conference on Motion in Games*, Barcelona Spain: ACM, Nov. 2017, pp. 1–6, ISBN: 978-1-4503-5541-4. [Online]. Available: https://dl.acm.org/doi/10.1145/3136457.3136461.

[26] C. Pidel and P. Ackermann, “Collaboration in Virtual and Augmented Reality: A Systematic Overview,” in Aug. 2020, pp. 141–156, ISBN: 978-3-030-58464-1.

[27] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst, “The Effects of Sharing Awareness Cues in Collaborative Mixed Reality,” *Frontiers in Robotics and AI*, vol. 6, p. 5, Feb. 2019.

[28] T. Piumsomboon, Y. Lee, G. Lee, and M. Billinghurst, “CoVAR: A collaborative virtual and augmented reality system for remote collaboration,” in *SIGGRAPH Asia 2017 Emerging Technologies*, Bangkok Thailand: ACM, Nov. 2017, pp. 1–2, ISBN: 978-1-4503-5404-2. [Online]. Available: https://dl.acm.org/doi/10.1145/3132818.3132822.

[29] Rec Room Inc, *Rec Room*, Jun. 2016.

[30] Respawn Entertainment and Panic Button Games, *Apex Legends*, Feb. 2019.

[31] Riot Games, *League of Legends*, Oct. 2009.

[32] T. Schultz, F. Putze, T. Fehr, *et al.*, *Linking Labs: Interconnecting Experimental Environments*, arXiv:2102.03684 [cs], Feb. 2021. [Online]. Available: http://arxiv.org/abs/2102.03684.

[33] Z. Szalavàri, D. Schmalstieg, A. Fuhrmann, and M. Gervautz, "?Studierstube?: An environment for collaboration in augmented reality," *Virtual Reality*, vol. 3, no. 1, pp. 37–48, Mar. 1998. [Online]. Available: http://link.springer.com/10.1007/BF01409796.

[34] F. Tecchia, L. Alem, and W. Huang, "3D helping hands: A gesture based MR system for remote collaboration," in *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry - VRCAI '12*, Singapore, Singapore: ACM Press, 2012, p. 323, ISBN: 978-1-4503-1825-9. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2407516.2407590.

[35] Vito Technology Inc., *Starwalk*, Nov. 2008.

[36] P. Wang, S. Zhang, X. Bai, *et al.*, "2.5DHANDS: A gesture-based MR remote collaborative platform," *The International Journal of Advanced Manufacturing Technology*, vol. 102, no. 5-8, pp. 1339–1353, Jun. 2019. [Online]. Available: http://link.springer.com/10.1007/s00170-018-03237-1.

[37] X. Wang and P. S. Dunston, "GROUPWARE CONCEPTS FOR AUGMENTED REALITY MEDIATED HUMAN-TO-HUMAN COLLABORATION," p. 7, 2006.