

Real-Time Augmented Reality Collaboration with Different Device Types

Master Thesis

Media Informatics
RWTH Aachen

Mixed and Augmented Reality Solutions
Fraunhofer FIT

Yücel Uzun
351062

2018-05-06

Examiners:

Prof. Dr. Wolfgang Prinz
Prof. Dr. Thomas Rose

Eidesstattliche Versicherung

Statutory Declaration in Lieu of an Oath

Yücel Uzun**351062**

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)

Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/
Masterarbeit* mit dem Titel

I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

Real-Time Augmented Reality Collaboration with Different Device Types

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Aachen, May 6, 2018

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:**Official Notification:****§ 156 StGB: Falsche Versicherung an Eides Statt**

Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

Para. 156 StGB (German Criminal Code): False Statutory Declarations

Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt

(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.

Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence

(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.

(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:

I have read and understood the above official notification:

Aachen, May 6, 2018

Ort, Datum/City, Date

Unterschrift/Signature

Abstract

In recent years, the popularity of Augmented Reality (AR) has increased considerably across the consumer market. One of the trending usage areas of AR is collaboration. Despite the comprehensive work on AR collaboration, possible effects of using different device types during collaboration, such as tablets and smart glasses, are only barely explored. During this thesis, a user study was designed to investigate the effects of device types during real-time, face to face AR collaboration on performance, group coordination and referenced object types. An AR application is implemented for HoloLens and iPad to use during the user study. Results indicate that the device type does not have a significant effect on performance but collaborators use more hand gestures and spatial expressions with the HoloLens. Furthermore, participants refer to virtual objects extensively, compared to physical objects. Finally, users prefer using combinations of the same devices rather than a mixture of them.

Contents

List of Figures	ix
List of Tables	xi
1 Introduction	1
2 Theoretical Background	5
2.1 Augmented Reality	5
2.1.1 Display Technologies	6
2.1.2 Input Technologies	7
2.1.3 Tracking Technologies	8
2.1.4 Commercial Products	9
2.2 Unity Editor and Game Engine	12
2.2.1 Augmented Reality and Unity	13
2.2.2 Real-time Collaboration and Unity	14
3 Related Work	15
3.1 General Overview	15
3.2 Studierstube	18
3.3 Handheld or Handsfree	18
3.4 Virtual Objects As Spatial Cues	20
4 Objective and Research Questions	23
5 User Study Design	27
5.1 User Study Requirements and Decisions	27
5.2 Low Fidelity Prototypes	28
5.2.1 First Iteration	29
5.2.2 Second Iteration	31
5.2.3 Requirements for the Implementation	32

6 Implementation	35
6.1 Overview and Development Environment	35
6.2 Augmented Reality	37
6.2.1 Displaying Objects	37
6.2.2 Moving Objects	38
6.3 Networking	39
6.3.1 Photon Networking	39
6.3.2 Networking Implementation	40
6.4 Real-time Collaborative AR Application	41
6.4.1 Initialization	42
6.4.2 Object Identification Task	43
6.4.3 Object Positioning Task	45
6.4.4 Logging	47
7 Evaluation	49
7.1 User Study	49
7.1.1 Evaluation procedure	50
7.1.2 Participants	51
7.1.3 Apparatus and Study Environment	51
7.1.4 Data Collection	51
7.2 Evaluation Results	53
7.2.1 Performance	53
7.2.2 Communication	55
7.2.3 Questionnaire	60
7.3 Discussion	61
7.3.1 Performance	61
7.3.2 Group Coordination	62
7.3.3 Object References	63
7.3.4 Research Questions	63
8 Conclusion and Future Work	67
8.1 Future Work	68
9 Bibliography	i
A Low Fidelity Study Design	ix
B Collected Data	xiii

List of Figures

1.1	AR collaboration with different devices	2
2.1	AR cheat sheet	6
3.1	Selected related work examples	16
3.2	Selected related work examples	17
3.3	Handheld or handsfree	19
3.4	Virtual objects as spatial cues	20
4.1	How to press a button with a HoloLens	24
4.2	How to press a button with a tablet	24
4.3	How to press a button with Epson Moverio BT-300	25
5.1	Paper prototype implementation for the Tower of Hanoi game.	29
5.2	Color filtered glasses and example shapes	30
5.3	Construction Task	32
6.1	Simplified system architecture	36
6.2	Displaying objects at the same positions	38
6.3	An example of network communication	41
6.4	Device preparation	42
6.5	Virtual objects in the test room	43
6.6	The cube positions and symbols	44
6.7	Decision flow chart for cube selection	45
6.8	Visualization of moving cubes	46
6.9	Calculating distances	46
7.1	Physical objects in the test room	52
7.2	The analysis of the object references	56

x

List of Tables

2.1	Comparison of some commercial AR products	10
7.1	Device combinations during the study	50
7.2	The analysis of the performance	54
7.3	The analysis of the deictic speech	56
7.4	The analysis of the object references	57
7.5	The analysis of the spatial vocal expressions	58
7.6	The analysis of the hand gestures	59
7.7	Questionnaire results for individual devices section	61
7.8	Questionnaire results for user preferences	61

Glossary

2D	Two-dimensional.
3D	Three-dimensional.
API	Application Programming Interface.
AR	Augmented Reality.
DoF	Degrees of Freedom.
GPS	Global Positioning System.
HMD	Head-mounted Display.
IMU	Inertial Measurement Unit.
PC	Personal Computer.
SDK	Software development kit.
SLAM	Simultaneous location and mapping.
VR	Virtual Reality.

Chapter 1

Introduction

Augmented Reality (AR) is a technology which aims to blend virtual content with the real environment seamlessly. With AR, virtual elements are added over the real world. For instance, a user might walk through a museum and see related information like text, videos and animations over the exhibited items while being followed by a 3D animated tour guide. AR offers a vast amount of potential use, both in business and entertainment. On the other hand, any AR system designed for everyday usage should be not only cost effective, but also mobile and comfortable enough to use for long duration. Due to the requirement of high computational power and cost of the display and tracking devices, the consumer market did not have any serious interest on AR for a long time. However in recent years, the popularity of AR has increased considerably thanks to improvements in mobile processors, sensors and tracking algorithms.

Today, there are many AR devices and software frameworks on the consumer market from various companies like Microsoft, Google, Vuforia and Apple. But this variety comes at a price, because none of the used technologies are standardized yet and all of these solutions have minor and major differences. The most obvious and drastic changes are between tablets and smart glasses. Tablets display virtual objects with a handheld device and get input via touch screen. On the other hand, smart glasses use head-mounted displays (HMD) and use separate touch-pad or hand gestures for input. Inevitably, these changes are also reflected in the developed AR applications and user experience; meaning that even the same application viewed through different devices might change the user experience vastly.

Real-time collaboration is one of the most anticipated areas of AR. It is not

only interesting for leisure, like multiplayer gaming, but also promising for the industry, like remote maintenance. There are already extensive amounts of academical work in this area, and consumer products are catching up as well. Microsoft is already promoting HoloLens with emphasis on real-time collaboration capabilities and providing all the necessary software frameworks for developers.



Figure 1.1: Example view of an AR collaboration app developed during this thesis. Users see the same virtual objects while using different devices.

At this point, several questions arise. What would happen if different collaborators use different devices? Would this affect their performance? How would their communication be affected? Would the physical environment or virtual objects be dominant in their conversation? These questions are important for any collaborative AR application that aims to provide the best user experience to their users without compromising the performance. However, effects of the different devices on collaboration have barely been explored and there are gaps in the literature. Therefore, the main aim of this thesis is to be able to answer these, and more of such, questions.

First of all, chapter 2 gives general information about AR technologies, consumer products and Unity Editor and game engine. The related work is presented in chapter 3. In chapter 4, the main objectives of this thesis and its research questions are formulated and discussed. Chapter 5 lists the requirements for the user study based on the research questions and explains the

design process of the main study concept. Implementation of the collaborative application, encountered problems and their solutions are discussed in chapter 6. In chapter 7, process and results of the evaluation are discussed. Finally, this thesis is concluded in chapter 8.

Chapter 2

Theoretical Background

This chapter provides a basic understanding about main concepts, technologies and terminology related to this thesis. Section 2.1 describes the concept of Augmented Reality (AR), the variety of technologies available to realize it and commercially available products. Section 2.2 focuses on the Unity game engine and its relation with AR and real-time collaboration.

2.1 Augmented Reality

AR is a technology which aims to blend virtual content with the real environment seamlessly. In his widely accepted definition, Azuma [1] specified three main characteristics of AR systems:

- Combine real and virtual
- Interactive in real time
- Registered in 3D

According to both Billinghurst and Schmalstieg, these characteristics define not only an AR system, but also the technical requirements for such a system. An AR system should have a display that combines the real and virtual context, be able to respond to user commands in real time and track user position to keep virtual context fixed in the real world. Furthermore, this definition does not limit any of these requirements with specific technologies and rather flexible on implementation aspect [2,3]. A general overview of the AR technologies are displayed on Figure 2.1.

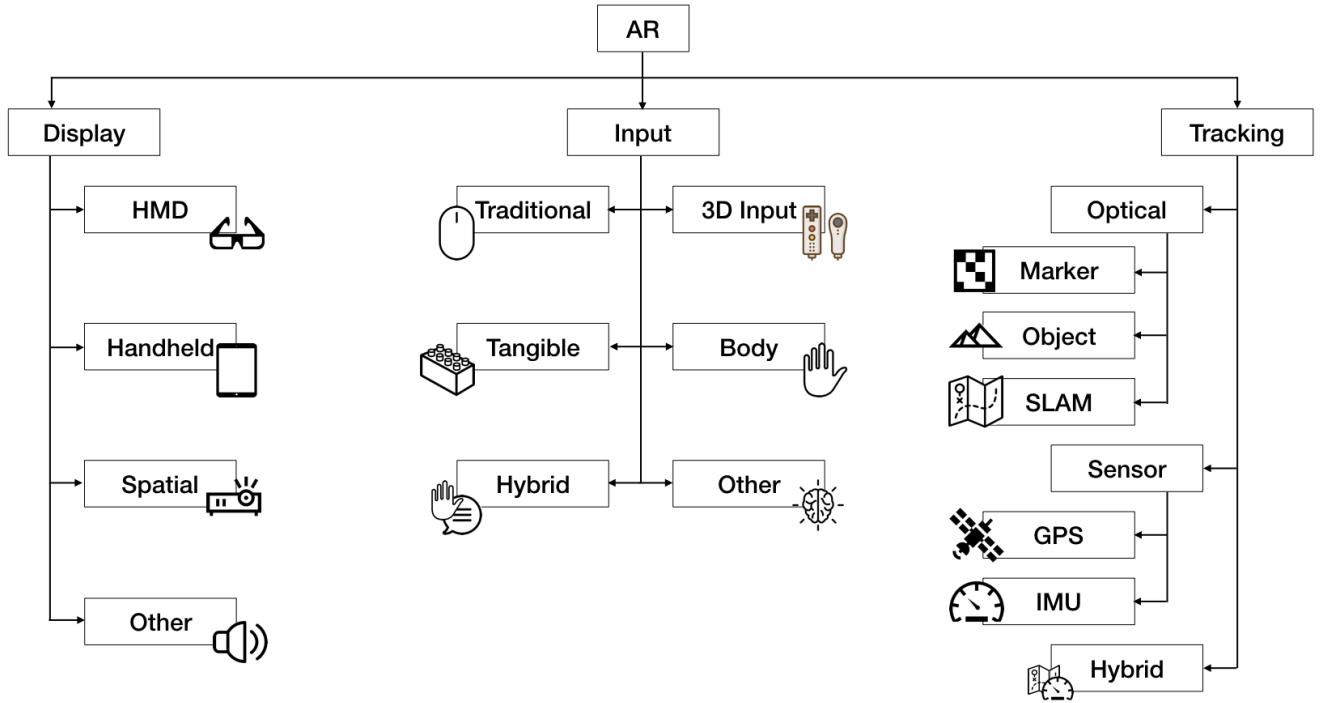


Figure 2.1: Overview of AR technologies.

2.1.1 Display Technologies

AR and visual displays are often thought of together, but as explained in the previous section, Azuma's definition does not specify the technology, hence other display types, such as audio [4], haptics [5] and even olfactory [6] or gustatory [7] can also be used. However, majority of the work in this area mainly focuses on the visual displays.

Visual displays are used to merge computer generated images with the real world and according to Billinghurst et al. can be categorized based on how they are placed between user's eyes and the real world [2]:

- **Head-attached:** Display is attached to user's head and there is nothing else between the display and the eyes. Can vary in size from size of a regular glass to a helmet. Ensures no other object comes between the view and eye, and leaves both hands of the user free.
- **Handheld and body-attached:** Display is either carried by the user or attached to the body. More socially accepted than head-attached displays and considered to be mobile and personal while still shareable.
- **Spatial:** Usually installed at a fixed location, very suitable as a public

display for multiple users.

2.1.2 Input Technologies

Like display technologies, interaction methods and techniques in AR also have huge variety. They highly depend on the application context and the use cases and can be categorized into six groups [2].

AR information browsers displays information that is tied into the physical world. Users can simply move to see the information and simple actions like filtering the information can be achieved with traditional input devices like keyboard, mouse, touch pad or touch screen. For instance, in the AR part of the Building Information Modeling issue tracking system of Biçer, 2D interactions are enough for all functions and solely touch screen of the Google Tango tablet is used for the input [8].

Over the years, many 3D interaction techniques for desktop and Virtual Reality (VR), such as 3D mouse, 6DoF joysticks or spaceballs have been developed, and they can easily be adopted to AR. With these controllers, users can select and manipulate virtual objects. For instance, Butz et al. used "3D pointing devices that combine a tracker target and two buttons to control a 3D arrow" to manipulate virtual objects [9].

In systems that use tangible user interfaces, a physical object represents a virtual one, and users can manipulate the virtual object via the physical one. For instance, Regenbrecht et al. used a plate-shaped device for displaying 3D objects which users can rotate by rotating the plate [10].

User's body motions and gestures can also be tracked and used as the main input source. Tracking might be done either with wearable sensors or image processing. For instance Lee and Höllerer created a system where users can directly interact with objects using hand gestures [11].

Naturally, different modalities can be used together. Most common combination is speech and gesture. For instance, Lee et al. compared speech-gesture combination with only gesture input systems, and found out combination is actually more usable and satisfying for the users [12].

In addition to these, some of the interaction methods are either too specific to use apart from their specific use case (e.g. whistling [13]) or still in early stages of the research (e.g. brain computer interfaces [14]).

It is also worth noting that display and input technologies are highly related

from a user experience aspect, for example a tangible input device that requires both hands to operate could not be used if the display is handheld.

2.1.3 Tracking Technologies

To keep virtual context fixed in the real world, any AR system is required to track positions of users and all the related objects such as input devices and displays, and update the context based on the changes. In the following sections, some of the tracking methods are discussed.

Optical Tracking

Optical tracking methods use camera images to calculate position of the camera or other objects of interest. This is especially convenient for systems that render virtual objects in front of the video image since the camera sensor is already present inside the system.

Depending on the use case, it is very popular to detect and track image markers or objects. In these methods, outstanding and unique features of the marker or the object are pre-detected, and features of candidate images that captured by the camera are compared with these to detect the marker or object.

Both marker and model tracking require some pre-trained data to begin with, which also might not be suitable for the use case. In these situations Simultaneous location and mapping (SLAM) tracking is used [15]. In a nutshell, SLAM algorithms first find features in the first captured frame of a video and track these features while continuously updating the tracked feature list with new captured frames. Afterwards, they calculate the relative camera position to the starting point based on the changes between frames.

In addition to regular optical cameras, sensors that are capable of detecting 3D depth of the environment have also become popular in recent years, especially after the release of Microsoft Kinect¹. By using infrared camera and projectors, these sensors create the depth map of the captured image, which is used to calculate position of the camera or the tracked objects [16,17].

Sensor Based Tracking

Besides optical sensors, other types of sensors can also be used for tracking.

¹<https://developer.microsoft.com/en-us/windows/kinect>

Multiple studies utilized Global Positioning System (GPS) for position tracking [18–21], but since GPS has accuracy limitations and can not provide rotational tracking, it was often used complementary to other tracking methods or when exact position had no critical importance.

Inertial Measurement Unit (IMU) sensors like accelerometers, gyroscopes and magnetometers can be used to track relative position of an object. Foxlin et al. lists several advantages of inertial sensors such as lack of range limitations, line-of-sight requirements and interference from any interference sources. They also can be as fast as required and have very low latency. On the other hand any noise or sensor bias easily cause gradual drift on their output, therefore like GPS sensors, IMU sensors are also used with other techniques [22].

Some of the earlier systems also used magnetic tracking [23].

Hybrid Tracking

To increase accuracy of the tracking, data from multiple sensors can be combined. These methods, also sometimes called "sensor fusion", are often used when sensor based tracking methods are not entirely reliable; e.g IMU sensors are clearly more error prone inside a moving vehicle.

Developments in mobile devices also created new approaches in hybrid methods. Not only do these devices usually have IMU and camera sensors together, but also have wireless communication capabilities. For instance, Yii et al. presented a system where smartphones send extracted features from the current camera image to a server, which compares these with the data collected by a stationary Microsoft Kinect [24].

2.1.4 Commercial Products

In recent years, popularity of AR has increased considerably in consumer market since technologies explained in previous sections are feasible with the current state of the art mobile processing units and sensors, some of which are already shipped with majority of smartphones and tablets.

Since many companies have not only different views of AR, but also different specialties and budgets, consumer products also show great variety in all aspects. For instance while Microsoft HoloLens generates tracking data internally with specialized sensors and serves it to the applications, Epson Moverio series smart glasses purely rely on third party libraries like Vuforia

for tracking. Comparison of some of the commercial AR products based on Azuma's characteristics can be seen at Table 2.1.

Device	Type	Display	Input	Tracking
Vuforia	Software	Depends on the device	Depends on the device	Marker and 3D model tracking, SLAM, IMU
ARKit	Software	Handheld display	Touch screen	SLAM with IMU
ARCore	Software	Handheld display	Touch screen	SLAM with IMU
Google Tango	Hardware	Handheld display	Touch screen	SLAM (depth camera) with IMU
Epson Moverio BT-300	Hardware	HMD	Touch pad	None, possible with Vuforia
Microsoft HoloLens	Hardware	HMD	Hand gestures, Voice	SLAM (depth camera) with IMU

Table 2.1: Comparison of some commercial AR products

The rest of this section discusses several AR solutions that are related to this thesis.

Tango platform and ARCore

Project Tango² is an AR platform developed by Google. It utilizes an IR projector, regular and RGB-IR cameras, as well as IMU sensors to create depth map of the environment and track the position of the device in the real world. Applications can get depth map by using Tango SDK.

After the initial development versions of a phone and a tablet, Google partnered

²[https://en.wikipedia.org/wiki/Tango_\(platform\)](https://en.wikipedia.org/wiki/Tango_(platform))

with other companies to release commercial products, and Lenovo Phab 2 Pro³ and Asus Zenfone AR⁴ were launched to customer market in 2016 and 2017.

Alongside with Tango, Google also developed ARCore. ARCore is a software framework for Android devices that can track device position and detect horizontal surfaces by using a SLAM algorithm which utilizes the camera and the IMU sensors of the smart phone. It can also estimate the environment's current lighting conditions, which allows developers to create more immersive applications. At the time of writing this thesis, ARCore only runs in handful of selected Android devices⁵.

Google deprecated Tango platform in favor of ARCore on March 1st 2018.

ARKit

ARKit is developed by Apple for iOS devices. Like ARCore, it also tracks the device position and lightning conditions in the environment. As of version 1.5, it can detect horizontal and vertical surfaces as well as image markers. In addition to these, it can also use the front cameras of iPhone X to track "position, topology, and expression of the user's face"⁶.

Vuforia SDK

Vuforia is one of the most popular tracking libraries for Augmented Reality. It supports detection and tracking of markers, custom images, custom 3D objects and also horizontal surfaces. One of the main reasons behind Vuforia's popularity is device support, Vuforia supports "the vast majority of smartphones and tablets running on Android, iOS and Windows 10" and also popular smart glasses such as Microsoft HoloLens, Epson Moverio BT300 and Vuzix M300⁷.

Moreover, Vuforia provides a wrapper around device tracking and horizontal surface detecting capabilities of ARKit and ARCore. This allows developers to use most optimized framework for the device with only using Vuforia.

³<https://www3.lenovo.com/us/en/smart-devices/-lenovo-smartphones/phab-series/Lenovo-Phab-2-Pro/p/WMD00000220>

⁴<https://www.asus.com/Phone/ZenFone-AR-ZS571KL>

⁵<https://developers.google.com/ar>

⁶<https://developer.apple.com/arkit>

⁷<https://www.vuforia.com/devices.html>

HoloLens

HoloLens smart glasses is an all-in-one AR solution from Microsoft.

Unlike some of the AR devices that outsource heavy processing to other stronger device, HoloLens is an independent and standalone computer that uses Intel Cherry Trail system on a chip (SOC) containing the Central Processing Unit (CPU) and Graphics Processing Unit (GPU). It features see-through lenses as displays and can automatically calibrate the distance of the pupils to focus content on the display.

HoloLens uses IMU sensors, four environment understanding cameras and a depth camera for tracking. It also includes light ambient sensor for detecting current lighting conditions and four microphones for speech recognition. Data from these sensors are processed at a custom-built Microsoft Holographic Processing Unit (HPU)⁸.

Besides the environment, HoloLens also tracks users hands and uses hand gestures as the main input source in addition to speech. A tracker can also be used instead of hands.

Microsoft provides APIs to developers for accessing spatial mapping, tracking information and user input⁹.

2.2 Unity Editor and Game Engine

Unity, developed by Unity Technologies, is a cross-platform development platform for 2D, 3D, VR and AR games and applications. It consists of state of the art graphics and physics engines as well as user friendly editor. It decreases development time vastly by providing easy to use features for rendering, physics and scripting¹⁰.

Initially announced only for Apple's OS X operating system, Unity has since been extended to target over 25 platforms, including Android, iOS and Universal Windows Platform¹¹. Moreover either directly working with Unity Technologies or indirectly by creating Unity compatible hardware and frameworks, cross-platform support is improved and strengthen by many different companies like Apple, Microsoft and Google. Thanks to this, developers can

⁸<https://developer.microsoft.com/en-us/windows/mixed-reality>

⁹<https://github.com/Microsoft/MixedRealityToolkit-Unity>

¹⁰<https://unity3d.com>

¹¹<https://unity3d.com/unity/features/multiplatform>

bring their games and applications to other platforms with relatively small changes. For instance, an AR game for Microsoft HoloLens can be ported to Google Tango platform by changing only the code specific to device features such as input handling or spatial mapping, instead of writing the app from scratch.

2.2.1 Augmented Reality and Unity

In practice, AR applications are based on the Azuma's characteristics that were explained in section 2.1. Therefore any application whether a simple game or a complex business solution, must render 3D graphics, handle user input and track positions of virtual objects and the user. Besides these, depending on the use case, application might also need some physics simulation to achieve more realistic user experience. And on the top of these, application logic itself should be built.

Unity has many features that simplify these steps and help developers to focus on the application itself instead of low level technical details.

- **3D Rendering:** Instead of creating a graphics rendering code by using low-level libraries such as OpenGL, Direct3D or Metal, developers can manage displayed objects by using Unity Editor interface. Displaying an object or changing texture of a 3D model is often only a single drag and drop action.
- **User Input:** Unity engine has built-in event system and supports inputs from keyboards, mouses, game controllers and touchscreens. Moreover devices that feature any other input method can utilize the event system to add input support, like HoloLens with hand gestures ¹².
- **Tracking:** While tracking heavily depends on the device hardware, provided frameworks by commercial products often support Unity and generate data compatible with Unity's coordinate system. Furthermore, both ARCore and ARKit have complete support for the Unity and Vuforia is officially included inside the Unity since October 2017.
- **Physics Simulation:** Physics bodies can be added to objects via Unity Editor and events like collisions can be listened by custom scripts without programming the physics calculations separately.

Moreover, fragmentation on the AR market explained in section 2.1.4 puts Unity

¹²<https://github.com/Microsoft/MixedRealityToolkit-Unity>

in a unique position; Unity can deploy applications to most of the devices on the market, and without Unity, developers should either develop their own engine or build separate applications for each platform by using different engines, which are neither trivial nor cheap.

2.2.2 Real-time Collaboration and Unity

Real-time collaboration requires consistent network connection between participants to keep workspaces of participants seamlessly synchronized. From the technical aspect, this is analogous with multiplayer gaming, where the game world is synchronized between different players.

Over the years multiple high-level real-time communication frameworks are developed for Unity to create multiplayer games, either by Unity Technology¹³ or third-party companies like Photon Engine¹⁴. These high-level frameworks allow configuring network parameters inside the Unity Editor, and hide low-level details like object serialization (i.e. converting an object to byte stream or creating an object from the received byte stream) from developers. Moreover, it is also possible to handle synchronization by totally custom code with using either transport layer APIs included in Unity engine or libraries provided by the programming language and target operating system.

Throughout this chapter, AR technologies and some of the consumer products were discussed. Any AR system must display 3D objects, get user input and track the positions of the user or the displaying device to keep the virtual objects in the correct location. Each of these functions can be realized by using different technologies, which often cause different user experiences. There are many off the shelf AR solutions, but as a result of the technological variety, consumer products also differ in many ways. From a developer's point of view, Unity Engine is very convenient for supporting as many different devices as possible. While this chapter provided a theoretical background, the following chapter presents the related work about the collaborative AR and identifies the gaps in the current literature.

¹³<https://docs.unity3d.com/Manual/UNetOverview.html>

¹⁴<https://www.photonengine.com>

Chapter 3

Related Work

In this chapter, state of the art collaborative Augmented Reality (AR) is presented. First section gives a general overview of these works, while the following sections provide details of the most related studies for this thesis.

3.1 General Overview

As early as mid-1990s, three pioneer studies focused on realizing collaboration in AR. Ahlers et al. implemented a collaborative design application using video recorders and desktop computers where users can place virtual furniture in the scene [25]. Rekimoto created a system called “Transvision”, where users can see virtual objects via handheld displays and interact with them using buttons on the display [26]. The most comprehensive study started around this time is “Studierstube”, where users can see 3D scientific data over the real world via head-mounted displays (HMDs) and projectors and manipulate the data using a two-handed pen and pad interface [27–29].

Several studies investigated collaborative AR in different contexts such as, scientific visualizations [30], education [31], museums [32, 33], archaeology [34], design [35], design error detection [36], construction [37, 38] remote guidance [39, 40], security and law enforcement [41, 42], emergency situations [43], navigation and over real-world annotation [19, 44–46].

Benefits of AR collaboration are documented in multiple studies. It does not require extra cognitive load [35], requires low mental work and can lead to faster task completion [36, 47], is superior to Virtual Reality (VR) while sharing work spaces [48], is better than a PC while viewing and manipulating 3D

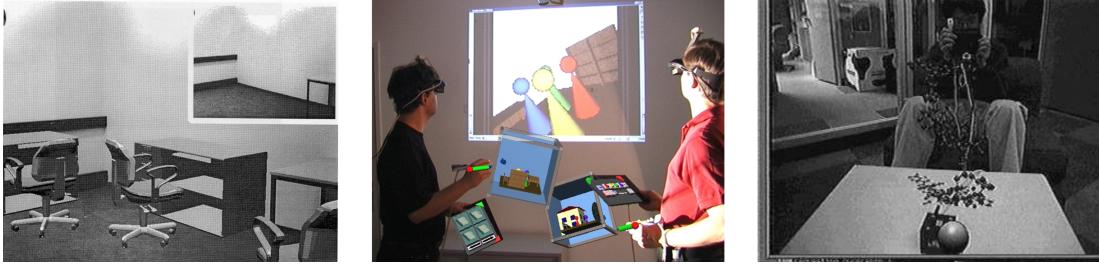


Figure 3.1: Left: System implemented by Ahlers et al. [25]. Middle: Studierstube [29]. Right: Transvision [26].

data [28] and is preferred or glorified by the users [49, 50]. On the other hand, Billinghurst et al. compared variations of face-to-face and remote settings with using AR and VR and found out users had the best performance when they can see each others avatars in VR [50].

Majority of research utilizes unique interaction techniques. Schmalstieg et al. and Reitmayr et al. used pen and touchpad as user input devices [27, 51]. Regenbrecht et al., Broll et al. and Regenbrecht and Wagner created collaborative AR systems with HMDs that use different tangible devices for user interactions [10, 52, 53]. Huynh et al. demonstrated an AR board game with tangibles for handheld devices [54]. Bauer et al. introduced a system, where remote expert can control a pointer that is connected to local user's HMD with a mouse [49]. Dong et al. developed a system where users can see models and animations through an HMD and interact with them using a mouse [55]. Butz et al. used optically tracked handheld pointer [9]. Multiple studies with handheld devices handled input with physical buttons of the device [26, 32, 33, 56]. Several others featured hand tracking and hand gestures [17, 39, 40, 57, 58].

Just a handful of the studies allowed collaboration between different device types. "Studierstube" was capable of displaying data both on HMDs and projected displays [29]. Datcu et al. used PCs and HMDs simultaneously during their different projects [42, 47, 59]. Butz et al. and Benko et al. introduced systems, where users can see and interact with 3D objects through HMDs and use handheld or tabletop devices for other related 2D tasks, such as changing the status of an object [9, 34]. In outdoor navigation guidance research of Höllerer et al., handheld devices are used for displaying maps, while users also wear HMDs [19]. Similarly, in "The Final TimeWarp" location-aware collaborative mobile AR game, Blum et al. used two tablets one of which is used for navigation with a map and the other displayed 3D virtual elements [60]. In "Implementation of god-like interaction techniques", Stafford et al. tested

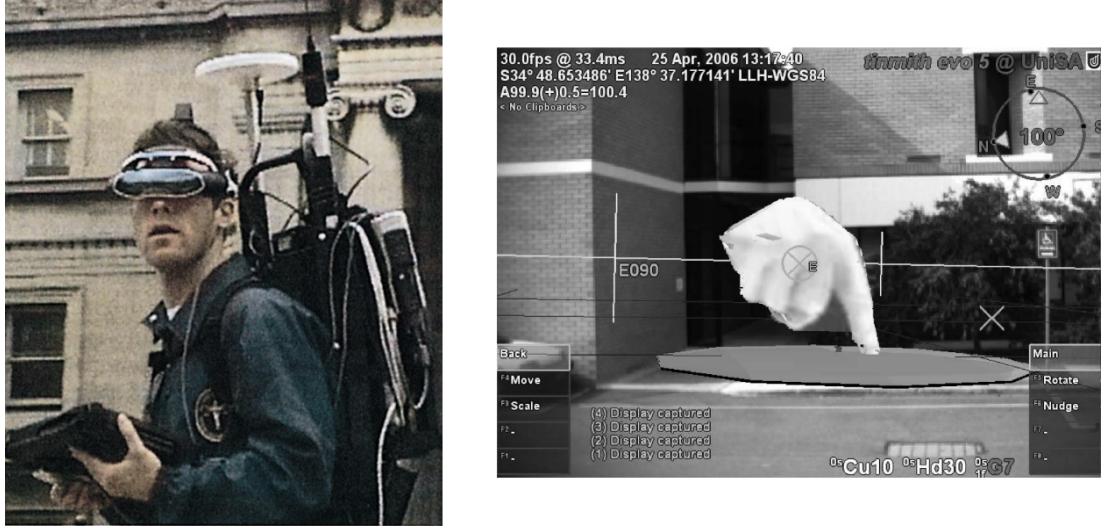


Figure 3.2: Left: Outdoor navigation guidance research of Höllerer et al. [19].
Right: A hand model in real world from Stafford et al's study [61].

a system that uses a tabletop display with a 3D scanner to scan objects and sends them to outdoor HMD wearing user's viewpoint [61]. Gauglitz et al. implemented an application that can infer 2D drawings made in touchscreen to 3D space of handheld AR devices [45]. Johnson et al. directly compared HMDs and handheld devices with users collaborating over physical tasks through video conferencing [62].

Naturally, user interactions and how they coordinate during the collaboration is also one of the research topics. Kirk and Fraser compared hand and pen during collaboration in 2D settings [57]. Kim et al. also compared drawing and pointing for collaboration and concluded drawing is easier and pointing requires more verbal communication [63]. In “Different Interaction Types in Augmented Reality”, Datcu et al. showed hands are not better than tangibles while interacting with the AR [64]. User studies of Chastine et al. showed users should have ability to point both virtually and physically and need reference points [65, 66]. Müller et al. investigated different reference points and found that people prefer and perform better with virtual reference points, compared to physical ones, while syncing their own spatial understanding with each other [67, 68]. Kiyokawa et al. compared different conditions to find best way to collaborate with HMDs and according to this study most natural interaction type is when users are facing each other and task space is between users [69].

In summary, very comprehensive work has been done on AR collaboration topic over the years that ranges from realizing such a system to comparing it with existing methods.

3.2 Studierstube

Studierstube is one of the pioneer collaborative AR systems, developed between 1996 and 2008. The system itself is designed as a complete framework to develop other AR applications; it provides not only graphical user interface APIs and widgets that can respond to 3D events, but also multi-tasking support as well as device driver abstractions to allow usage of different display and tracking devices. Because of these capabilities, it's even referred to as "Augmented Reality operating system" by the developers [29].

In Studierstube, users wear magnetically tracked see-through HMDs to display 3D environment, but desktop and wall displays are also supported. A magnetically tracked pen and panel interface is used for the interaction. The pen is used as a 6DoF pointer and users can perform gestures with the pen (e.g. drawing a circle) on the panel. Panel shows different content based on the running application, such as buttons, sliders or 3D widgets. The system also can display different data for different users which allows them to personalize the interface based on their needs [27,28].

Various AR applications are developed using the framework. Construct3D is designed to help high school students during mathematics and geometry education. Users, who wear HMDs, can create 3D shapes using the pen and check normal lines and intersection points by using buttons in the panel [70]. In another prototype for storyboard design, users wearing HMDs can manipulate and relocate objects, actors and camera in a scene which is represented by a 3D window. Another 2D projected screen is used to show view from the camera in the scene [29].

In essence, Studierstube demonstrated that collaborative AR is technically feasible and applicable in different contexts. Moreover, modular structure of the developed application allowed usage of different input and display modalities, although possible effects of these are not investigated during the project.

3.3 Handheld or Handsfree

Devices with different input and display modalities might be used during computer supported collaboration. Johnson et al. conducted a user study to investigate effects of using handheld or head-mounted displays on "collaborative behaviours, perceptions and performance" during remote collaboration. In the



Figure 3.3: Workers' environment in study of Johnson et al. [62].

study, participants are required to perform a construction task collaboratively via video conferencing. One user is called the "helper" and uses a desktop computer to instruct the worker user. The "worker" is in another room and equipped with either a Google Nexus 7 tablet or a Google Glass. In static setting, components to assemble are located in one table, and in dynamic one they are distributed amongst three.

Results indicate that in the dynamic setting, participants were able to complete tasks faster with HMD because it allowed them to use "more frequent directing commands and more proactive assistance". On the other hand for the static setting, helpers favored tablets over HMDs as oppose to workers. Johnson et al. suggest this is most likely due to lack of feedback for currently captured video in Google Glass interface. Users performed better with HMDs in general, but it is only marginally faster in dynamic task settings [62].

These results confirm that device type affects collaboration between users, both for performance and perception, which raises the question if these effects also happen during AR collaboration.

3.4 Virtual Objects As Spatial Cues

During communication, spatial coordination is often achieved by referring physical objects. Augmented reality allows collaborators to share virtual environments with virtual objects, which raises some questions in referring context, for instance would users prefer physical or virtual objects as reference points when co-located or how is coordination affected when users are not co-located and remotely collaborating. Müller et al. conducted user studies to answer these questions.

To understand spatial cues in the co-located collaborative Augmented Reality, participants are asked to perform object identification and positioning tasks with Project Tango tablets. In object identification task, a version of memory card game¹ is used, where different symbols are attached to floating cubes which reveal the symbol when selected. For the object positioning, dyads are asked to position cubes based on the positions in the previous game. For both tasks, different physical cues such as a waste paper basket and a clothes hook and virtual cues such as a vending machine are used. Results showed participants extensively used virtual objects as spatial cues over physical ones, and existence of virtual objects not only positively affected their communication behavior but also decreased user task load while improving user experience [67].



Figure 3.4: Visual objects used as spatial cues in the study of Müller et al. as they are displayed on the tablet display [67].

Müller et al. repeated the object identification task in remote settings. In this experiment, virtual objects or "shared virtual landmark (SVL)", like potted tree

¹[https://en.wikipedia.org/wiki/Concentration_\(game\)](https://en.wikipedia.org/wiki/Concentration_(game))

and plant are used as spatial cues. Users are asked to play the game with and without SVLs in different sessions. Results show SVLs "reduced the occurrence of ambiguous deictic expressions which could cause conflict situations" and "participants reported a significantly increased user experience and favored the SVL condition" [68].

Work of Müller et al. clearly shows collaborative AR experiences should provide virtual landmarks to prevent possible communication issues between collaborators. On the other hand, this research does not clarify if these landmarks are also essential for the collaboration with HMDs, since HMDs also change users' perception of the environment.

This chapter presented the related work. AR collaboration is a very popular research topic and very comprehensive work has been done on this area. Work of Johnson et al. established that different display modalities affect the collaboration in remote video conference settings. On the other hand, majority of the research about AR collaboration, even the ones that support different input and display modalities did not investigate possible effects of these. In the following chapter, objectives of this thesis and the questions it strives to answer to fill this gap in the current literature are introduced.

Chapter 4

Objective and Research Questions

As explained in section 2.1, Augmented Reality (AR) systems are essentially a combination of display, input and tracking subsystems. There are multiple ways to realize these subsystems. The display device can be handheld, head-mounted or spatial (see section 2.1.1). There are several input techniques, including traditional techniques like the keyboard and mouse, tangible interfaces and hand gestures (see section 2.1.2). The tracking method might be optical, sensor based or combination of both, and these methods also have subcategories (see section 2.1.3). These variety reflects on the consumer market too and consumer hardware and software solutions vary in many aspects (see section 2.1.4).

This variation of the consumer devices also affects the developed applications and, inevitably, user experience. To reach the maximum audience, developers try to support as many devices as possible for their applications. However, user experience of the same application in different devices might vastly change. Even a simple action, like pressing a button, has major differences in different devices, which are illustrated and explained in Figures 4.1, 4.2 and 4.3. Furthermore, even for an application that works only with voice commands, device type affects the user experience deeply; for example if an AR system assists user with overlaying information on an engine, while user is maintaining the engine with both hands, a head mounted display (HMD) would be much more convenient compared to a handheld one.

Amount of work referenced in chapter 3 clearly indicates that collaborative AR is a very popular research topic. Besides the scientific community, it is also

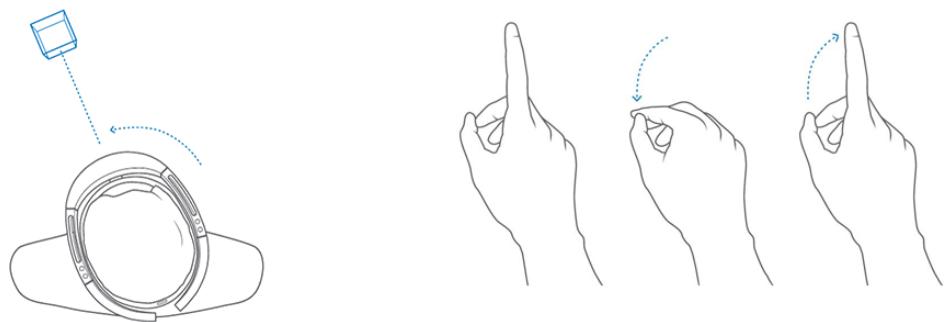


Figure 4.1: How to press a button with a HoloLens. Left: User turns her head to gaze at the button. Right: User makes an air tap gesture; raises her hand with index finger pointing upwards, flexes her index finger down and then backs it up again. Images are taken from <https://support.microsoft.com/> and used with permission from Microsoft.

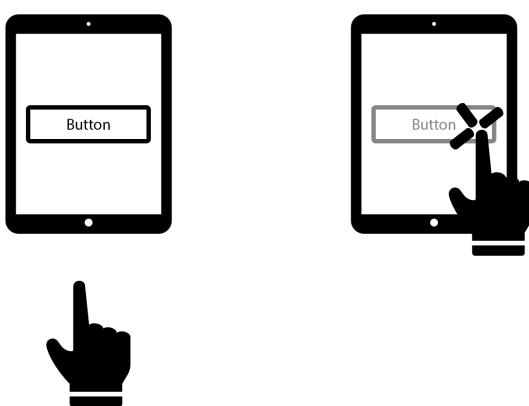


Figure 4.2: How to press a button with a tablet. Left: User gets ready to tap to the screen. Right: User taps the button with a finger.

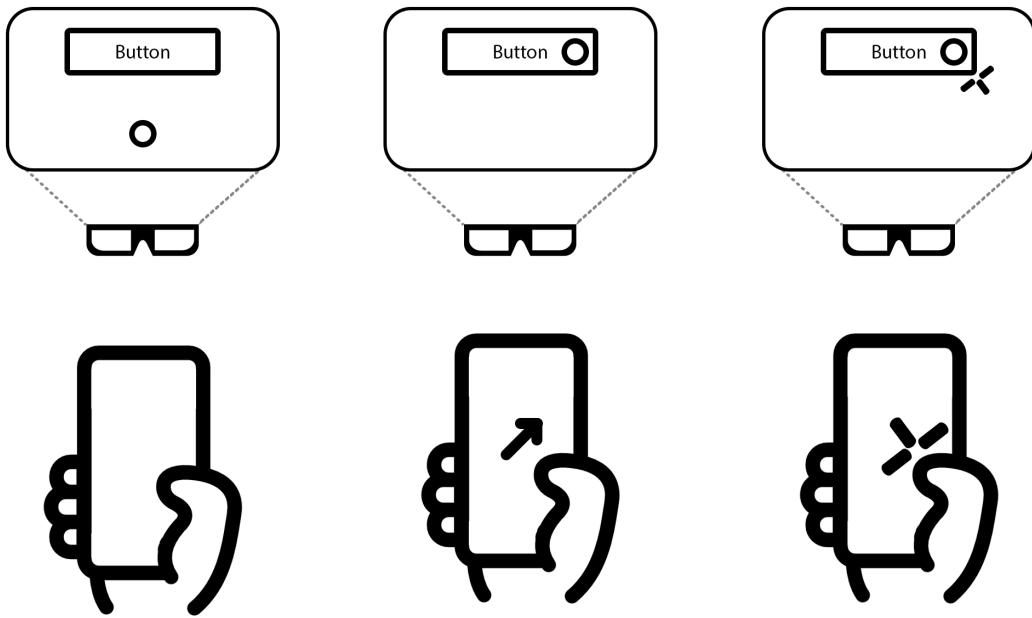


Figure 4.3: How to press a button with Epson Moverio BT-300. Left: User wears the headset and holds the touchpad. Middle: User moves the cursor over the button using the touchpad. Right: User taps onto the touchpad.

gaining interest in consumer products; for instance ScopeAR¹ is providing cross platform AR applications for real-time guidance and Microsoft is promoting HoloLens with emphasis on real-time collaboration capabilities and providing all the necessary software frameworks to the developers².

Work of Johnson et al. established that different display modalities affect the collaboration in remote video conference settings [62]. However, effects of different display, input or tracking modalities to real-time collaboration in AR environments are not thoroughly researched. In addition, Müller et al. suggests AR collaboration highly benefits from providing additional virtual objects in the environment [67], though their tests explicitly used handheld tablets and it is not clear if HMDs would produce different results, since users experience virtual environment in a different way with them.

Based on this background, this thesis strives to answer following research questions through a user study:

¹<https://www.scopear.com>

²https://developer.microsoft.com/en-us/windows/mixed-reality/shared_experiences_in_mixed_reality

- **Q1:** How is the group coordination and performance affected by different device types during the real-time collaboration for different tasks?
- **Q2:** Do users prefer virtual or real reference points for coordination while collaborating with different device types?

This chapter outlined the motivation of this thesis and the questions it seeks to research through a user study. The following chapter elaborates on the design of the user study.

Chapter 5

User Study Design

In chapter 4, the research questions this thesis aims to answer are introduced. These questions shape the design and the implementation of the user study. This chapter discusses the design process of the user study in detail. Section 5.1 elaborates on the requirements of the study to answer research questions. Section 5.2 presents the ideas for the user study and low fidelity prototype iterations of them. And finally section 5.2.3 lists the requirements for the Augmented Reality (AR) implementation.

5.1 User Study Requirements and Decisions

To assert that the user study corresponded with the research questions, four requirements were derived:

1. Different device types should be used.
2. Tasks should require users to collaborate and communicate.
3. There should be at least two different identifiable tasks.
4. Task should require participants to use spatial references.

Requirements 1 and 2 were necessitated by the general theme of the topic and requirement 3 and 4 were inferred from the research questions that were introduced in chapter 4.

Besides these theoretical requirements, there were also several practical decisions and concerns, which were interconnected. One method to answer the research questions of this thesis is collecting and analyzing quantitative data,

such as task completion times and amounts of vocal communication between users. Compared to qualitative methods, collecting quantitative data requires a higher number of participants. However number of potential participants for this study was limited, therefore using a within group design would be the most beneficial. Moreover, the duration of the user study had to be as short as possible to minimize possible scheduling conflicts and appeal to more volunteers. On the other hand, each task had to be repeated with all possible devices and participant permutations due to the study's within group design, thus using more than two device types would not be feasible.

Current tracking technologies are often hidden from the users and have minimal impact on the user experience. Therefore display and input modalities were prioritized during the device selection. As a result of its highly accurate tracking and native hand gesture support, Microsoft HoloLens was selected as the HMD - hand gesture input device type. Initially during the concept design stage, Google Tango had been chosen as the handheld - touch input device type since it also provides very high accuracy tracking, but later during the implementation, because of the technical reasons (see section 6.1) it was switched with 10.5-inch iPad Pro 2017.

There are no clear limitations for spatial locations of users based on the research questions, and they could be located separately or together. However, considering requirements 2 and 4, as well as practicality, it was decided that having a face to face setting would be more advantageous to answer the research questions.

In summary, during this thesis a quantitative, within group user study was designed. It followed the listed requirements and features a collaborative real-time AR application that run on Microsoft HoloLens and iPad.

5.2 Low Fidelity Prototypes

Based on the requirements and preliminary decisions listed in section 5.1, several ideas were developed. To decide which one to implement, these ideas were converted to low fidelity prototypes and improved or eliminated with each iteration. This section presents these ideas and the iteration process.

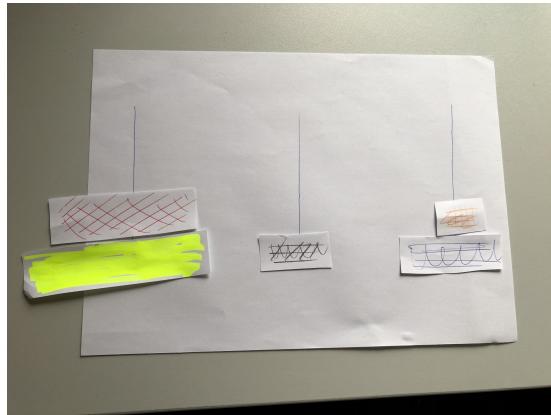


Figure 5.1: Paper prototype implementation for the Tower of Hanoi game.

5.2.1 First Iteration

During the first iteration, four different tasks were defined and examined. Two of these ideas were eliminated but some elements from those ideas were merged into the others during the next iteration.

Tower of Hanoi

Tower of Hanoi is a mathematical puzzle in which the player tries to move a stack of different sized disks from one stick to another. The player can only move one disk at a time and cannot place a disk on top of a smaller disk. For this user study, a second player was added to this game and users were allowed to move disks only if both of the users agree to the move. Since this game lacks spatial references besides its own elements, it was initially planned to be used along with "*finding objects*" task.

A very simple paper prototype, as displayed in Figure 5.1, quickly revealed that novelty in each repetition is very low for this game. Therefore, mainly due to this reason, *Tower of Hanoi* idea was repealed.

Finding Objects

The *finding objects* idea was designed to test spatial reference points. In this task, objects with different shapes are scattered randomly around the environment. One user sees which object is randomly selected by the system and directs the other user to collect it. On the next iteration, this idea was integrated inside the "*construction*" task.

Construction

The *construction* idea was mainly inspired by study of Johnson et al., where users assembled a model via video conferencing [62]. In this version, however, the physical model to assemble was replaced by a 3D model. Like the original study, only one user can see the instructions and the other user can manipulate the objects. A simple, non-formal test with LEGO bricks showed that users engage in verbal and gestural communication to coordinate while assembling LEGO bricks together, therefore it was carried over to the next iteration.

Find And Move

The *find and move* idea is a combination of two consecutive tasks. Preliminarily, 3D models, like pyramids and spheres, are scattered inside the test environment randomly in pairs. The position of each model appears to the users as a cube, but the real shape is hidden as default. When a user selects a cube, the other user can see the actual shape of the model. When both users select models with matching shapes, the users need to move and collect the models at some predetermined position. If the models do not match, they turn back into cubes.

To simulate different views of the same environment and see if this example was suitable to answer the research questions, a paper prototype with color filters was prepared. Two paper anaglyph glasses were cut in the middle, and matching colors were combined. Several shapes were printed out in a way that they were only distinguishable with one of the two glasses. Examples of this can be seen in Figure 5.2

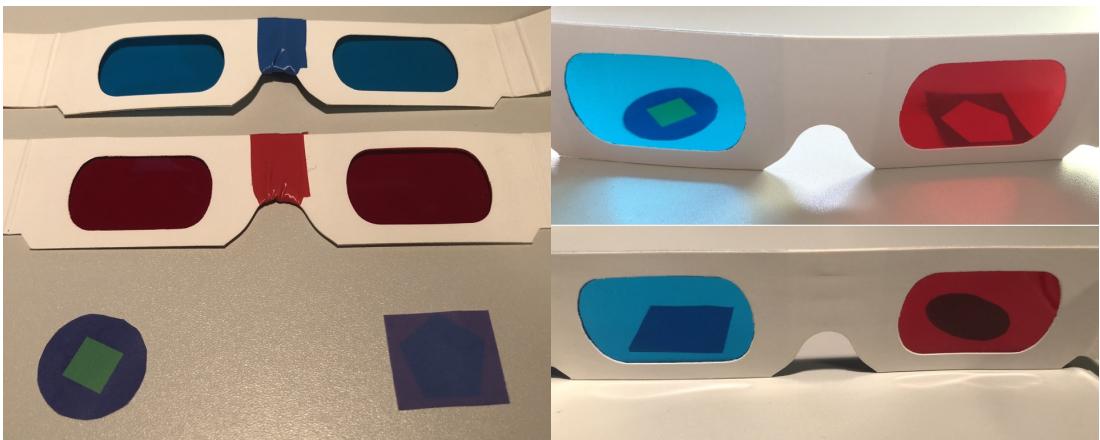


Figure 5.2: Left: Color filtered glasses and example shapes. Top-Right: When shapes are distinguishable. Bottom-Right: When shapes are hidden.

While these tasks require a high amount of collaboration and spatial references, the effects of personalized views on collaborative AR are not within the scope of this thesis. Hence, after some consideration, these tasks were replaced with a version of study of the Müller et al. about virtual landmarks [67] in the next iteration.

5.2.2 Second Iteration

During the second iteration, previous ideas were reshaped. Afterwards, a semi-formal user study with low-fidelity prototypes was conducted to make the final decision about the main user study.

The *construction* task was used as in the first iteration, but, additionally, the *finding objects* idea was integrated as a pre-task. Before the construction itself, users are asked to locate and collect the required 3D objects that are randomly distributed into the test room. *Find and move* was redesigned as a memory cards game¹, where users try to find matching symbol pairs. White 3D cubes are scattered in the test environment. Each cube reveals a symbol when any of the users selects the cube. A user can only select one cube, and, unlike the previous idea, the symbol of the selected cube is visible to both users. When both users select a cube, the displayed symbols are compared; if they match, the cubes are removed from the game and if not, the symbols are hidden again. Once the users have found all of the matching pairs, they are asked to reposition the symbols to their original positions.

User Study

To determine the most suitable task to implement, a semi-formal user study with four voluntary participants was conducted. For the construction tasks, LEGO bricks were randomly placed in the test room and participants were given a list of required pieces, as well as instructions to build a model, which can be seen in Figure 5.3. For the memory cards game, 10 Unicode symbols were printed in pairs and, also, randomly placed in the test room. Users were asked to flip papers to reveal symbols. In the second part, participants were asked to put the papers back to their original locations. Subsequently, users were asked about their impressions regarding the tasks and collaboration in a small interview. The complete design of this study can be seen in appendix A.

¹[https://en.wikipedia.org/wiki/Concentration_\(game\)](https://en.wikipedia.org/wiki/Concentration_(game))



Figure 5.3: The final model users built with the given instructions during the construction task.

Afterwards, the spatial expressions of participants were analyzed. Each instance of deictic speech (phrases that "can't be fully understood with speech alone" like "here" and "over there" [69]), references to regions inside the room (e.g. "next to wall", "center of the table"), other test objects and other objects inside the room, as well as pointing gestures without speech were counted. The analysis immediately indicated participants used nearly twice as many expressions in the memory cards game (36 and 64 instances) than in construction task. Participants also mentioned that communication during the construction tasks felt more one-sided than in the matching cards game, since one user kept giving the orders. Moreover, participants found repositioning the symbols more cooperative and harder than the other tasks.

Although it would have been possible to modify construction task to have more balanced communication between the users based on the feedback, the matching cards game was selected for the main implementation, since it already had all the required properties.

5.2.3 Requirements for the Implementation

Before the implementation of the memory cards game for the AR devices, parts of the user study which directly concern the implementation were also designed. The theme of the study is highly inspired from the work of Müller et al. [67].

The game was separated into two main tasks:

- **Object identification:** White cubes are distributed in the environment around physical and virtual objects. These cubes reveal their symbol if selected. One user can select one cube at a time. If users select the cubes with matching symbols, cubes are removed, if not, their symbols are hidden again.
- **Object positioning:** Users are asked to place the cubes back into their original positions.

Independent variables of the study were identified as following:

- **Device type:** Device type is the main research point of the study. As the section 5.1 stated, the study had been designed as within group experiment, and HoloLens and a tablet were selected as the main devices. Therefore all users were planned to complete experiment tasks using all possible combinations of two HoloLenses and tablets.
- **Physical and virtual objects in the environment:** Both physical and virtual objects in the test environment had major importance since reference point usage was also investigated. These objects were held constant to have comparable data for the reference points by different user groups.
- **Positions and symbols of the cubes:** Because of the within group design, users had to complete tasks at least 5 times. Keeping positions and symbols the same between all of these iterations would be prone to create undesired learning effects. For instance users could have confused position and symbol of a cube with an another iteration. To avoid these issues, the cube positions were shuffled and symbol set was changed for every repetition. On the other hand, these positions and symbol sets were held constant between different participant groups.

Based on the research questions and the tasks, the data that was to be collected through the application, or dependent variables, was decided as following:

- **Task completion times:** For the *object identification*, the task completion time is defined as the the time between the cubes being displayed and the last matching pair being selected. For the *object positioning*, it is defined as the time between the first cube pair being displayed and the last cube pair being placed. Any significant difference between device combinations clearly shows collaboration is effected by the device type.
- **Error rate:** For the *object identification* task, the error rate is defined as the number of times a cube was selected with a non-matching pair after it already has been selected once. For the *object positioning* the error rate is defined as the distance of the cube from the original position. Higher error

rates indicate that the device combination is not suitable for collaborative work.

- **Participants' communication:** The spatial expressions used by the participants can directly answer what kind of objects people refer to more. All verbal and gestural expressions are also directly connected with the participants' coordination.

In addition, users should have completed the tasks twice as a tutorial with both devices to decrease the possible learning effects.

In summary, AR implementation should support several features to correspond with the user study:

- Collaborative object identification and positioning tasks
- Running on HoloLens and iPad
- Displaying static virtual objects besides the cubes
- Shuffling positions and symbols of the cubes in each iteration, including the tutorial, but keeping them the same for different groups
- Logging durations of tasks, as well as error rates

During this chapter, the design process of the user study was explained. Based on the research questions, as well as the practical reasons, it was decided to have a quantitative, within group user study with a collaborative AR application that runs on Microsoft HoloLens and iPad. Several ideas were developed as the main theme of the user study and low-fidelity prototypes were produced to test these ideas. As a result of the testing, a version of the memory cards game was selected. Before the implementation stage, parts of the user study which directly concern the implementation were also designed. The game was separated into two tasks and dependent and independent variables were identified. Based on these, the features of the implementation were outlined. Details of the implementation are discussed in the next chapter.

Chapter 6

Implementation

In order to conduct the user study, an Augmented Reality (AR) application was developed based on the requirements that outlined in section 5.2.3. This chapter explains the implementation of this application.

Section 6.1 outlines the general architecture of the project, including the software tools that were utilized during the development. The implemented application can be separated into three main components: AR, networking, and the application logic. Further, AR technologies that were used during this thesis and how they work together with different devices are presented in section 6.2. Subsequently, section 6.3 describes the network implementation. Finally, section 6.4 explains the implementation of tasks and how AR and networking components are utilized to realize these tasks in a collaborative AR environment.

6.1 Overview and Development Environment

As detailed in section 2.2, Unity Engine provides a developer friendly environment for the AR application development and supports many existing platforms and devices, including HoloLens and iOS. Because of these reasons, Unity was selected as the main development environment.

At a preliminary stage, Google Tango was chosen as the handheld device for the implementation. However, after the introduction of ARCore, Google officially discontinued the Tango platform on 1 March 2018. While it would still be technically possible to develop an application for Tango, the lack of support from Google, Unity, and Vuforia made it overall less favorable. As a result, 10.5-inch iPad Pro 2017 was selected as the handheld device.

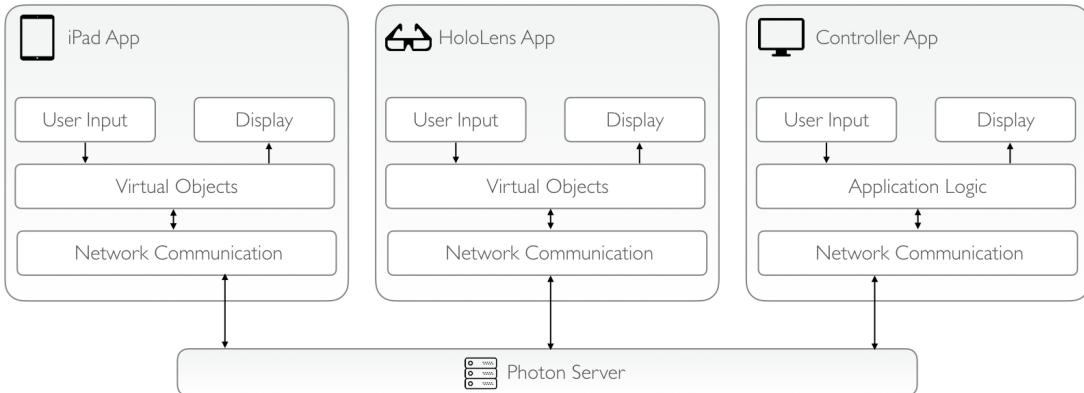


Figure 6.1: Simplified system architecture. Arrows indicate data flow directions.

The developed application consists of three different clients, comprising client applications for HoloLens and iOS, which users can use to complete tasks, as well as a controller application that controls the state of the tasks that can be run on both Windows and MacOS. These clients used Photon software development kit (SDK) for networking and communication through a Photon self-hosted server¹. In this study, the Photon server was merely used to transmit data between clients. A general system architecture is presented in Figure 6.1.

Virtual objects were displayed at the same position on both HoloLenses and iPads by using image markers. Image markers were detected with Vuforia on both device types. Vuforia was also used for device tracking on iOS.

HoloLens applications can only be deployed to the device from a Windows 10 operating system with Microsoft Visual Studio. Therefore, Visual Studio Community 2017 was used as the main code editor for the HoloLens client. A similar limitation is also exist for iOS and MacOS; iOS applications can only be deployed by using Xcode on MacOS. Hence, the iOS application was deployed to iPads via Xcode 9.2 on MacOS High Sierra. However, since Xcode does not support code editing for Unity projects, Visual Studio Code was used for this purpose on MacOS. Since switching platforms on Unity takes too long to be feasible in long-term development, different projects were created for each target client and files they had to share were synchronized between these projects.

¹<https://www.photonengine.com/en-US/OnPremise>

6.2 Augmented Reality

The main challenge of the AR implementation was to display and move objects relative to the real-world coordinates and not the ones generated by the devices. Virtual objects were displayed at the same position by using image markers, which were detected with Vuforia SDK. While moving an object, the position was updated based on its relative position and rotation to an image marker. The following sections explain these problems and their solutions in more detail.

6.2.1 Displaying Objects

Since HoloLens provides all the tracking information to the application itself, there was no particular need to use additional libraries for tracking on HoloLens. On the other hand during the implementation stage, the two main options to realize AR on iOS were ARKit 1.0 and Vuforia 7.0. ARKit 1.0 has no image marker detection (which was added later in version 1.5) and does not work together with Vuforia. Moreover, the Vuforia tracking already employs ARKit for tracking when possible (i.e. iOS 10 and below does not support ARKit), and therefore Vuforia was used for tracking on iOS.

Both Vuforia and HoloLens set the initial location and rotation of the device as the origin of the coordinate system when the application is started. Because of this reason, any device that runs the same application generates a completely different coordinate system. Furthermore, the environment information gathered by these devices are also not compatible. While HoloLens creates a 3D map of the environment, Vuforia can only detect horizontal surfaces and does not provide any specific information about the surfaces, like their size. In order to overcome these differences and ensure that virtual objects were displayed at the same physical location in different devices, image markers were used. The virtual objects were placed inside the physical environment relative to these markers.

Vuforia tracking allows developers to add anchor points into the environment, which are tracked more excessively. These anchors can then be added to arbitrary points which are called *mid-air anchors*, or onto the detected planes, which are called *plane anchors*. Since the device tracking and image markers are independent in Vuforia 7.0, objects connected to an image marker are only rendered while the marker is inside the camera view and detectable, which is highly undesirable for this project. To be able to render objects at the correct positions at all times, a mid-air anchor was initially added at the position of



Figure 6.2: Displaying objects at the same positions. Left: View of the test room on HoloLens. Right: View of the test room on iPad.

the image marker when it was detected, and objects were moved under that anchor. To allow error correction, the position and rotation of the anchor was updated every time a marker was re-detected. Further tests revealed mid-air anchors are not reliable without the plane detection, therefore plane anchors were used instead.

As Vuforia image markers directly support HoloLens tracking, no additional steps were required for the HoloLens. After an image marker was detected, the tracking of spawned objects was handled by the HoloLens, regardless of the marker's detection status.

6.2.2 Moving Objects

The second task of the user study required users to move objects. The problem of non-aligned coordinate systems in different devices also existed in this scenario, therefore only the relative position of the moving object to image markers was synchronized between the clients.

In Unity, objects are placed inside the scenes in a tree hierarchy. Each object consists of different components which control the object's visualization and behaviour. All objects must have a "Transform" component that keeps track of the position, rotation and scale of the object. The Transform component also tracks the local position and rotation of the object, which are relative values of these to the parent object.

As explained in section 6.2.1, objects were moved under the plane anchors when an image target had detected. To display moving objects physically at the same location on different devices, only the local position and rotation of these objects' Transform component were shared with the other clients. Even though the position of the moving object might shift by up to 30 cm in any direction on

the other device, these displacements were visually hard to detect and deemed good enough for the user study.

The accuracy of object positions during their movements decreases when the distance between the parent and the moving object is increased, especially on the iPad. To minimize this effect, moveable cubes in the second task were spawned under an image marker that had been placed in the middle of the test room. On iOS, a mid-air anchor was added to the new position of the cube to reduce further dispositions.

6.3 Networking

Networking is the backbone of the real-time collaboration. Without networking, it would not be possible to update all clients and display the most recent information automatically. Following sections elaborate on the decision of the networking library and the implementation.

6.3.1 Photon Networking

Trivial but practical reasons played major roles in selecting the networking framework, rather than their technical capabilities. During the initial feasibility research, Unity Networking was excluded from the potential framework list, since it requires all clients to be compiled with the same Unity version, which is not possible for Tango and HoloLens with Vuforia libraries. At a later stage, when the iPad had been selected as the tablet, Unity Networking was preferred for networking since version limitation does not apply for this device combination and it is directly integrated into Unity Editor. However, as explained in section 6.1, the overall application consists of three different projects and Unity Networking is not designed to support this case. Although it can connect clients from different projects, it does not run stable enough and requires lots of workarounds. Because of these reasons, Photon Networking, which can be run between different projects and fully supports both iOS and HoloLens, was selected amongst available frameworks.

Photon Networking is specifically designed for multiplayer gaming. Different clients can connect to the server and create or join rooms to start games. With this room system, multiple games can be run simultaneously on the same server without affecting one another. Exit Games, the developer of the Photon Networking, offers different versions of the server. The main difference between

these versions is where the server is running, Exit Games offers cloud solutions, as well as self-hosted servers.

When a client is connected to a room, different objects can synchronize variables of components and send messages to other clients. All objects with a network connection has an owner client. Only the owner of an object can update the variable values of that object. Different clients can request ownership of specific objects from others.

6.3.2 Networking Implementation

The first major decision regarding the networking was the selection of the server type. In order to avoid any internet connectivity related issues, the self-hosted server was used in order to avoid any internet connectivity related issues. No modifications were made on the server, and it mainly managed the client connections and data transmissions.

The object and variables to synchronize through the different devices were directly connected to application logic. The following information was shared to keep all clients in sync during the game:

- Types of the connected clients (Controller, iOS, HoloLens)
- Active task
- Identification task: Positions and symbols of the cubes
- Identification task: User selection, result of the selection
- Positioning task: Owners of the cubes, ownership requests
- Positioning task: Positions of the moved cubes

The clients used Photon Networking libraries to connect to the server and communicate among each other. One client, named *controller*, took the responsibility for managing the tasks and the other clients. It also owned all objects related to the tasks. Clients sent user actions to the *controller* which then updated related variables and sent necessary commands based on these actions. For instance, if users selected matching cubes, clients took no action until the controller sent a command to remove selected cubes. This ensured that all clients were displaying the same objects with the same status. The only exception to this was when moving cubes during the positioning task, where clients took ownership of the cubes from the controller and updated their positions while users were moving them. An example network communication flow is illustrated in Figure 6.3.

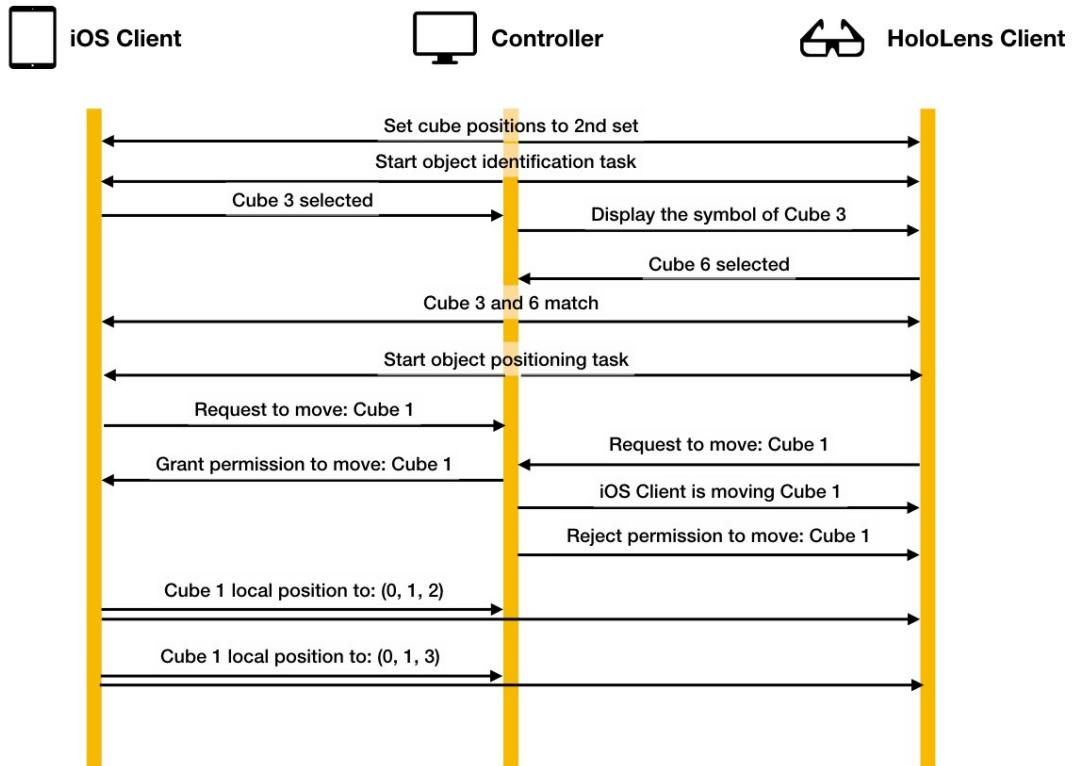


Figure 6.3: An example of network communication flow. Time flows from top to bottom.

6.4 Real-time Collaborative AR Application

The application was developed based on the requirements of the user study that had been introduced in section 5.2.3. As previously explained in the section 6.1, the application has iOS and HoloLens implementations, as well as a controller app. In this section, iOS and HoloLens implementations were referred to as *clients*, and the controller client was referred to as *controller*.

The application run cycle during the user study was planned as the following: after a short initialization process where all *clients* are connected to the server and the respective image markers are scanned, participants are asked to complete the tutorial twice. Subsequently, they are asked to complete the tasks with the all possible device combinations. The *controller* which also logs the important data during the study, switches between the active tasks. The following sections elaborate on these steps.

6.4.1 Initialization

The initialization of the app can be subdivided into two parts: network connection and device preparation.

The network connection was performed automatically by all clients without any user interaction. The server address and the room information was shared by all *clients* and the *controller*. Upon connecting to the server, the *controller* creates the room. *Clients* join to the room if it exists, or wait until it is created by the *controller*.

The device preparation is differed on both devices because of the different requirements. As explained in section 6.2.1, the iPads need to detect a plane first to have the best tracking possible. Therefore, the image marker detection was only enabled after a plane was found. To inform the user, a short message was displayed on-screen, as shown in Figure 6.4.

The proper adjustment of the HoloLens display is often difficult for people when they are using the device for the first time. Moreover, the individual nature of the device makes detecting problems related to the device orientation and helping users harder. For this purpose, the HoloLens client displayed a frame on the borders of its field of view the first time the application started, aiding users to arrange the glasses accordingly or explain the issues more clearly. Furthermore, a floating button was displayed in front of users to introduce the relevant hand gesture. Once it has been pressed, the frame that served as an adjustment aid disappeared and enabled the image marker detection. The frame and the button can be seen in Figure 6.4.

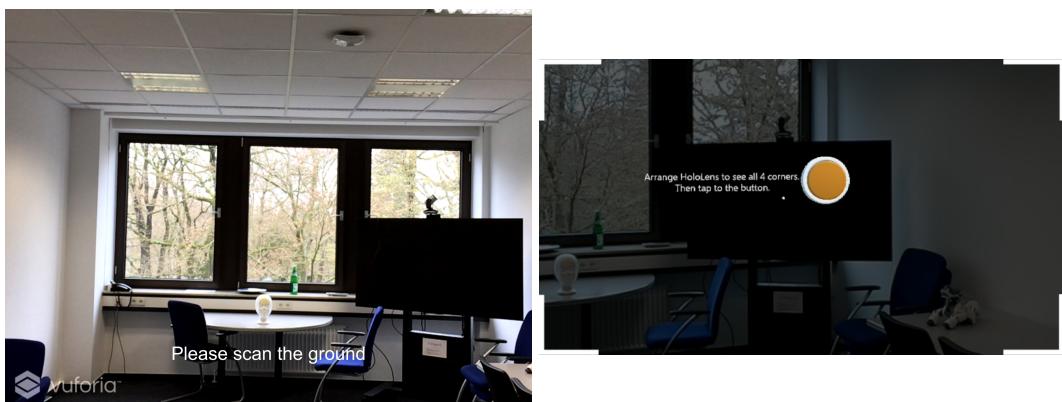


Figure 6.4: Preparing devices for the tasks. Left: Asking users to scanning ground first on the iPad. Right: Display frame and test button on HoloLens.

After a plane had been detected with the iPad and the button had been pressed

by the wearer of the HoloLens, the *clients* were required to scan all image markers that present inside the test room. Overall, five different image markers were used during the study. The markers were printed on DIN-A4-sized sheets in order to increase the detection height, enabling the participants to complete image detection process from a comfortable distance. Figure 6.5 shows all objects that spawned over these markers.

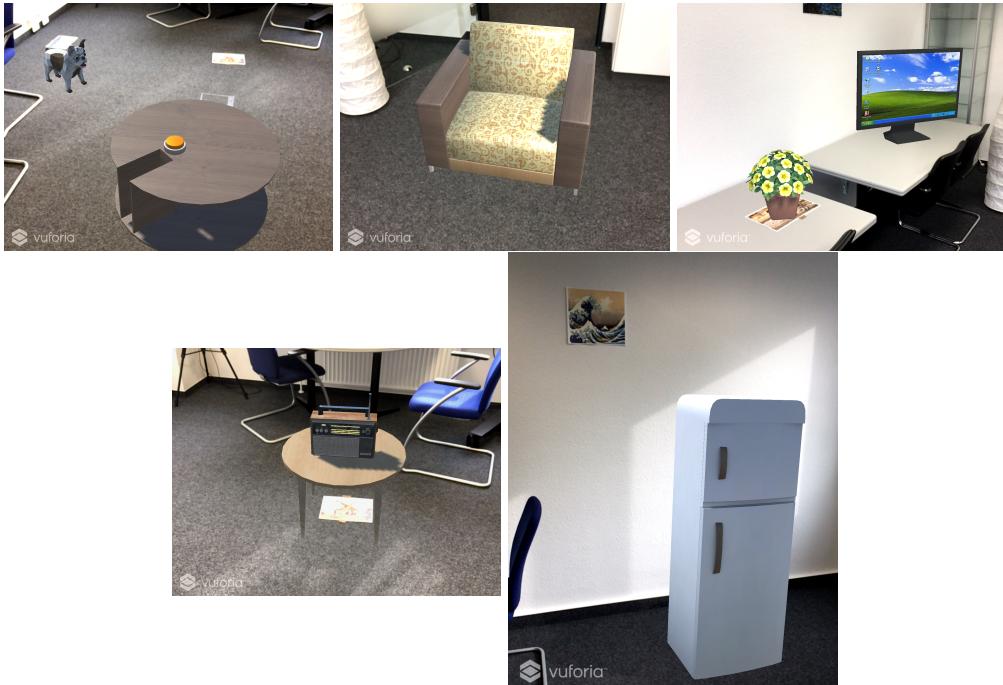


Figure 6.5: Virtual objects spawned over the image markers as seen on the iPad.

6.4.2 Object Identification Task

During the *object identification* task, participants played a matching symbols game with 3D boxes. The implementation followed the rules and requirements that were previously explained in section 5.2.3.

During this task, eight pairs of white cubes were displayed to the participants. The edge length of the cubes was set around 30cm to avoid congestion inside the test room. For each image marker, 27 different possible cube positions were designated as combinations of relative positions;

- Behind, center and front with 70 cm intervals, center is the marker
- Left, center and right with 70 cm intervals, center is the marker

- Bottom, center and top with 60 cm intervals, bottom is the marker

For each image marker, four cubes were placed at one of these combinations. For each repetition of the game, using a different device combination, the cube positions were changed. The positions were determined manually based on the other image markers as well as the physical objects inside the test room; they were not randomized to avoid possible overlaps with both virtual and physical objects. These positions were kept the same between different participant groups. The *controller* app was responsible for changing the active set. An example of the position set with symbols can be seen in Figure 6.6.



Figure 6.6: An example of the cube positions and their assigned symbols. All symbols are displayed here only for demonstration.

Just as with the cube positions, a different set of symbols were used for each repetition. These symbols were required to be easily distinguishable from each other, as well as recognizable with a single blink of an eye. The symbols were selected amongst Unicode characters for the tutorial, and emoji for the rest, as they meet the listed features. The symbol of each cube was pre-determined for each repetition and assigned by the *controller* when the position set changes.

The cubes were selected with default selection input on both devices; tapping on the screen at a cube position on the iPad and a tapping gesture that is visualized in Figure 4.1 on HoloLens. When a *client* had selected a cube, a message containing this information was sent to the *controller*, which then sent the proper response to the *clients*. The decision flow for the proper response to

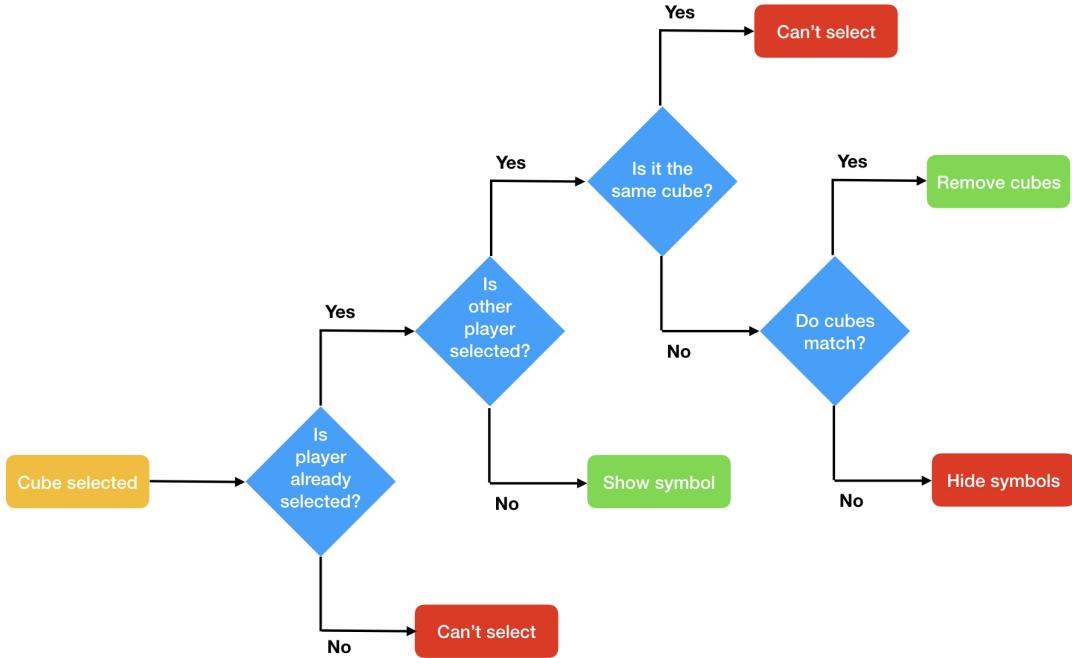


Figure 6.7: Decision flow chart for cube selection

the selection is illustrated in Figure 6.7.

When two cubes were selected, regardless whether or not they were a matching pair, the symbols remained visible for a second before they were hidden with an animation. In addition, a different audio feedback was given for each response of the *controller*.

6.4.3 Object Positioning Task

During the second task, the users were expected to return the cubes back to their original positions based on their symbols. As explained in the section 6.2.2, the location of the moving cubes were synchronized based on their relative position to a centrally located image marker.

After the users found all matching symbol pairs, the task was manually switched to the *object positioning* via the *controller* app. One randomly selected cube pair was automatically displayed on a coffee table which can be seen at Figure 6.5 on left-top. Users could receive further cubes by pressing at the button on the

table.

The cubes were selected via default selection input on both devices. When a user had selected a cube, the *client* requested ownership of that cube from the *controller*. When ownership was granted, the cube was moved in front of the device with one meter distance and started to follow the device until the user selected it again. Upon the second selection, the *client* gave the ownership of the cube back to the *controller*. Figure 6.8 shows the visualization of the cube movement on both devices.



Figure 6.8: Cubes follow the device while moving.

When the users had finished positioning, the distance of all cubes to their original locations were calculated. Each cube can be located at two different locations, but the selection of the matching position was contingent on the location of the other cube. For instance, as illustrated in Figure 6.9, only A-B or C-D pairs can be selected. During the study, this selection was made based on the closest cube to the original position.

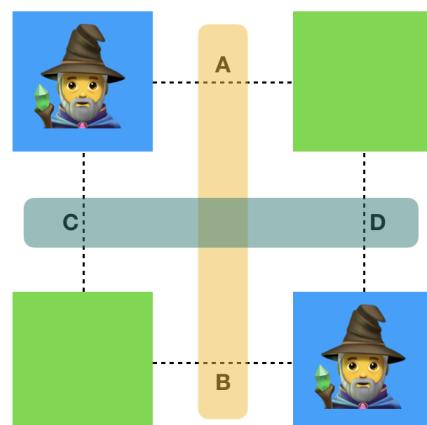


Figure 6.9: Possible selections while calculating distances of repositioned cubes to original positions. Green rectangles represent original positions.

To compensate accuracy issues that have been mentioned in section 6.2.2, distance calculation was made in all active devices, and the average of these values were used during the data analysis.

6.4.4 Logging

To use later in the evaluation, all important events during the tasks were logged, including:

- The tasks' start and stop time, as well as the duration of each task
- Selected cubes and the actions taken during the *object identification* task
- The number of times a cube was selected at least twice without a match during the *object identification* task
- Selected cubes and their new positions during the *object positioning* task
- At the end of the positioning task, distances of the all cubes to their original positions for each active device

The logging was done centrally by the *controller*, which also saved the log to a text file for its further use.

In this chapter, the developed collaborative AR application was discussed. The application was developed based on the user study and consisted of iOS and HoloLens implementations, as well as a controller app. It was developed with the Unity Engine. Vuforia SDK was used to realize tracking on iPad and Photon Networking was used as the networking library. The biggest challenge of the implementation was displaying different virtual objects at the same position, which was solved by using image markers. The following chapter discusses the evaluation process and the evaluation results.

Chapter 7

Evaluation

To answer the research questions of this thesis, a user study was conducted. In this chapter, the user study and its results are presented. Section 7.1 explains the evaluation process. Section 7.2 reports the evaluation results. Finally in section 7.3 evaluation results and their implications on the research questions are discussed.

7.1 User Study

As previously explained in chapter 4, this thesis strives to answer the following research questions:

- **Q1:** How is the group coordination and performance affected by different device types during the real-time collaboration for different tasks?
- **Q2:** Do users prefer virtual or real reference points for coordination while collaborating with different device types?

To answer these questions, a user study was designed and conducted. The concept of the user study, its implementation and the decisions related to it were previously explained in section 5.2.3 and chapter 6. This section describes the evaluation procedure.

7.1.1 Evaluation procedure

The study used a within-group design. During the study, users were asked to complete two collaborative tasks with using HoloLens and iPad combinations:

- **Object identification:** 3D Cubes with white textures were distributed in the environment around physical and virtual objects. These cubes revealed their symbol if selected. One user could select one cube at a time. If users selected the cubes with matching symbols, cubes were removed, if not, their symbols were hidden again.
- **Object positioning:** Users were asked to place the cubes back into their original positions.

Order of devices to users was switched for every group to prevent order effects. Order of devices for each four dyads is listed in Table 7.1.

	Session 1	Session 2	Session 3	Session 4
Group 1	 HoloLenses	 HoloLens iPad	 iPad HoloLens	 iPads
Group 2	 HoloLens iPad	 iPad HoloLens	 iPads	 HoloLenses
Group 3	 iPad HoloLens	 iPads	 HoloLenses	 HoloLens iPad
Group 4	 iPads	 HoloLenses	 HoloLens iPad	 iPad HoloLens

Table 7.1: The order of used device combinations during the study

To decrease the possible bias towards any device and familiarize participants with the environment, users completed the tasks twice as a tutorial. During the tutorial, both HoloLens and iPad were introduced, order of which was also switched during each session. Regular sessions did not begin till participants confirmed that they were familiar with the environment and the application.

It is worth noting that although mental workload evaluation could have provided some important insights on effects of the device combination, it was not included in this study due to time constraints.

After all the tasks were completed in all combinations, users filled a questionnaire about their opinions. This survey aimed to identify user preferences, as well as their perceptions of their own performance.

7.1.2 Participants

20 participants (5 female, 15 male) between 23-63 years of age ($M = 33.65$, $SD = 10.91$) were recruited for the user study. 11 participants had previous experience with the HoloLens, and all participants reported previous tablet usage.

7.1.3 Apparatus and Study Environment

User study was conducted at Fraunhofer Institute for Applied Information Technology (FIT) between 6th and 25th of April 2018. Fraunhofer FIT allocated a room with $7 \times 4 \times 2.65$ meters physical size. Study equipment, two HoloLenses and two iPads (10.5-inch, 2017), were also provided by the Fraunhofer FIT.

Tables and chairs, a flat screen TV, a floor lamp, Sony ERS-7 robotic dog, two posters, a mannequin head and five image markers were presented in the room as physical objects. Besides these, participants also referred to other physical features of the room like windows, the floor and the door. The view of the room without virtual objects can be seen in Figure 7.1.

Besides the cubes, seven virtual objects were displayed: a coffee table with a button, a dog, a sofa, a flower, a monitor, a table with a radio, a refrigerator. These objects can be seen in Figure 6.5.

7.1.4 Data Collection

As listed in section 6.4.4, task completion times, as well as amount of wrongly selected cubes for the *object identification* task and distances of cubes to original positions for the *object positioning* task were collected through the application.

To analyze the communication behaviour, each study was video recorded without the addition of virtual objects. Afterwards, spatial expressions and gestures of users were counted for seven different categories:



Figure 7.1: The view of the test room without virtual objects. Note that Sony ERS-7 robotic dog is missing in this picture and normally located at far side of the table at center-right.

- **Deictic Speech:** Phrases that "can't be fully understood with speech alone, like "here", "over there" and "this one" [69].
- **Participant:** Referring to the other participant or self, e.g. "behind you", "at my foot".
- **Region:** Referring to part of the test room, e.g. "in the middle of the room", "at that corner".
- **Physical objects:** Referring to physical objects, e.g. "in front of the desk", "on the floor".
- **Virtual objects:** Referring to virtual objects, e.g. "next to the fridge", "on the sofa".
- **Hand gestures:** Pointing a location or directing an area using hands.
- **HoloLens pointing:** Occurs with HoloLens when it is not clear if the user intended to point or was keeping their hand ready for the selection gesture.

In addition to this, participants also provided feedback through a survey.

The collected data through the application and video are presented in appendix B without any additional comments. The questionnaire results can be seen at tables 7.7 and 7.8.

7.2 Evaluation Results

Overall results show different device combinations have only minor effects on the performance and communication. The evaluation methods and results are reported in three sections: performance, communication and questionnaire. Results are discussed in detail later in section 7.3.

It should be noted that during the study users completed tasks twice with HoloLens - iPad combination. In this chapter rounded up averages of results from these two iterations are used. However, results from both iterations of HoloLens - iPad combination were also separately analyzed to ensure that they do not produce different results from the ones presented here.

In addition, device combinations are referred to as following through the rest of this chapter:

- HoloLens - HoloLens combination: HoloLenses
- HoloLens - iPad combination: Mixture
- iPad - iPad combination: iPads

7.2.1 Performance

User performance was defined as task completion times and error rates.

Task Completion Times

On average, users completed both tasks faster with *iPads*. Moreover, the number of pairs who completed the tasks faster using *iPads* is higher compared to the other combinations (*Object identification* = 4, *object positioning* = 6).

However, further analysis with Wilcoxon signed rank test showed time difference is significant only between *iPads* and *mixture* while comparing overall completion time of both tasks. Details of the analysis are displayed in Table 7.2.

Object Identification							
Task Duration	HoloLenses	Mixture	iPads		HoloLenses	Mixture	iPads
	Mean 213 s	209,2 s	196,7 s		Mean 4	4,8	5
	SD 97,72	88,17	104,13		SD 3,77	5,15	4,76
				z p			
	Mixture vs HoloLenses 0,45	0,69			Mixture vs HoloLenses 0,92	0,37	
	Mixture vs iPads 1,17	0,27			Mixture vs iPads 0,17	0,91	
	HoloLenses vs iPads 1,27	0,23			HoloLenses vs iPads 0,89	0,42	
(Completion time of object identification task in seconds)				(Number of times a cube selected at least twice without a match)			
Object Positioning							
Task Duration	HoloLenses	Mixture	iPads		HoloLenses	Mixture	iPads
	Mean 194,2 s	199,3 s	175,4 s		Mean 0,76 m	0,79 m	0,89 m
	SD 71,55	88,80	104,69		SD 0,284	0,289	0,31
				z p			
	Mixture vs HoloLenses 0,25	0,50			Mixture vs HoloLenses 0,15	0,92	
	Mixture vs iPads 1,07	0,32			Mixture vs iPads 1,57	0,13	
	HoloLenses vs iPads 0,77	0,49			HoloLenses vs iPads 1,88	0,06	
(Completion time of object positioning task in seconds)				(Average distance of memory cubes to original positions)			
Overall							
Total Duration	HoloLenses	Mixture	iPads		HoloLens	iPad	
	Mean 407,2 s	408,5 s	372,1 s		Mean 2,7	3,5	
	SD 166,96	171,35	179,64		SD 3,12	3,98	
				z p			
	Mixture vs HoloLenses 0,23	0,85			HoloLens vs iPad 1,79	0,07	
	Mixture vs iPads 2,51	0,009					
	HoloLenses vs iPads 1,56	0,12					
(Completion time of both tasks together in seconds)				(Number of times a cube selected at least twice without a match for all HoloLens and iPad users)			
<p>p-value $\leq 0,05$ is significant p-value $\leq 0,01$ is very significant p-value $\leq 0,001$ is extremely significant</p>							

Table 7.2: The analysis of the performance

It is worth noting that several users had issues with the selection gesture of HoloLens and this might have affected their task completion times.

Error Rates

On average, users selected fewer wrong cubes during *object identification* and had more accuracy for *object positioning* while using *HoloLenses*. On the other hand, Wilcoxon signed rank test showed no significant difference for any of these cases. Details of the analysis are displayed in Table 7.2.

7.2.2 Communication

Communication behaviour was examined under deictic speech, object references, overall vocal expressions, gestures and other behaviour.

Deictic Speech

The Wilcoxon signed rank test revealed deictic speech is used significantly less with *iPads* compared to *mixture* for *object identification* task. Also overall, *HoloLens* users used significantly more deictic speech than *iPad* users. Details of the analysis are displayed in Table 7.3.

Object References

Spatial expression and gesture categories were previously listed in section 7.1.4. While evaluating referenced physical objects, both region and participant categories were included into physical objects. Wilcoxon signed rank test revealed that collaborators only referenced physical objects significantly more with *mixture* compared to *iPads* during the *object positioning*. Wilcoxon test also revealed there is no significant difference for referencing virtual objects for any of the device combinations. Details of the analysis are displayed in Table 7.4.

Further analysis with Wilcoxon signed rank test showed participants referenced virtual object significantly more than physical objects for all possible combinations, except with *HoloLenses* on *object positioning* task. Details of the analysis can be seen in Figure 7.2.

Overall Vocal Expressions

Total amount of spatial vocal expressions (deictic speech, referencing other participants, regions, physical and virtual objects) used by the participants have

Object Identification			Object Positioning				
	HoloLenses	Mixture	iPads		HoloLenses	Mixture	iPads
Mean	11,2	11,65	7,2	Mean	12,6	10,85	10,1
SD	5,67	6,58	5,53	SD	5,5	3,8	5,54
	z	p			z	p	
Mixture vs HoloLenses	0,15	0,92			Mixture vs HoloLenses	8,44	0,42
Mixture vs iPads	2,19	0,027			Mixture vs iPads	0,47	0,65
HoloLenses vs iPads	1,78	0,08			HoloLenses vs iPads	1,32	0,19

(Number of times participants used deictic speech during object identification task)

Overall			HoloLens vs iPad				
	HoloLenses	Mixture	iPads	HoloLens	iPad		
Mean	23,8	22,5	17,3	Mean	48,6		
SD	10,32	9,18	9,83	SD	20,47		
	z	p			z	p	
Mixture vs HoloLenses	0,45	0,69			HoloLens vs iPad	2,14	0,027
Mixture vs iPads	1,88	0,06					
HoloLenses vs iPads	1,78	0,08					

(Total number of times participants used deictic speech with HoloLenses and iPads)

p-value $\leq 0,05$ is significant
 p-value $\leq 0,01$ is very significant
 p-value $\leq 0,001$ is extremely significant

Table 7.3: The analysis of the deictic speech

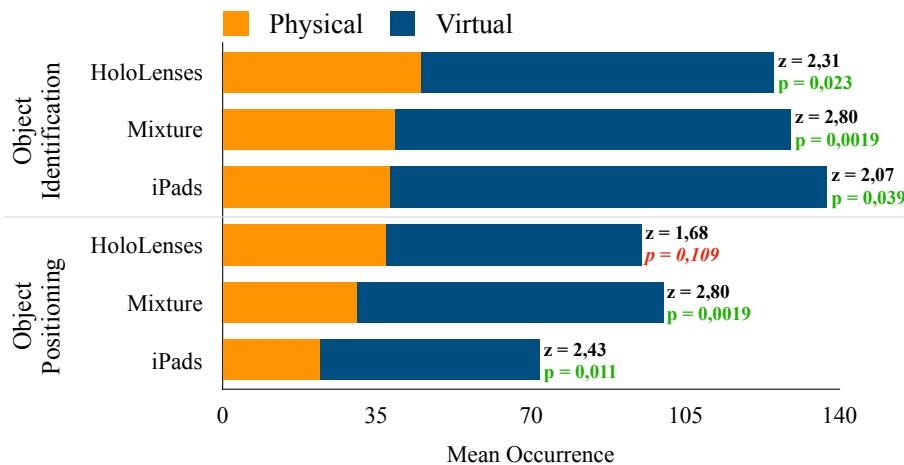


Figure 7.2: The analysis of the object references

Object Identification																														
Physical Reference	HoloLenses			Mixture			iPads		Virtual Reference	HoloLenses			Mixture			iPads														
	Mean		5,4	Mean		5,45	Mean		4,8	Mean		5,84	Mean		6,53	Mean		9,9												
	SD		5,25	SD		4,38	SD		4,61	SD		5,84	SD		6,53	SD		9,79												
			<hr/>		<hr/>		<hr/>			<hr/>		<hr/>		<hr/>		<hr/>														
	z		p																											
	Mixture vs HoloLenses		0,21		<hr/>		0,84			Mixture vs HoloLenses		0,89		<hr/>		0,42														
	Mixture vs iPads		0,77		<hr/>		0,49			Mixture vs iPads		0,65		<hr/>		0,57														
	HoloLenses vs iPads		0,59		<hr/>		0,65			HoloLenses vs iPads		0,94		<hr/>		0,43														
	(Number of times participants referenced physical objects during object identification task)									(Number of times participants referenced virtual objects during object identification task)																				
Object Positioning																														
Physical Reference	HoloLenses			Mixture			iPads		Virtual Reference	HoloLenses			Mixture			iPads														
	Mean		4,5	Mean		4,35	Mean		2,2	Mean		5,8	Mean		6,95	Mean		5												
	SD		4,22	SD		3,06	SD		1,98	SD		3,91	SD		5,07	SD		4,44												
			<hr/>		<hr/>		<hr/>			<hr/>		<hr/>		<hr/>		<hr/>														
	z		p																											
	Mixture vs HoloLenses		0,17		<hr/>		0,91			Mixture vs HoloLenses		1,07		<hr/>		0,32														
	Mixture vs iPads		2,49		<hr/>		0,011			Mixture vs iPads		1,89		<hr/>		0,06														
	HoloLenses vs iPads		1,54		<hr/>		0,14			HoloLenses vs iPads		1,10		<hr/>		0,37														
	(Number of times participants referenced physical objects during object positioning task)									(Number of times participants referenced virtual objects during object positioning task)																				
Overall																														
Physical Reference	HoloLenses			Mixture			iPads		Virtual Reference	HoloLenses			Mixture			iPads														
	Mean		9,9	Mean		9,8	Mean		7	Mean		13,8	Mean		15,95	Mean		14,9												
	SD		9,18	SD		7,29	SD		6,01	SD		9,21	SD		11,14	SD		13,16												
			<hr/>		<hr/>		<hr/>			<hr/>		<hr/>		<hr/>		<hr/>														
	z		p																											
	Mixture vs HoloLenses		0,07		<hr/>		1,05			Mixture vs HoloLenses		1,32		<hr/>		0,19														
	Mixture vs iPads		1,73		<hr/>		0,08			Mixture vs iPads		1,30		<hr/>		0,25														
	HoloLenses vs iPads		1,12		<hr/>		0,27			HoloLenses vs iPads		0,56		<hr/>		0,64														
	(Total number of times participants referenced physical objects during both tasks)									(Total number of times participants referenced virtual objects during both tasks)																				
p-value $\leq 0,05$ is significant p-value $\leq 0,01$ is very significant p-value $\leq 0,001$ is extremely significant																														

Table 7.4: The analysis of the object references

no significant difference for *object identification* task, but for the *object positioning*, *HoloLenses* used significantly more spatial expressions than *iPads*. For overall, during both tasks, users used significantly more spatial expressions when the combination included a *HoloLens*. *HoloLens* users also used more spatial expressions than *iPad* users. Details of the analysis are displayed in Table 7.5.

Object Identification			Object Positioning				
	HoloLenses	Mixture	iPads		HoloLenses	Mixture	iPads
Mean	24,6	26,1	21,9	Mean	22,9	22,15	17,3
SD	12,92	11,73	13,40	SD	10,77	9,13	7,28
	<u>z</u>	<u>p</u>			<u>z</u>	<u>p</u>	
<i>Mixture vs HoloLenses</i>	0,35	0,76		<i>Mixture vs HoloLenses</i>	0,05	1	
<i>Mixture vs iPads</i>	1,36	0,20		<i>Mixture vs iPads</i>	1,68	0,10	
<i>HoloLenses vs iPads</i>	1,30	0,25		<i>HoloLenses vs iPads</i>	2,09	0,037	

(Number of times participants used a spatial expression during object identification task)

Overall			HoloLens vs iPad			
	HoloLenses	Mixture	iPads		HoloLens	iPad
Mean	47,5	48,25	39,2	Mean	100,3	82,9
SD	22,43	19,61	18,85	SD	42,73	35,26
	<u>z</u>	<u>p</u>			<u>z</u>	<u>p</u>
<i>Mixture vs HoloLenses</i>	0,15	0,92		<i>HoloLens vs iPad</i>	2,7	0,0039
<i>Mixture vs iPads</i>	2,19	0,027				
<i>HoloLenses vs iPads</i>	2,24	0,027				

(Total number of times participants used a spatial expression during both tasks)

p-value $\leq 0,05$ is significant
 p-value $\leq 0,01$ is very significant
 p-value $\leq 0,001$ is extremely significant

Table 7.5: The analysis of the spatial vocal expressions

Gestures

Wilcoxon signed rank test clearly confirmed that participants used hand gestures more often while using *HoloLens*. Only case that had no significant difference is the comparison of *mixture* with *HoloLenses* in which both combinations already have at least one *HoloLens*. When usage of hand gestures for different tasks were compared, Wilcoxon test revealed participants used

significantly more hand gestures with *HoloLenses* on *object identification* task (*Object identification*: $M = 11.7$, $SD = 6.89$, *Object positioning*: $M = 8.4$, $SD = 6.22$, $z = 2.43$, $p = 0.011$). Note that including "HoloLens pointing" which is explained in section 7.1.4 to the HoloLens data does not change the results. Details of the analysis without "HoloLens pointing" can be seen in Table 7.6.

Object Identification			Object Positioning				
	HoloLenses	Mixture	iPads		HoloLenses	Mixture	iPads
Mean	11,7	9,95	3,7	Mean	8,4	8	4,9
SD	6,89	4,97	1,88	SD	6,22	3,38	3,31
	z	p			z	P	
<i>Mixture vs HoloLenses</i>	0,84	0,46		<i>Mixture vs HoloLenses</i>	0,35	0,82	
<i>Mixture vs iPads</i>	2,65	0,0058		<i>Mixture vs iPads</i>	2,38	0,01	
<i>HoloLenses vs iPads</i>	2,54	0,0058		<i>HoloLenses vs iPads</i>	1,93	0,04	

(Number of times participants used a hand gesture during object identification task)

Overall			HoloLens vs iPad			
	HoloLenses	Mixture	iPads		HoloLens	iPad
Mean	20,1	17,95	8,6	Mean	42,9	21,7
SD	12,72	7,24	4,78	SD	18,77	8,55
	z	p			z	p
<i>Mixture vs HoloLenses</i>	0,88	0,42		<i>HoloLens vs iPad</i>	2,70	0,003
<i>Mixture vs iPads</i>	2,70	0,003				
<i>HoloLenses vs iPads</i>	2,49	0,009				

(Total number of times participants used a hand gesture during both tasks)

p-value ≤ 0.05 is significant
 p-value ≤ 0.01 is very significant
 p-value ≤ 0.001 is extremely significant

Table 7.6: The analysis of the hand gestures

Other

Besides the listed situations, the most noteworthy behaviour of participants was directing an iPad as a pointing gesture, usually accompanied with a deictic expression (e.g. directing iPad to a specific position and saying "the cube is over there"). This instance occurred in 7 out of 10 groups at least twice. Another striking behaviour of participants was checking other participant's iPad. This happened in 7 out of 10 groups, total of 19 occurrences.

In several occasions, people used their hands or feet as anchor points and asked other user to place the memory cube exactly at that point. Multiple users also remarked that depth perception in iPad was non-existent and field of view of the HoloLens display was too narrow. One user checked the height of the cubes by moving the cube inside the other virtual objects while using iPad. A common request from the collaborators was to display where the other HoloLens user is looking at, or at least have a feature like a laser pointer to point objects in the virtual space. Similarly, some users complained about the lack of rulers and guidelines.

All groups tried to develop strategies for both tasks, but they often failed to follow them through these. For instance, several groups tried to memorize how many cubes were around specific objects in the beginning of *object identification* task, but during the *object positioning*, they either did not mention these or misremembered the locations and amounts. Two different groups used associations for this purpose and created stories or memorable quotes based on symbols and the environment, such as "dog is eating the octopus but octopus is trying to escape" or "dinosaur is extinct, so it flies". Both groups had the best accuracies for the *object positioning* tasks ($M_{Group\ 6} = 0.35m$, $SD_{Group\ 6} = 0.27$, $M_{Group\ 5} = 0.49m$, $SD_{Group\ 5} = 0.28$).

In addition to these, based on the observation, majority of instances where participants stumbled or hit objects happened while they were using iPads, although these instances were not systematically counted.

7.2.3 Questionnaire

First section of the user survey asked users about their opinions on individual devices. More participants reported that they perceived both digital and physical objects well with the iPad, but estimated distances easily with the HoloLens. Full results of this section are reported in Table 7.7.

In the second section of the questionnaire, users were asked about their preferences and self-estimated performance. 12 out of 20 participants preferred the combination of same devices more. More users reported that they believe they performed better with *iPads*. All user answers are shown in Table 7.8.

iPad						Questions		HoloLens					
1	2	3	4	5	Average			Average	1	2	3	4	5
16	4	0	0	0	1,2	Selection of cubes was easy		2,6	4	5	7	3	1
7	9	2	2	0	1,95	Moving cubes was easy		2,1	7	4	9	0	0
14	3	1	2	0	1,55	I was able to perceive physical objects very well		1,7	10	6	4	0	0
16	3	1	0	0	1,25	I was able to perceive digital objects very well		2,25	5	7	7	0	1
3	5	5	3	4	3	I was able to estimate distances easily		2,2	6	8	2	4	0
9	8	3	0	0	1,7	I was able to collaborate well with my team partner while using the device		1,55	11	7	2	0	0

Table 7.7: Questionnaire results for the individual devices section. 1 represents "I agree", 5 represents "I do not agree".

	HoloLens	iPad	Both same	None of them
Which device did you like most for these tasks?	6	10	2	2
Which device was physically more comfortable?	2	13	4	1
	HoloLenses	Mixture	iPads	All the same
Which combination did you like most?	7	2	5	6
In your opinion, in which combination you found all matching cubes faster?	1	2	11	6
In your opinion, in which combination you placed all cubes back more accurately?	4	1	9	3
In your opinion, which combination created the best collaboration?	6	2	8	4
				None of them

Table 7.8: Questionnaire results for user preferences

7.3 Discussion

The results reported in section 7.2 are interpreted to answer the research questions. The first research question can be examined under performance and group coordination categories and it is directly related to all of the collected data. Second question is about object references and can be directly answered using the data presented in section 7.2.2.

7.3.1 Performance

In this study, performance is defined by task completion times and error rates, which are presented in Table 7.2.

Results clearly show that even though users completed both tasks faster with *iPads* on average, difference is not statistically significant, except in one case. When the total duration of tasks for different device types was compared, *iPads* were significantly faster than *mixture*. However, this relation does not occur between *mixture* and *HoloLenses* or *iPads* and *HoloLenses*. Therefore this result implies, usage of iPad - iPad combination is time-wise more favorable compared to iPad - HoloLens combination, but other than that, device type has no significant effect on task completion time.

Analysis of error rates also show device type has no major effect on the accuracy, though on average, *HoloLenses* selected fewer wrong cubes during the first task and had more positional accuracy during the second one. It is worth noting that *p-value* for *HoloLenses* vs *iPads* is 0.06 for distances of memory cubes to original positions in *object positioning* task, which is very close to significant value. Interestingly, in the survey more collaborators claimed they have placed cubes in the original positions more accurately with *iPads* (table 7.8), even though more users reported they were able to estimate distances well with *HoloLens* on average. This shows lack of depth perception on tablets decreases not only accuracy, but also spatial-awareness.

7.3.2 Group Coordination

First and foremost, as displayed in Table 7.5 and 7.6, *HoloLens* induces usage of significantly more spatial expressions and hand gestures. This might be due to display modality of *HoloLens*, since it can convey 3D space better compared to the tablet screen.

Mixture users, because of the *HoloLens*, used significantly more hand gestures compared to *iPads*. On the other hand, total amount of spatial expressions had no significant difference for separate tasks. However average of *mixture* was always higher than *iPads*. Specific spatial expression categories like deictic speech and object references are also compatible with these results.

Additionally, participants reported they were able to collaborate well with their partner regardless of the device type. However, based on the answers of the questionnaire, as listed in Table 7.8, *mixture* is the least preferred combination amongst all.

7.3.3 Object References

As Figure 7.2 clearly laid out, participants referenced virtual objects significantly more in all device combinations during both tasks, except *HoloLenses* during *object positioning*. Nevertheless, average amount of virtual object references is higher than physical objects for this case too.

7.3.4 Research Questions

In this section, answers of the research questions based on the analyzed data are presented.

Q1: How is the group coordination and performance affected by different device types during the real-time collaboration for different tasks?

For the group performance, results of this study shows the following:

- User performance is not affected by device type for separate tasks.
- Using iPad - HoloLens combination yields faster task completion time compared to iPad - iPad combination, when tasks are evaluated together.
- Users estimate distances worse with iPads, compared to HoloLenses.

For the group coordination, this study suggests following results:

- Users use more hand gestures and spatial expressions when HoloLens is involved.
- During object identification, users use more deictic speech with iPad - HoloLens combination than iPad - iPad combination.
- While positioning objects, users refer to physical objects more with iPad - HoloLens combination than iPad - iPad combination.
- Participants prefer using same devices together more than iPad - HoloLens combination.

These results might be interpreted in several ways.

First of all, device type does not affect the time required to perform tasks. Even though the average values were shorter for iPad - iPad combination during the study, device familiarity might have also played a role here; many users without previous HoloLens experience had problems with the selection gesture

and lost time while repeating the gesture. Therefore difference between average task completion times might even get shorter if the users are more familiar with the device.

Moreover, error rates are also not majorly affected by the device type. But average error rates on *object positioning* task, as well as questionnaire results and user comments during the study show having a HoloLens in the combination would be beneficial if physical accuracy is desired. This also alternatively suggests that tablets need a proper user interface to convey the depth of the environment.

As opposed to performance, group coordination is clearly affected by the device type. Even though comparing combinations shows no serious differences, collaborators used significantly more hand gestures and spatial expressions when HoloLens was involved. Obviously, carrying a tablet limits the usage of hand gestures and is therefore definitely expected. Furthermore, increase in usage of spatial expressions might be a result of the continuous immersive illusion created by HoloLens. On iPad, users see virtual objects through a 2D screen. Whenever user looks away from the screen, immersion is broken, which makes it harder to perceive objects as part of the real world. On the other hand, despite the narrow field of view, virtual objects are visible through first person view on HoloLens, and easier to register as part of the real environment, which might have caused the increase.

And finally, users prefer using combinations of same devices more than the mixture of them. For the specific devices, only slightly more users liked iPad more, but majority of them reported iPad was physically more comfortable.

Q2: Do users prefer virtual or real reference points for coordination while collaborating with different device types?

Results of this study show users refer to virtual objects significantly more in all device combinations during all tasks. Only exception of this is using HoloLens - HoloLens combination while positioning objects, though during the study average amount of virtual object references was higher for this condition too.

This result indicates virtual objects have critical importance for collaboration which is also consistent with the study of Müller et al. [67].

This chapter discussed evaluation process and the results of the evaluation. The user study was conducted with 20 participants and their performance and communication were analyzed. Results show that the device type does not

have a significant effect on performance, participants use more hand gestures and spatial expressions with the HoloLens and refer to virtual objects more than physical objects. They also prefer using combinations of the same devices more than a mixture of them. The following chapter summarizes and concludes this thesis.

Chapter 8

Conclusion and Future Work

The popularity of Augmented Reality (AR) is increasing considerably across the consumer market due to recent developments on processing, tracking and display technologies. These technologies have many variations, which has caused a notable fragmentation on consumer product types.

Collaborating with AR devices is one of the trending usage areas of the AR. Over the years, very comprehensive work has been done on AR collaboration topic that ranges from realizing such a system to comparing it with existing methods. However, possible effects of using different device types during the collaboration, such as tablets and smart glasses, are only barely explored.

During this thesis, a user study was designed to investigate the effects of device types during real-time, face to face AR collaboration on performance, group coordination and referenced object types. A memory card game was selected as the main theme of the study and a multiplayer AR version of it was implemented for HoloLens and iPad.

The application was developed with Unity. On iPad, tracking was achieved with Vuforia software development kit (SDK). The main challenge of the implementation was displaying objects at the exact physical positions since both devices generate individual coordinate systems. This was solved by using image markers. Image markers were detected with Vuforia SDK on both devices. Photon Networking was used as the networking library.

The user study was conducted with 20 volunteers. During the study, task completion times and error rates were logged. Sessions were video recorded and afterwards the individual participant's usage of spatial expressions as well as hand gestures were counted. Users also filled a questionnaire about their

opinions.

Results indicate that the device type does not have a significant effect on performance. However, if physical accuracy is required, using a HoloLens in combination is beneficial. Collaborators use more hand gestures and spatial expressions with the HoloLens, which suggests virtual objects are perceived as a part of the physical environment more with head-mounted displays compared to handheld ones. Furthermore, data clearly shows users refer to virtual objects more than physical objects. And finally, during the study, users preferred using combinations of the same devices more than a mixture of them.

8.1 Future Work

While this study laid out the effects of different device types on AR collaboration, it also raises several questions for potential future work. First of all, effects of different device types on remote AR collaboration is still an open question. Especially the high usage of hand gestures and spatial expressions by HoloLens users implies that there might be significant performance effects during the remote collaboration. Secondly, mental workload of using different devices was not measured during this thesis and is yet to be explored. Moreover, it could be beneficial to repeat the presented user study with more participants and between-groups design. And finally, the number of device types used during this study was limited. How other input and display modalities, such as speech input and projector displays, affect collaboration are still required to be investigated.

Chapter 9

Bibliography

- [1] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoper. Virtual Environ.*, vol. 6, no. 4, pp. 355–385, Aug. 1997. [Online]. Available: <http://dx.doi.org/10.1162/pres.1997.6.4.355>
- [2] M. Billinghurst, A. Clark, and G. Lee, "A survey of augmented reality," *Foundations and Trends® in Human–Computer Interaction*, vol. 8, no. 2-3, pp. 73–272, 2015. [Online]. Available: <http://dx.doi.org/10.1561/1100000049>
- [3] D. Schmalstieg and T. Höllerer, *Augmented Reality - Principles and Practice*, 06 2016.
- [4] R. W. Lindeman, H. Noma, and P. G. de Barros, "Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nov 2007, pp. 173–176.
- [5] O. Bau and I. Poupyrev, "Revel: Tactile feedback technology for augmented reality," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 89:1–89:11, Jul. 2012. [Online]. Available: <http://doi.acm.org/10.1145/2185520.2185585>
- [6] S. A. Seah, D. Martinez Plasencia, P. D. Bennett, A. Karnik, V. S. Otrocol, J. Knibbe, A. Cockburn, and S. Subramanian, "Sensabubble: A chrono-sensory mid-air display of sight and smell," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 2863–2872. [Online]. Available: <http://doi.acm.org/10.1145/2556288.2557087>
- [7] H. Iwata, H. Yano, T. Uemura, and T. Moriya, "Food simulator: a haptic interface for biting," in *IEEE Virtual Reality 2004*, March 2004, pp. 51–57.
- [8] D. Biçer, "Supporting collaboration in construction settings with a mixed reality issue tracker system in a building information modeling process," Master's thesis, RWTH Aachen, Germany, 2017.
- [9] A. Butz, T. Hollerer, S. Feiner, B. MacIntyre, and C. Beshers, "Enveloping users and computers in a collaborative 3d augmented reality," in *Augmented Reality, 1999. (IWAR '99) Proceedings. 2nd IEEE and ACM International Workshop on*, 1999, pp. 35–44.
- [10] H. Regenbrecht, M. Wagner, and G. Baratoff, "Magicmeeting: A collaborative tangible augmented reality system," *Virtual Reality*, vol. 6, no. 3, pp. 151–166, Oct 2002. [Online]. Available: <https://doi.org/10.1007/s100550200016>

- [11] T. Lee and T. Hollerer, "Handy ar: Markerless inspection of augmented reality objects using fingertip tracking," in *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers*, ser. ISWC '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 1–8. [Online]. Available: <https://doi.org/10.1109/ISWC.2007.4373785>
- [12] M. Lee, M. Billinghurst, W. Baek, R. Green, and W. Woo, "A usability study of multimodal input in an augmented reality environment," *Virtual Real.*, vol. 17, no. 4, pp. 293–305, Nov. 2013. [Online]. Available: <https://doi.org/10.1007/s10055-013-0230-0>
- [13] R. W. Lindeman, G. Lee, L. Beattie, H. Gamper, R. Pathinarupothi, and A. Akhilesh, "Geoboids: A mobile ar application for exergaming," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, Nov 2012, pp. 93–94.
- [14] K. Kansaku, N. Hata, and K. Takano, "My thoughts through a robots eyes: An augmented reality-brain-machine interface," *Neuroscience Research*, vol. 66, no. 2, p. 219–222, 2010.
- [15] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2007.1049>
- [16] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality*, ser. ISMAR '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 127–136. [Online]. Available: <http://dx.doi.org/10.1109/ISMAR.2011.6092378>
- [17] R. S. Sodhi, B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci, "Bethere: 3d mobile collaboration with spatial input," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 179–188. [Online]. Available: <http://doi.acm.org/10.1145/2470654.2470679>
- [18] S. Feiner, B. MacIntyre, T. Hollerer, and A. Webster, "A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment," in *Proceedings of the 1st IEEE International Symposium on Wearable Computers*, ser. ISWC '97. Washington, DC, USA: IEEE Computer Society, 1997, pp. 74–81. [Online]. Available: <http://dl.acm.org/citation.cfm?id=851036.856454>
- [19] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway, "Exploring mars: developing indoor and outdoor user interfaces to a mobile augmented reality system," *Computers & Graphics*, vol. 23, no. 6, pp. 779 – 785, 1999. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S009784939900103X>
- [20] B. Thomas, V. Demczuk, W. Piekarski, D. Hepworth, and B. Gunther, "A wearable computer system with augmented reality to support terrestrial navigation," in *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, Oct 1998, pp. 168–171.
- [21] G. A. Lee, A. Dünser, S. Kim, and M. Billinghurst, "Cityviewar: A mobile outdoor ar application for city visualization," in *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, Nov 2012, pp. 57–64.
- [22] E. Foxlin, Y. Altshuler, L. Naimark, and M. Harrington, "Flighttracker: A novel optical/inertial tracker for cockpit enhanced vision," in *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ser. ISMAR '04.

- Washington, DC, USA: IEEE Computer Society, 2004, pp. 212–221. [Online]. Available: <http://dx.doi.org/10.1109/ISMAR.2004.32>
- [23] M. Bajura, H. Fuchs, and R. Ohbuchi, “Merging virtual objects with the real world: Seeing ultrasound imagery within the patient,” in *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH ’92. New York, NY, USA: ACM, 1992, pp. 203–210. [Online]. Available: <http://doi.acm.org/10.1145/133994.134061>
- [24] W. Yii, W. H. Li, and T. Drummond, “Distributed visual processing for augmented reality,” in *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, ser. ISMAR ’12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 41–48. [Online]. Available: <http://dx.doi.org/10.1109/ISMAR.2012.6402536>
- [25] K. H. Ahlers, A. Kramer, D. E. Breen, P.-Y. Chevalier, C. Crampton, E. Rose, M. Tuceryan, R. T. Whitaker, and D. Greer, “Distributed augmented reality for collaborative design applications,” *Computer Graphics Forum*, vol. 14, no. 3, pp. 3–14, 1995. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.1995.cgf143_0003.x
- [26] J. Rekimoto, “Transvision: A hand-held augmented reality system for collaborative design,” 1996.
- [27] M. Gervautz, D. Schmalstieg, Z. Szalavri, K. Karner, F. Madritsch, and A. Pinz, “Studierstube-a multi-user augmented reality environment for visualization and education,” *Technical report TR-186-2-96-10*, 1996.
- [28] Z. Szalavári, 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: <https://doi.org/10.1007/BF01409796>
- [29] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavári, L. M. Encarnaçāo, M. Gervautz, and W. Purgathofer, “The studierstube augmented reality project,” *Presence: Teleoper. Virtual Environ.*, vol. 11, no. 1, pp. 33–54, Feb. 2002. [Online]. Available: <http://dx.doi.org/10.1162/105474602317343640>
- [30] A. Fuhrmann, H. Loffelmann, and D. Schmalstieg, “Collaborative augmented reality: exploring dynamical systems,” in *Visualization ’97., Proceedings*, Oct 1997, pp. 459–462.
- [31] S. Oh, K. Park, S. Kwon, and H.-J. So, “Designing a multi-user interactive simulation using ar glasses,” in *Proceedings of the TEI ’16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI ’16. New York, NY, USA: ACM, 2016, pp. 539–544. [Online]. Available: <http://doi.acm.org/10.1145/2839462.2856521>
- [32] D. Schmalstieg and D. Wagner, “Experiences with handheld augmented reality,” in *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ser. ISMAR ’07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 1–13. [Online]. Available: <http://dx.doi.org/10.1109/ISMAR.2007.4538819>
- [33] D. Wagner, T. Pintaric, F. Ledermann, and D. Schmalstieg, “Towards massively multi-user augmented reality on handheld devices,” in *Proceedings of the Third International Conference on Pervasive Computing*, ser. PERVASIVE’05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 208–219. [Online]. Available: http://dx.doi.org/10.1007/11428572_13
- [34] H. Benko, E. W. Ishak, and S. Feiner, “Collaborative mixed reality visualization of an archaeological excavation,” in *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ser. ISMAR ’04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 132–140. [Online]. Available: <http://dx.doi.org/10.1109/ISMAR.2004.23>

- [35] L. F. Güл and S. M. Halıcı, "Collaborative design with mobile augmented reality," in *34th eCAADe Conference*, 2016.
- [36] X. Wang and P. S. Dunston, "Comparative effectiveness of mixed reality-based virtual environments in collaborative design," *Trans. Sys. Man Cyber Part C*, vol. 41, no. 3, pp. 284–296, May 2011. [Online]. Available: <http://dx.doi.org/10.1109/TSMCC.2010.2093573>
- [37] A. Hammad, H. Wang, and S. P. Mudur, "Distributed augmented reality for visualizing collaborative construction tasks," *Journal of Computing in Civil Engineering*, vol. 23, no. 6, pp. 418–427, 2009.
- [38] U. Riedlinger, "Visualisierung von inneneinrichtungen auf google tango tablets unter verwendung von umgebungsbeschreibungen," Master's thesis, Hochschule Bonn-Rhein-Sieg, Germany, 2017.
- [39] O. Oda, M. Sukan, S. Feiner, and B. Tversky, "Poster: 3d referencing for remote task assistance in augmented reality," in *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, March 2013, pp. 179–180.
- [40] W. Huang, L. Alem, and F. Tecchia, *HandsIn3D: Supporting Remote Guidance with Immersive Virtual Environments*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 70–77. [Online]. Available: https://doi.org/10.1007/978-3-642-40483-2_5
- [41] R. Poelman, O. Akman, S. Lukosch, and P. Jonker, "As if being there: Mediated reality for crime scene investigation," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, ser. CSCW '12. New York, NY, USA: ACM, 2012, pp. 1267–1276. [Online]. Available: <http://doi.acm.org/10.1145/2145204.2145394>
- [42] D. Datcu, M. Cidota, H. Lukosch, and S. Lukosch, "On the usability of augmented reality for information exchange in teams from the security domain," in *2014 IEEE Joint Intelligence and Security Informatics Conference*, Sept 2014, pp. 160–167.
- [43] S. Nilsson, B. Johansson, and A. Jonsson, "Using ar to support cross-organisational collaboration in dynamic tasks," in *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, Oct 2009, pp. 3–12.
- [44] G. Reitmayr and D. Schmalstieg, "Collaborative augmented reality for outdoor navigation and information browsing," 2004.
- [45] 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*, ser. VRST '14. New York, NY, USA: ACM, 2014, pp. 197–205. [Online]. Available: <http://doi.acm.org/10.1145/2671015.2671016>
- [46] S. Kasahara, V. Heun, A. S. Lee, and H. Ishii, "Second surface: Multi-user spatial collaboration system based on augmented reality," in *SIGGRAPH Asia 2012 Emerging Technologies*, ser. SA '12. New York, NY, USA: ACM, 2012, pp. 20:1–20:4. [Online]. Available: <http://doi.acm.org/10.1145/2407707.2407727>
- [47] D. Datcu, S. G. Lukosch, and H. K. Lukosch, "Comparing presence, workload and situational awareness in a collaborative real world and augmented reality scenario," in *IEEE ISMAR Workshop on Collaboration in Merging Realities (CiMeR), Adelaide, Australia, 1 October 2014*, 2013.

- [48] K. Kiyokawa, H. Iwasa, H. Takemura, and N. Yokoya, "Collaborative immersive workspace through a shared augmented environment," in *Proc. SPIE*, vol. 98, 1998, pp. 2–13.
- [49] M. Bauer, G. Kortuem, and Z. Segall, ""where are you pointing at?" a study of remote collaboration in a wearable videoconference system," in *Proceedings of the 3rd IEEE International Symposium on Wearable Computers*, ser. ISWC '99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 151–. [Online]. Available: <http://dl.acm.org/citation.cfm?id=519309.856501>
- [50] M. Billinghurst, S. Weghorst, and T. Furness, Iii, "Shared space: An augmented reality approach for computer supported collaborative work," *Virtual Real.*, vol. 3, no. 1, pp. 25–36, Mar. 1998. [Online]. Available: <https://doi.org/10.1007/BF01409795>
- [51] G. Reitmayr and D. Schmalstieg, "Mobile collaborative augmented reality," in *Proceedings IEEE and ACM International Symposium on Augmented Reality*, 2001, pp. 114–123.
- [52] W. Broll, M. Stoerring, and C. Mottram, "The augmented round table-a new interface to urban planning and architectural design." in *INTERACT*, 2003.
- [53] H. T. Regenbrecht and M. T. Wagner, "Interaction in a collaborative augmented reality environment," in *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '02. New York, NY, USA: ACM, 2002, pp. 504–505. [Online]. Available: <http://doi.acm.org/10.1145/506443.506451>
- [54] D.-N. T. Huynh, K. Raveendran, Y. Xu, K. Spreen, and B. MacIntyre, "Art of defense: A collaborative handheld augmented reality board game," in *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games*, ser. Sandbox '09. New York, NY, USA: ACM, 2009, pp. 135–142. [Online]. Available: <http://doi.acm.org/10.1145/1581073.1581095>
- [55] S. Dong, A. H. Behzadan, F. Chen, and V. R. Kamat, "Collaborative visualization of engineering processes using tabletop augmented reality," *Advances in Engineering Software*, vol. 55, pp. 45 – 55, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0965997812001287>
- [56] A. Henrysson, M. Billinghurst, and M. Ollila, "Face to face collaborative ar on mobile phones," in *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ser. ISMAR '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 80–89. [Online]. Available: <http://dx.doi.org/10.1109/ISMAR.2005.32>
- [57] D. Kirk and D. Stanton Fraser, "Comparing remote gesture technologies for supporting collaborative physical tasks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06. New York, NY, USA: ACM, 2006, pp. 1191–1200. [Online]. Available: <http://doi.acm.org/10.1145/1124772.1124951>
- [58] W. Huang and L. Alem, "Handsair: A wearable system for remote collaboration on physical tasks," in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, ser. CSCW '13. New York, NY, USA: ACM, 2013, pp. 153–156. [Online]. Available: <http://doi.acm.org/10.1145/2441955.2441994>
- [59] D. Datcu, S. G. Lukosch, and H. K. Lukosch, "A collaborative game to study the perception of presence during virtual co-location," in *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW Companion '14. New York, NY, USA: ACM, 2014, pp. 5–8. [Online]. Available: <http://doi.acm.org/10.1145/2556420.2556792>

- [60] L. Blum, R. Wetzel, R. McCall, L. Oppermann, and W. Broll, "The final timewarp: Using form and content to support player experience and presence when designing location-aware mobile augmented reality games," in *Proceedings of the Designing Interactive Systems Conference*, ser. DIS '12. New York, NY, USA: ACM, 2012, pp. 711–720. [Online]. Available: <http://doi.acm.org/10.1145/2317956.2318064>
- [61] A. Stafford, W. Piekarski, and B. Thomas, "Implementation of god-like interaction techniques for supporting collaboration between outdoor ar and indoor tabletop users," in *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ser. ISMAR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 165–172. [Online]. Available: <http://dx.doi.org/10.1109/ISMAR.2006.297809>
- [62] S. Johnson, M. Gibson, and B. Mutlu, "Handheld or handsfree?: Remote collaboration via lightweight head-mounted displays and handheld devices," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15. New York, NY, USA: ACM, 2015, pp. 1825–1836. [Online]. Available: <http://doi.acm.org/10.1145/2675133.2675176>
- [63] S. Kim, G. A. Lee, and N. Sakata, "Comparing pointing and drawing for remote collaboration," in *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Oct 2013, pp. 1–6.
- [64] D. Datcu, S. Lukosch, and F. Brazier, "On the usability and effectiveness of different interaction types in augmented reality," *International Journal of Human–Computer Interaction*, vol. 31, no. 3, pp. 193–209, 2015. [Online]. Available: <http://dx.doi.org/10.1080/10447318.2014.994193>
- [65] J. W. Chastine, K. Nagel, Y. Zhu, and L. Yearovich, "Understanding the design space of referencing in collaborative augmented reality environments," in *Proceedings of Graphics Interface 2007*, ser. GI '07. New York, NY, USA: ACM, 2007, pp. 207–214. [Online]. Available: <http://doi.acm.org/10.1145/1268517.1268552>
- [66] J. Chastine and Y. Zhu, "The cost of supporting references in collaborative augmented reality," in *Proceedings of Graphics Interface 2008*, ser. GI '08. Toronto, Ont., Canada, Canada: Canadian Information Processing Society, 2008, pp. 275–282. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1375714.1375760>
- [67] J. Müller, R. Rädle, and H. Reiterer, "Virtual objects as spatial cues in collaborative mixed reality environments: How they shape communication behavior and user task load," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 1245–1249. [Online]. Available: <http://doi.acm.org/10.1145/2858036.2858043>
- [68] ——, "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*, ser. CHI '17. New York, NY, USA: ACM, 2017, pp. 6481–6486. [Online]. Available: <http://doi.acm.org/10.1145/3025453.3025717>
- [69] K. Kiyokawa, M. Billinghurst, S. E. Hayes, A. Gupta, Y. Sannohe, and H. Kato, "Communication behaviors of co-located users in collaborative ar interfaces," in *Proceedings. International Symposium on Mixed and Augmented Reality*, 2002, pp. 139–148.
- [70] H. Kaufmann, D. Schmalstieg, and M. Wagner, "Construct3d: A virtual reality application for mathematics and geometry education," *Education and Information Technologies*,

vol. 5, no. 4, pp. 263–276, Dec. 2000. [Online]. Available: <http://dx.doi.org/10.1023/A:1012049406877>

Appendix A

Low Fidelity Study Design

Comparing Different Collaboration Scenarios

Problem

For the master thesis “Real-Time Augmented Reality Collaboration with Different Device Types”, a user study will be conducted. The study will require dyads and the majority of the potential participants can't spare more than an hour for the study. Moreover, due to the limited number of potential participants, within-group design will be used and participants will switch 2 devices in all 4 different combinations, which limits each variant with maximum 15 minutes. When the introduction, switching devices, survey and practice time included, each session will obviously have much less time. Therefore, it makes most sense to implement one test scenario that can answer research questions of the master thesis, instead of going through different ones. This user study is conducted to decide on which scenario to implement for the later user study.

Study Design

The master thesis has following research questions:

- How is the group coordination and performance affected by different device types during the real-time collaboration for different tasks?
- Do users prefer virtual or real reference points for coordination while collaborating with different device types?

Based on these questions, requirements for the main study are derived as:

1. Tasks should force users to collaborate and communicate.
2. Multiple device types should be used.
3. User study must have at least 2 different identifiable tasks.
4. Task should require participants to use spatial references.

Based on these requirements, two different scenarios are formed. To decide the most suitable one, low-fidelity prototypes of these scenarios are created. Since the 2nd question is the main topic of the later user study, it is ignored during this study.

Scenario 1: Construction

In this scenario, the users will build a simple model with using lego bricks.

Steps

1. Set of lego pieces will be scattered over different tables. The pieces will be located around the distinctive non-lego objects.
2. One user, “instructor”, will be given list of required pieces and not be allowed to touch lego pieces.
3. The instructor will direct the other user, “worker”, to collect the required pieces.
4. When the worker selects a piece, asks instructor to confirm if that is indeed the correct piece.
5. When all pieces are collected, the instructor will be given set of instructions that show how lego pieces should be attached.
6. The instructor will tell the worker what to do for each step till the model is completed.

Scenario 2: Matching Cards Game

In this scenario, users will play a version of matching cards game. This study is highly inspired by Müller et al. [1]. The pair of symbols will be printed on paper.

Steps

1. The symbols will be scattered over different tables. Empty sides of the papers will be visible and they will be located around the distinctive objects.

2. Users will decide who should select a paper, then the selected user will flip a random paper.
3. The second user will flip another random paper. If the symbols match, papers will be removed. If not, they will be turned back.
4. Steps 2-3 will be repeated until all symbol pairs are found.
5. Afterwards, users will be given a random symbol and asked to put papers back to the original positions.
6. Step 5 is repeated till all symbols will be re-located.

Interview

After the tasks are completed, a semi-structured interview with the participants will be done. Following questions will be asked:

- In your opinion, which of the tasks was the hardest?
- Did any of the tasks confuse you?
- Is there anything you'd change to improve completion time and accuracy in these scenarios?
- In which scenario you thought you have collaborated more? In your opinion, on which one would you perform better alone?

Research Questions

This study aims to compare collaborative aspects of two different scenarios. AR implementation will have some differences, for instance, during the first scenario, user will assemble a model with using a tablet or a headset, instead of using their hands directly or second scenario will use floating boxes instead of papers. That being said, the main aspects of these scenarios are similar in both the low-fidelity and AR versions. Results from this study will guide the design of the implementation.

Following questions will be answered based on the user study;

1. Is there any significant difference between the amount of verbal and non-verbal communication during different scenarios?
2. Is there any significant difference between the amount of verbal and non-verbal spatial expressions during different scenarios?
3. Does any of the scenarios take significantly longer to complete compared to another?
4. Do users think any of these scenarios required more collaboration?

To answer first two questions, user study will be video recorded and user behaviour, including verbal communication and physical gestures, will be analyzed. 3rd question will be answered with measuring task competition times. Answers of the interview will clarify the 4th question.

Hypotheses

1. Amount of total communication is expected to be same for both scenarios.
2. Second scenario have more spatial expressions, mainly because of the step 5.
3. Second scenario takes longer to finish.
4. Users think first scenario requires more collaboration.

Participants

Participants will be IT professionals who are familiar with the AR and collaboration.

References

- [1] Jens Müller, Roman Rädle, and Harald Reiterer. 2016. Virtual Objects as Spatial Cues in Collaborative Mixed Reality Environments: How They Shape Communication Behavior and User Task Load. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 1245-1249. DOI: <https://doi.org/10.1145/2858036.2858043>

Additional Notes

Matching Cards Game

Symbols used for the matching game: ☀️ 🎵 ⚡ € ♡ 🌸 ✈️ ➔

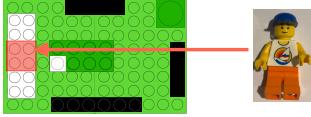
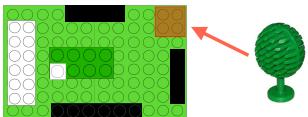
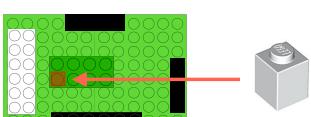
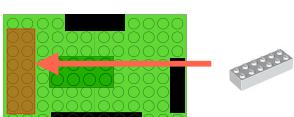
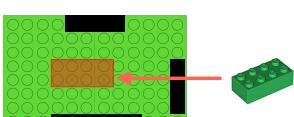
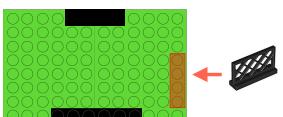
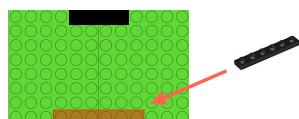
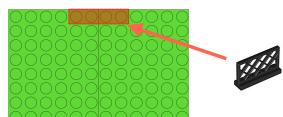
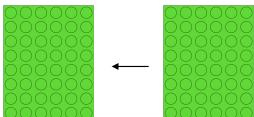
Construction

Required Pieces

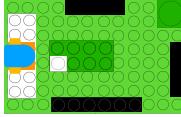
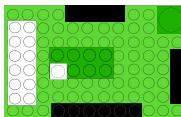
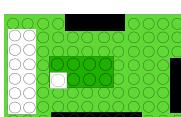
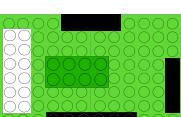
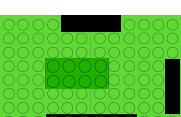
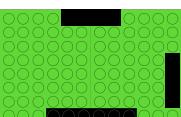
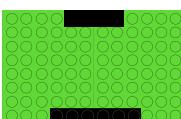
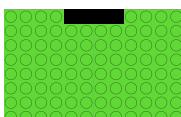
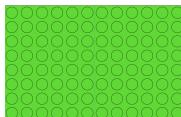
2x		2x	
Black Fence		6x8 green plate	
1x		1x	
1x6 black plate		2x4 green brick	
1x		1x	
Round green tree		Figure with blue hat and orange pants	
1x		1x	
2x6 white brick		1x1 white brick	

Instructions

Instruction



Result



Appendix B

Collected Data

Application Data

		Task completion times											
		Object Identification				Object Positioning							
		HoloLens	iPad	HoloLens	iPad	HoloLens	iOS	HoloLens	iPad	HoloLens	iPad	HoloLens	iPad
Device 1	HoloLens	05:56	05:51	05:56	05:51	05:10	04:45	06:06	06:37	06:19	01:49	01:31	01:31
Device 2	HoloLens	02:24	02:16	02:24	02:16	01:51	02:08	02:11	01:49	01:49	03:13	04:25	04:25
Study 1	05:09	02:45	03:07	03:32	02:18	02:51	03:44	02:21	02:21	02:08	02:08	01:47	01:47
Study 2	03:02	02:23	01:59	01:33	01:45	02:57	02:13	02:13	02:13	02:03	02:03	01:48	01:48
Study 3	03:07	02:49	04:08	03:02	02:48	02:48	02:13	02:13	02:13	02:03	02:03	01:48	01:48
Study 4	07:05	05:18	06:38	07:18	05:25	05:25	03:34	03:34	03:34	06:36	06:36	03:09	03:09
Study 5	02:24	03:07	02:22	02:27	02:27	01:50	02:08	02:08	02:08	01:26	01:26	01:09	01:09
Study 6	04:47	04:04	04:25	03:48	04:12	03:45	04:19	04:19	04:19	04:36	04:36	04:36	04:36
Study 7	01:44	02:05	02:12	02:37	02:08	03:59	02:21	02:21	02:21	01:12	01:12	01:12	01:12
Study 8	03:04	03:11	02:47	02:31	03:18	03:01	02:52	02:52	02:52	03:18	03:18	02:55	02:55
Average	03:33	03:24	03:34	03:17	03:14	03:18	03:20	03:20	03:20	02:55	02:55	02:55	02:55

Table 1: The task completion durations of different groups. Values are formatted as minutes:seconds.

		Error rates											
		iPads - HoloLens				Device 1 Device 2				HoloLens HoloLens		HoloLens iPad	
Device Types	HoloLenses	HoloLens	iPad	iPad	HoloLens	iPad	HoloLens	iPad	HoloLens	HoloLens	HoloLens	HoloLens	iPad
Study 1	1	2	2	1	1	0	0	Study 1	0,795 m	1,112 m	0,905 m	1,373 m	1,373 m
Study 2	14	6	8	13	9	16	16	Study 2	1,196 m	1,117 m	1,015 m	1,113 m	1,113 m
Study 3	5	3	5	5	2	3	4	Study 3	0,527 m	0,748 m	0,621 m	0,824 m	0,824 m
Study 4	1	1	2	1	0	4	4	Study 4	1,033 m	0,736 m	1,094 m	1,225 m	1,225 m
Study 5	2	1	0	2	5	1	1	Study 5	0,532 m	0,426 m	0,369 m	0,667 m	0,667 m
Study 6	3	0	0	0	0	4	4	Study 6	0,282 m	0,326 m	0,419 m	0,392 m	0,392 m
Study 7	4	3	3	3	2	4	4	Study 7	0,787 m	0,754 m	0,615 m	0,651 m	0,651 m
Study 8	4	0	0	1	2	2	2	Study 8	1,108 m	0,878 m	0,902 m	1,145 m	1,145 m
Study 9	2	0	2	4	0	10	10	Study 9	0,675 m	1,525 m	1,02 m	0,675 m	0,675 m
Study 10	4	0	2	1	0	6	6	Study 10	0,745 m	0,745 m	0,542 m	0,888 m	0,888 m
Total	40	40	52	52	50	50	50	Average	0,768 m	0,837 m	0,750 m	0,895 m	0,895 m

Table 2: The number of times a cube was selected with a non-matching pair after it already has been selected once.

Table 3: The average distance of the cubes from their original positions for each study

Video Data

Device Types	HoloLenses	Object Identification						Deictic speech			Object Positioning		
		HoloLens	HoloLens	iPad	iPad	iPad	HoloLens	iPad - HoloLens	iPads	HoloLenses	HoloLens	iPad	iPad - HoloLens
Study 1	9	7	5	1	2	1	14	9	1	7	10	10	10
Study 2	14	20	8	6	22	18	13	9	7	16	18	17	17
Study 3	2	1	3	7	3	3	7	7	7	12	8	16	16
Study 4	13	11	1	6	2	12	21	3	10	12	11	17	17
Study 5	17	4	4	9	8	10	17	8	3	5	11	5	5
Study 6	17	10	5	15	2	8	16	4	7	8	18	10	10
Study 7	3	7	2	7	1	0	2	1	4	5	4	0	0
Study 8	17	9	1	11	4	6	10	3	10	11	12	9	9
Study 9	7	3	4	3	2	10	10	4	4	10	7	11	11
Study 10	13	2	1	5	3	4	16	9	5	12	20	6	6
Average	11,2	5,4	7	7,2	12,6	5,75	10,85	10,1					

Table 4: The number of times participants used deictic speech (terms like "here" or "there") during different tasks.

Referring to Physical Objects

Device Types	HoloLenses	Object Identification						Object Positioning			Object Positioning		
		HoloLens	HoloLens	iPad	iPad	iPad	HoloLens	iPad - HoloLens	iPads	HoloLenses	HoloLens	iPad	iPad - HoloLens
Study 1	14	6	7	2	7	7	11	1	6	2	6	2	2
Study 2	0	1	0	1	0	0	3	2	0	0	1	1	0
Study 3	8	0	0	6	8	4	5	0	2	3	8	3	3
Study 4	0	1	0	1	0	1	2	0	1	1	1	0	0
Study 5	2	1	5	5	2	5	3	3	2	3	3	6	6
Study 6	14	11	3	4	6	13	12	2	3	7	7	3	3
Study 7	2	2	1	1	0	0	0	0	0	0	0	0	0
Study 8	6	0	4	0	5	3	1	0	0	1	4	1	1
Study 9	2	1	0	0	0	3	1	1	0	0	4	2	3
Study 10	6	9	1	1	1	12	7	3	1	4	2	4	4
Average	5,4	2,65	2,1	4,8	4,5	1,35	2,95	2,95					

Table 5: The number of times participants referred to physical object (e.g. "over the desk") during different tasks.

Device Types	Referring to Virtual Objects						Object Positioning					
	Object Identification			Object Identification			Object Identification			Object Positioning		
	HoloLenses	HoloLens - iPad	iPad	iPad	HoloLens	HoloLens	iPads	HoloLenses	HoloLens	iPad	iPad	iPads
Study 1	12	8	5	8	6	6	21	6	6	10	5	5
Study 2	1	0	0	1	0	0	1	0	0	1	0	1
Study 3	12	3	7	5	7	9	7	3	4	2	2	5
Study 4	5	4	1	2	0	3	4	3	3	3	0	4
Study 5	14	1	7	10	7	18	12	5	2	2	2	15
Study 6	16	8	12	6	18	29	10	2	8	7	9	8
Study 7	0	2	0	1	1	0	0	0	0	1	0	0
Study 8	10	3	4	4	4	8	4	0	1	6	3	2
Study 9	1	7	1	3	1	1	7	3	1	1	0	2
Study 10	9	9	3	6	0	9	8	0	0	10	2	8
Average	8	4,25	4,6	9,9	5,8	3	3,8	3,8	3,8	5	5	4,9

Table 6: The number of times participants referred to virtual object (e.g. "next to fridge") during different tasks.

Device Types	Hand Gestures						Object Positioning					
	Object Identification			Object Identification			Object Identification			Object Positioning		
	HoloLenses	HoloLens - iPad	iPad	iPad	HoloLens	HoloLens	iPads	HoloLenses	HoloLens	iPad	iPad	iPads
Study 1	24	11	8	1	6	3	19	2	6	0	11	5
Study 2	15	10	5	7	8	4	7	5	1	1	5	6
Study 3	3	0	2	10	14	6	2	3	5	7	12	5
Study 4	12	12	0	6	2	6	12	6	2	4	2	7
Study 5	8	3	2	0	11	1	6	3	2	1	4	2
Study 6	14	8	5	3	12	2	13	4	6	1	8	4
Study 7	2	5	1	4	0	1	3	3	2	1	0	0
Study 8	19	8	2	0	1	5	16	7	3	7	4	12
Study 9	7	1	3	2	1	4	1	4	2	3	0	2
Study 10	13	5	1	4	4	5	5	7	4	7	6	6
Average	11,7	4,6	5,9	3,7	8,4	3,85	4,2	4,2	4,2	4,9	4,9	4,9

Table 7: The number of times participants used hand gestures during different tasks.