HOLOVISION: ASYNCHRONOUS MIXED REALITY GROUPWARE TO SUPPORT PHYSICAL WORK TASKS

by

Alexander William Kienzle

Bachelor of Computer Science, University of New Brunswick, 2018

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Computer Science

in the Graduate Academic Unit of Computer Science

Supervisor: Scott Bateman Ph.D., Faculty of Computer Science

Examining Board: Saqib Hakak, Ph.D., Faculty of Computer Science, Chair

Weichang Du, Ph.D., Faculty of Computer Science

Zhen Lei, Ph.D., Department of Civil Engineering

This thesis accepted by the

Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

May, 2021

© Alexander Kienzle, 2021

ABSTRACT

Software that supports collaboration (groupware) is becoming ubiquitous in the workplace. The ability to share documents, images, and videos, or to have face-to-face conversations almost anywhere and anytime has transformed the workplace and increased productivity. At the same time, mixed reality devices are beginning to gain traction as viable platforms to support the production and consumption of new forms of information like spatial data and in-situ 3D objects. Current groupware systems focus largely on supporting virtual work, that is information work and manipulating information that can be displayed on the screen. However, many work tasks are physical in nature, they require manipulating, repairing, and assembling physical objects. The ability to share these new forms of information offers new and relatively under explored opportunities for collaboration around physical work tasks. Drawing on inspiration from previous research in the field, as well as 3D guidance techniques in video games, the mixed reality groupware system, "HoloVision" was created to take advantage of this new collaboration medium and provide a proof-of-concept for the design of future asynchronous mixed reality groupware systems. In this thesis, I document the background research, design, development, and testing of HoloVision as a novel mixed reality groupware system that combines speech-to-text, gaze, video, and spatial data, in a way that allows a user to intuitively author persistent "hypermedia" annotations to asynchronously support the guidance of future users in both collocated and remote workspaces.

Table of Contents

ABSTRACT ii

Table of Contents iii

List of Figures vi

1 Introduction 1

2 Related Work 8

2.1 Extended Reality Systems 8

2.2 CSCW Groupware 9

2.3 MR Groupware 13

2.4 Asynchronous MR Groupware 15

2.5 MR Guidance 16

2.6 Annotation Authoring 18

2.7 MR Session Recording and Replay 20

3 System Design 26

3.1 Summary of Previous Work 26

3.2 Conception 29

3.3 System Overview 33

3.3.1 Main Menu 33

3.3.2 Record Mode 34

3.3.3 Replay Mode 37

3.4 Interactions 41

3.4.1 Record Mode Interactions 41
3.4.2 Replay Mode Interactions 43

3.5 Design Considerations 45

4 Task Demonstrations 52

4.1 Scenario 1: Making Cereal 52

4.1.1 Record Mode Session: Cereal Making Example 53
4.1.2 Replay Mode Session: Cereal Making Example 59

4.2 Scenario 2: Apartment Tour 63

4.2.1 Record Mode Session: Apartment Tour 64
4.2.2 Replay Mode Session: Apartment Tour 71

4.3 Scenario 3: Plant Watering Instructions 77

4.3.1 Record Mode Session: Plant Watering Instructions 78
4.3.2 Replay Mode Session: Plant Watering Instructions 84

4.4 Scenario 4: Video Game System Setup 88

4.4.1 Record Mode Session: Video Game System Setup 89
4.4.2 Replay Mode Session: Video Game System Setup 95

4.5 Summary 99

5 Planned Study 101

5.1 Study Design 101

5.2 Study Part 1: Replay Mode 102
5.3 Collocated Replay Participants 103
5.4 Remote Replay Participants 104
5.5 Study Part 2: Record Mode 105
5.6 Results 106
5.6.1 Part 1 (Replay Mode) 106
5.6.2 Part 2 (Record Mode) 117
5.7 Summary 119
6 Discussion 120
6.1 Is HoloVision successful? 120
6.2 When would HoloVision be most beneficial? 123
7 Limitations and Future Work 125
8 Conclusion 127
Bibliography 129
CURRICULUM VITAE

List of Figures

Figure 2.1: The CSCW Groupware Matrix (Baecker, 1995). 11

Figure 2.2: A user wearing a HoloLens optical see through, head-mounted display. The wearer can see and interact with virtual objects placed in the physical world. Image from https://www.flickr.com/photos/sndrv/38564811131/in/photostream/ included under CC BY 2.0 30

Figure 2.3. HoloVision Data Diagram 35

Figure 2.4: Illustration of logged Annotation Data. 37

Figure 2.5: Apartment Spatial Map 50

Figure 2.6: Annotated Apartment Spatial Map 51

Figure 3.1: Scenario 1, HoloVision Menu (Record Session) 53

Figure 3.2: Scenario 1, Record Mode Menu 54

Figure 3.3: Scenario 1, Bowls 55

Figure 3.4: Scenario 1, Spoons 56

Figure 3.5: Scenario 1, Cereal 57

Figure 3.6: Scenario 1, Milk 58

Figure 3.7: Scenario 1, HoloVision Menu (Replay Session) 59

Figure 3.8: Scenario 1, Replay Mode 60

Figure 3.9: Scenario 1, Kitchen Annotations 61

Figure 3.10: Scenario 1, Spoons Annotation 62

Figure 3.11: Scenario 1, Spoons Video 63

Figure 3.12: Scenario 2, Kitchen 65

Figure 3.13: Scenario 2, Bathroom 66

Figure 3.14: Scenario 2, Office 67

Figure 3.15: Scenario 2, Bedroom 68

Figure 3.16: Scenario 2, Living Room 69

Figure 3.17: Scenario 2, Balcony 70

Figure 3.18: Scenario 2, HoloVision Menu (Desktop) 71
Figure 3.19: Scenario 2, Replay Mode (Desktop) 72
Figure 3.20: Scenario 2, Kitchen Annotation 73
Figure 3.21: Scenario 2, Living Room Annotation 74
Figure 3.22: Scenario 2, Living Room Video 75
Figure 3.23: Scenario 2, Balcony Annotation 76
Figure 3.24: Scenario 2, Office Annotation 77
Figure 3.25: Scenario 3, Window Plants 79
Figure 3.26: Scenario 3, Office Plant 80
Figure 3.27: Scenario 3, Watering Can 81
Figure 3.28: Scenario 3, Living Room Plants 82
Figure 3.29: Scenario 3, Tomato Plant 83
Figure 3.30: Scenario 3, All Office Plant Annotations 84
Figure 3.31: Scenario 3, Window Plant Annotations 85
Figure 3.32: Scenario 3, Window Plants Video 86
Figure 3.33: Scenario 3, Office Plant Annotation 87
Figure 3.34: Scenario 3, Living Room Plants Annotation 88
Figure 3.35: Scenario 4, Setup Tutorial 89
Figure 3.36: Scenario 4, HDMI to TV 90
Figure 3.37: Scenario 4, HDMI to Console 91
Figure 3.38: Scenario 4, Power to Console 92
Figure 3.39: Scenario 4, Power Bar 93
Figure 3.40: Scenario 4, Controller 94
Figure 3.41: Scenario 4, Setup Tutorial Annotations 95
Figure 3.42: Scenario 4, Components Overview Video 97
Figure 3.43: Scenario 4, Power Bar Annotation 98
Figure 3.44: Scenario 4, Controller Annotation 99

Figure 4.1: HoloVision Recording Session Data Snippet 109
Figure 4.2: Pilot Study, Task Set 110
Figure 4.3: Pilot Study, Drawing Task Annotation 111
Figure 4.4: Pilot Study, Calculator Task Annotation 112
Figure 4.5: Pilot Study, Scissors Task Annotation 113
Figure 4.6: Pilot Study, Scissors Task Video 114
Figure 4.7: Pilot Study, Scissors Task Annotation (Desktop) 115
Figure 4.8: Pilot Study, Calculator Task Video (Desktop) 116
Figure 4.9: Pilot Study, Drawing Task Video 117

1 Introduction

People rely on human-to-human communication in many aspects of life. As technology advances, so too does our ability to overcome temporal and spatial barriers to communication. The core focus of computer supported cooperative work (CSCW) is to overcome these two barriers when designing computing systems that focus on collaboration. Augmented reality (AR) systems (a subset of mixed reality (MR) systems) offer new ways to produce and consume information about our workspaces. Currently, much of the work done in MR focuses on remote, synchronous collaboration; this typically involves an expert remotely providing guidance to an MR user (Ens et al., 2019). In these scenarios, information such as video, audio, gestures, position, gaze, and field-of-view (FOV), can be captured and streamed between collaborators in an attempt to reintroduce information that is lost when collaborators are separated by time or space (Kim et al., 2019; Piumsomboon, Day, et al., 2017; Piumsomboon, Lee, et al., 2017; Ranjan et al., 2007; Sun et al., 2018). In addition, MR devices can display virtual objects (or holograms) that may be shared between collaborators. Some systems even support the recording of temporal and spatial information from a session, which can then be replayed to be analyzed or experienced asynchronously, in both remote and collocated scenarios (Nebeling et al., 2020; Thoravi Kumaravel et al., 2019). With this new ability to produce and consume rich MR media across both space and time, our task as human-computer interaction (HCI) researchers is to find applications that are both useful and usable.

For MR devices, the design and development of groupware (software systems that support collaboration) is an active area of research as a possible alternative or supplementary option for collaborators (Ens et al., 2019). For asynchronous collaborations, MR solutions can offer richer information than traditional forms of sharing information (e.g., sticky notes or videos) can provide, but they are still not widely adopted (Palmarini et al., 2018). Even while accessible devices like smartphones are now capable of supporting some MR applications, rich MR-specific features like spatial mapping do not yet seem to be robust or comfortable enough to outweigh the costs required to both produce and consume the information. Under the assumption that the physical hardware will improve, combined with the fact that there have been promising demonstrations of the benefits of using rich MR media in formal, controlled collaborative scenarios, one focus should be on improving user experience (UX) to increase adoption of these MR systems.

Information seekers prefer informal, unplanned interactions in collaborative work scenarios (Kraut et al., 2002). This implies that in many situations people will prefer to consume information when and where it is needed, which facilitates interactions and collaborations. However, this raises a question: "Why would the information producers choose MR over less costly, more familiar methods of asynchronous communication (like sticky notes or video recordings)?" Since people do not typically use MR technology, this means the startup costs of even simply accessing the technology (e.g., putting on a MR

headset, starting it up, waiting for software to launch and interacting with the software) is often too high to justify the benefits. To fully understand the potential benefits of MR groupware, we should focus on reducing interaction costs for both the information producers and consumers while leveraging MR technology features.

When designing a system to support the information needs of asynchronous collaborators previous work has proposed questions relating to who, what when, and where can provide guidance (Gutwin & Greenberg, 2002). For example, when designing new asynchronous AR groupware we should ask, "when and where would it be useful to place virtual information in a real workspace?". Imagine you are an information seeker, a newly-hired office worker; you just printed a document, but you don't know the location of the printer. To find it, you could ask a coworker or you could explore the workspace on your own. Asking a coworker would certainly be a quicker solution, but everyone nearby seems busy with meetings or calls, and you would prefer to not interrupt them. On the other hand, an independent search for the printer would likely take more time but would not require help from others. Depending on your personality, the work environment, and other contextual factors, you may try asking for help immediately, or you may search independently until you find the printer (or at least until someone offers help). Evidence suggests that you are more likely to do the latter (Kraut et al., 2002). Asynchronous groupware is useful when collaborators are not available at the same time; meetings, leaves of absence, or time zone differences are just a few reasons why real-time assistance

may not be available in a shared workspace. As a solution to this trivial problem scenario, AR groupware could be used to place persistent virtual annotations that could guide the user to the location of the printer, such that anyone could easily access the spatial information, if and when it is needed. While a solution like this may seem unnecessary in this case, consider all of the other information-seeking tasks a new hire may encounter. How would the worker locate a bathroom, a particular meeting room, or a coworker's desk? How would the worker accomplish these tasks if no other coworkers were available? Could similar information-seeking tasks apply in other workspaces, like warehouses, oil rigs, or hospitals? While people currently contend with these problems now, they could potentially be supported in more efficient ways (requiring less direct interaction and more independent work).

Asynchronous AR groupware should ideally be used in situations where it is useful for collaborators to produce and consume information about a workspace. AR can be differentiated from other media by its ability to capture and display spatial information (e.g., an interactive 3D model of an appliance, directly next to the real-world appliance). Not only should it be easy for the information seeker to successfully consume the information they seek, it should also be easy for the information producers to share relevant information within the space. For example, the procedure to repair an appliance can be demonstrated on a virtual 3D model, next to the real appliance to be repaired.

In many real world and MR applications, annotations are used to help highlight physical locations in space or on objects, and to share information relevant to those locations. These are used to replace the face-to-face equivalent. For example, a collaborator pointing at a location and sharing relevant information through a voice description. In terms of usability, annotation production should be just as easy as the face-to-face equivalent. It should also be easier and more efficient than creating signs, maps, or videos to inform users about a space. If an AR device is used like a video camera to record video, spatial, gaze, and voice data over time, it can be used to generate rich, persistent AR annotations about the space (Irlitti et al., 2013). If AR annotation authoring could be as simple as recording a video with a single AR head-mounted display (ARHMD), perhaps this could sufficiently reduce the barrier to producing such information. Similarly, this information-rich "hypermedia" could allow consumers to easily and independently explore workspaces for other relevant information. Overall, such a system would offer the production and consumption of persistent spatially distributed audio-visual information describing task-relevant features within a workspace. This system would allow users to produce and consume information about a workspace when other collaborators are not available.

In this thesis, we propose an asynchronous AR system called, "HoloVision", that introduces a novel approach to the production and consumption of AR session data to facilitate workspace-dependent task guidance. HoloVision is an application that runs on

the Microsoft HoloLens (1st generation), HoloLens 2, and Windows 10 PCs. The HoloLens versions of the application allows a user to record an AR session, similar to how one might record a video with a camera; the "expert" recording user moves around with the HMD and verbally comments about task components as they encounter the environmental components of the task within the workspace. Both the HoloLens (collocated) and the PC (remote) versions of the app allow a user to explore the information recorded during the AR sessions as persistent annotations placed within either the real space (ARHMD) or the virtual rendering (PC or virtual reality (VR)) of that space, respective to the device used. These computer-generated speech-to-text annotations are automatically placed on the recording-user's subject of speech, based on their gaze at the time of the voice recording. Annotations can also be selected to view a short point-of-view (POV) video clip that corresponds to the comment denoted in the annotation, such that the viewer may acquire deeper contextual information about a task component. The remainder of this thesis is laid out in the following Chapters:

- **Chapter 2 - Related Work:** A summary of research work related to this thesis. This work directly informed the design considerations of the system contribution of the thesis.
- **Chapter 3 - System Design:** An in-depth explanation of the design and development process of HoloVision, the system contribution for this thesis.

- **Chapter 4 - Task Demonstrations:** A practical demonstration of arbitrary use cases for HoloVision. This section illustrates the features and practical usage of the system contribution of this thesis using labeled images describing the use of HoloVision to accomplish trivial real-world asynchronous communication tasks.
- **Chapter 5 - Pilot Study:** A mock pilot study of the system contribution of this thesis. Due to restrictions on lab and participant access as a result of COVID-19, this section illustrates how HoloVision could be used in a formal user study. Study "results" are observations, anecdotal insights, and images generated from the primary researcher's performance with actual tasks using HoloVision. This section serves to inform future studies using HoloVision or similar systems.
- **Chapter 6 - Limitations and Future Work:** A summary of limitations of the system and thesis, as well as suggestions for improvement of the system or future systems in a relevant research area.
- **Chapter 7 - Conclusion:** A summary of the contributions of this thesis and the extent to which they were achieved.

2 Related Work

Below we cover background research to more fully describe the space of mixed reality-based collaborations. To do this, we first introduce the space of "extended reality systems", describe seminal work from the field of computer supported cooperative work and describe how these fields have been combined in Mixed Reality Groupware.

2.1 Extended Reality Systems

It is important to understand the distinction between the systems frequently mentioned in this thesis. The term, "Extended Reality" (XR) broadly encompasses systems that facilitate some combination of virtual and physical spaces. Some commonly referenced XR systems include: Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). Definitions for these systems vary slightly due to overlaps and gaps in hardware functionality and terminology decided on by manufacturers, but many definitions of XR systems commonly reference Milgram's concept of a "Virtuality Continuum" (Milgram et al., 1995). Milgram's continuum places XR systems on a one-dimensional axis ranging from reality (the real world) to virtual reality (a completely virtual interactive space). In between, we have systems that contain a mix of real and virtual; this is where some confusion may arise. For example, AR systems typically involve the real world enhanced by virtual objects, but the manufacturers' chosen names

and type of display, such as light projection, HMDs with passthrough cameras, or optical see-through (OST) HMDs can vary between devices. Depending on the device, virtual objects can be obscured by real objects, or real objects can be overlaid with virtual ones. Furthermore, some devices are configurable in such a way that they support interactions across the spectrum, so it can be difficult to place a particular device into any one category. This thesis focuses on an AR application created for the HoloLens and HoloLens 2 devices, which are OST HMDS that are commonly referred to as MR devices. From this point on, the thesis may at times refer to either MR or AR. Usage of MR is likely a broad reference to devices like the HoloLens, whereas AR likely indicates a narrower reference to the field of AR-specific research.

2.2 CSCW Groupware

Despite the innovations in workplace communication technology over the years, serious accidents resulting from miscommunication still occur (Palmarini et al., 2018). Groupware (software that allows people to work together to complete shared work) is common in many workplaces. For example, shared digital whiteboards allow people to sketch and develop ideas in an office, email and collaboration software (like Slack and MS Teams) allows people to work with others across great distances and/or at different times. Many different groupware systems, features, and standards exist for desktop computers, which make them popular and valuable in the workplace. However, desktop

computers, tablets, laptops, or smartphones typically require active, explicit effort to allow clear communication between workers. HMDs like the HoloLens 2 (which are designed with the intent to enhance workplace productivity) offer interactions based in physical environments, and leverage user gaze, gesture and voice input. However, standard interactions that leverage these types of sensed input are still being defined for these increasingly robust technologies, as they are actively being explored and improved upon.

A prominent framework for evaluating CSCW is described by the groupware matrix (Baecker, 1995); (Figure 2.1). This matrix divides groupware into four categories composed of two dimensions: space and time. In the spatial dimension, groupware tasks can be either collocated or remote. This means that tasks are performed by users in the same or different physical spaces. Each has unique obstacles and affordances when it comes to system design. The other dimension of the CSCW matrix is time. This dimension divides collaboration tasks into those that are synchronous and asynchronous. A synchronous task is that which involves collaborators working at the same time, whereas an asynchronous task consists of collaboration between individuals working at different times.

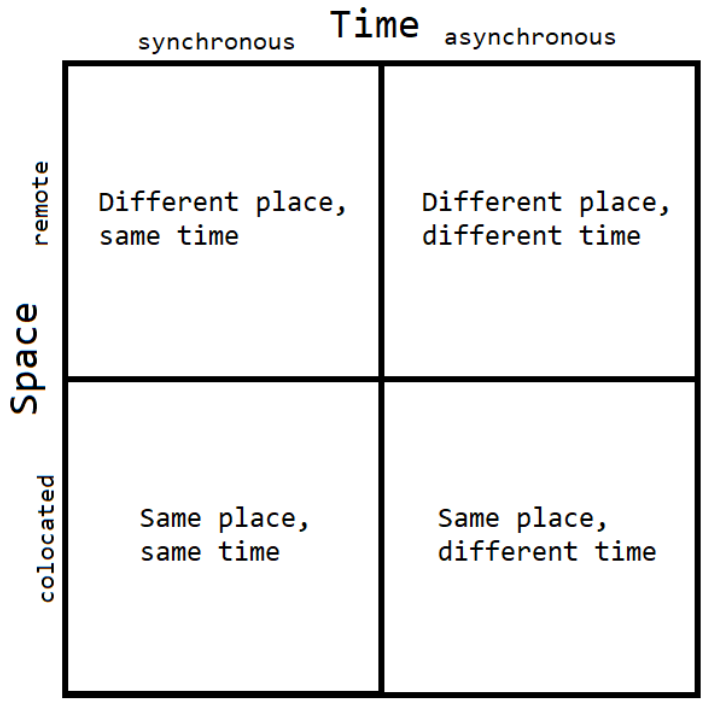

Figure 2.1: The CSCW Groupware Matrix (Baecker, 1995).

Spatial and temporal barriers often prevent us from communicating effectively, but these barriers can be overcome with groupware. Technologies like telephones, computers, and the Internet have allowed us to easily share more forms of information. To an extent, we have overcome the spatial barrier by sharing information in the form of text, audio, and video. The same applies for temporal barriers; asynchronous interactions within email, instant messaging, and group document systems act as effective forms of communication at different times. However, the adoption of this technology presents new

challenges. One such challenge is *awareness.* Awareness can be defined as "the up-to-date knowledge of the actions of others within a shared workspace" (Gutwin & Greenberg, 2002) . In physical collaborations, awareness is naturally conveyed through people's actions in a shared space (e.g., a mechanic can see the actions of another mechanic working on the same engine). However, when groupware does not sufficiently support awareness, it can lead to cumbersome collaborations and miscommunication (because people must focus on describing and sharing their actions, rather than the primary task they are working on).

During collaboration, a glance or gesture can provide additional context to a verbal comment. These subtle, in-person communications may passively update collaborators about the status of their workmate. In CSCW, "loosely-coupled" tasks involve this sort of infrequent communication between workers (S. Bateman, 2012; S. S. Bateman et al., 2013; Gutwin & Greenberg, 2002; Pinelle, 2004). Collaborators with these tasks often have distinct roles that may occasionally benefit from information from other collaborators. With a typical in-person workspace, this information is produced and consumed naturally. With groupware, this information encounters a "bottleneck" of sorts; it needs to be sensed and digitized by a computing system, and some of this subtle, implicit communication information is lost along the way. CSCW research is focused on informing design of groupware to create intuitive and efficient collaboration solutions, given the current hardware limitations. As new technology (like MR) becomes available, our

limitations are reduced; this provides us with opportunities to expand groupware systems to support work in new collaborative environments.

For workers who complete most of their tasks using a desktop computer, the frequent use of some form of computer supported collaboration is commonplace. The same may not currently be true for other types of work, but as the effectiveness and adoption of new computing technologies continues to rise, it is reasonable to expect that they will integrate into industries that can benefit from them (Ens et al., 2019; Palmarini et al., 2018). With improvements to technology that allows us to transmit more information, we may discover new ways to design our systems to further improve collaboration. Hands-free computers like MR HMDs do not require mouse and keyboard or touch input, they are capable of both capturing and displaying information within a physical space, and they can communicate wirelessly with other computing devices. Altogether, this suggests that MR groupware may have opportunities to expand into workplaces that have historically not been the focus of CSCW research (Ens et al., 2019).

2.3 MR Groupware

MR HMDs offer a different way to interact with computers. Traditionally, the 2D monitor, mouse, and keyboard have been the standard input-output computer peripherals. With smartphones and tablets, voice recognition and touch interfaces allow us to perform many of the same tasks that would normally be handled on a desktop computer. In

contrast, MR devices allow us to interact using information from gestures, gaze, or our physical environment. Instead of images on 2D screens, information can now be gathered from and displayed within a real workspace, where it is needed. Furthermore, HMDs support hands-free interaction, allowing users to perform physical tasks uninterrupted. When designing groupware for this platform, it will be important to focus on improving current collaboration conditions by leveraging these new MR-specific features.

Much of the current research in MR groupware focuses on remote, synchronous collaboration (Ens et al., 2019). The asymmetric "remote expert" use case of MR technology is a common example of this type of groupware (Billinghurst & Kato, 1999; Kwon et al., 2019; Piumsomboon, Lee, et al., 2017). Remote expert scenarios often involve a user wearing a MR HMD while performing some sort of task and are referred to as the "worker" or "local worker". Another individual, referred to as the "remote expert", can use a desktop computer, tablet, or a VR headset, to support the worker's task remotely, by observing and interacting with the worker through the worker's HMD. The remote expert attempts to provide guidance to help complete the task. Unlike with previous collaboration technologies, MR HMDs allow the expert to view the workspace from the worker's perspective through a live camera feed. The expert can also guide the worker using their voice or by providing support materials such as images, documentation, and/or virtual 3D objects. The MR HMD greatly facilitates access to support materials that by placing them virtually and directly in the worker's workspace (Piumsomboon, Lee,

et al., 2017). This type of collaboration is synchronous, as the interactions are typically performed while both the expert and user are actively collaborating. The collaboration is also performed remotely, leveraging the cameras on the MR device to allow the expert to see what the remote worker is doing and seeing.

2.4 Asynchronous MR Groupware

MR groupware research has focused heavily on remote, synchronous collaboration (Ens et al., 2019). This may be due to the obvious benefits of overcoming spatial barriers, reducing the cost of in-person interaction, and addressing many scenarios where email, voice, or video calls are insufficient. Conversely, there has not been much research focused on asynchronous MR groupware.

Work that has examined supporting asynchronous collaboration through MR, has often leveraged virtual annotations. Using spatial data from a workspace, MR users are able to place virtual objects in a physical space to act as guidance for others to view in the future (Hansen, 2006; Moghadam et al., 2016; Poupyrev et al., n.d.; Thoravi Kumaravel et al., 2019; Tseng et al., 2019; Vayanou et al., 2014). Unlike physical objects, these virtual objects exist as coordinates within a virtual version of the room. Due to their virtual nature, properties of these objects can be manipulated in ways physical ones cannot. For example, a life-sized virtual aircraft engine may be easily lifted, scaled down, stretched, or rotated to facilitate viewing and referencing. Virtual objects can also be programmed to appear or

behave differently for certain users. They can be placed passively within a room based on implicit user behaviour like moving, gesturing, speaking, or looking. Using position data, an MR system can be programmed to drop holograms on the floor of a workspace, like footprints. To design effective asynchronous MR groupware, we must further explore guidance using virtual annotations.

2.5 MR Guidance

In loosely-coupled task scenarios, collaborators seek information from each other to inform their decisions, but do so less frequently in more tightly-coupled forms of collaboration (S. S. Bateman et al., 2013). In these scenarios, problems can arise when access to this information is limited due to the availability of collaborators. When working remotely or asynchronously, collaborators can use groupware to overcome the space or time barriers, but the effectiveness of the collaboration is dependent on the quality of the information shared. With groupware, text, audio, and video are often shared to help collaborators build awareness, but for physical tasks, this information can lack spatial context.

MR guidance is the use of MR features, like spatial data, to help users complete tasks (Katz et al., 2012). Guidance approaches have been shown to be useful for "macro-navigation" tasks, like finding a room in a building, and for "micro-navigation" tasks, like

searching for a tool on a workbench (Katz et al., 2012). MR guidance cues could be generated using external data, like maps used by Google Maps AR.

Researchers have also looked for approaches in building MR guidance systems, by leveraging the large body of successful guidance techniques that have been implemented in commercial video games (Dillman et al., 2018). The intention behind the design of these guidance systems was to allow users to navigate a 3D environment in a way that is preferable to the user. By analyzing the systems used in 49 video games, Dillman has provided a framework to describe the types of visual interaction cues to classify how players are guided within 3D virtual space. This was done by defining three dimensions for visual cues: Purpose, Markedness, and Trigger. Each of these classifiers are further broken down into categories that describes a range of possible interaction scenarios in 3D space. This framework aids in informing the design of MR guidance features.

Social navigation cues can also be generated by collaborators; position, gaze, and hand gestures are just a few examples of MR information that collaborators may produce to communicate synchronously (Thoravi Kumaravel et al., 2019). Once produced, this data can be processed, visualized, and consumed immediately as live-updated holograms that simulate the behaviour of the producer. The data may also be stored to be consumed asynchronously, at a later time. For asynchronous visualizations, the data can be used to generate a “replay” of producer behaviour (Thoravi Kumaravel et al., 2019). Consuming

this replay is like watching a “holographic video” of what had occurred during a MR collaboration session. Since all this MR data can be captured and stored, perhaps then it could also be processed to generate persistent cues that summarize the behaviours of the producer as traces; where they went, where their gaze lingered, and what they spoke about.

2.6 Annotation Authoring

MR guidance relies on the placement of visual cues within a physical space. In games, these are often placed as traces of an actor’s past behaviour in that space. For instance, an animal’s footprints may be highlighted, a ghost may repeat a past conversation over his corpse, or a beacon of light may trigger a brief vision of a solution to a future problem in a particular room. In all of these scenarios, the player is provided with a visual annotation within the game world that informs about what happened (or what will happen) in that particular space (Wuertz et al., 2018). These visual annotations often represent an access point to a wealth of deeper contextual knowledge about a space. The footprints can guide the player to a new location. The ghost can trigger an audio clip. The beacon of light can trigger a video clip. In games, these annotations are placed both explicitly by the designers or players (Toups et al., 2014), or implicitly by the behaviour of players and other in-game entities (Wuertz et al., 2018).When designing MR groupware, it is important to have a clear understanding of annotations and to consider how they are produced and consumed.

In groupware, annotations are often explicitly placed by collaborators. In shared documents like Google Docs, a user can attach comments to relevant text or images. Similarly, in the MR space, collaborators may place holograms in relevant 3D locations within the workspace. This is one of the core features of synchronous MR groupware; the expert user may draw overlaid annotations to highlight real-world objects or manipulate virtual models to demonstrate a task. In this environment, it may be important to understand how MR annotations can be represented. With the video game examples, guidance cues are often updated in real-time; old footprints may fade away as new footprints are generated, or ghosts may disappear and reappear based on player proximity. These virtual objects are anchored to the space, typically based on previous actors' implicit behaviour, like moving or speaking. While these objects do not always carry explicit text, they can still carry commentary about the interaction history within a space. For MR applications, it may be helpful to broaden our definition of annotations to include hypermedia; virtual drawings, models, text, gestures, gaze, sound, and video each present a unique medium to comment about a physical space. With hypermedia MR annotations in mind, we can further explore how we might author them using MR devices.

Semantic Annotation of Georeferenced Environments (SAGE) is a virtual annotation system that supports the implicit authoring of text annotations within a virtual 3D version of a real physical space (Moghadam et al., 2016). A user may record their voice and their movement through a physical space using a passive handheld camera

system which also captures a spatial map of the environment. Essentially, this allows a user to create a labelled 3D map of a space by walking around and talking about it. Once the data has been captured, the labelled 3D map may be explored on a PC. If we apply this semantic annotation authoring concept to MR devices, we may be able to simplify MR annotation production. Furthermore, in addition to spatial maps, user position, and voice data, MR HMDs like the HoloLens supports the hands-free capture of gesture, video, and gaze data. This MR-specific data may be helpful in generating and placing the aforementioned MR hypermedia annotations. For instance, user gaze could be projected onto the spatial map to highlight key focal points, or first-person video clips could be placed to provide commentary that is more informative than text. Finally, with MR devices, these annotations may also be explored in-situ.

2.7 MR Session Recording and Replay

Like video data recorded with a camera and microphone, MR data can be recorded with MR-capable devices. A video is composed of a chronological sequence of 2D images to be observed in sync with the recorded audio. Together, this information acts as a window into a memory of a certain place and time. Like a video, a MR "session" can be captured using an MR device, like a HoloLens HMD. In addition to audio and video, this recording device also captures information about the physical space and the user's interactions with that space. Like 360-degree recordings, the information can be freely

captured with the appropriate device, but for some use cases, it may seem like an unnecessary novelty to explore that data. What value does this session information add, and does it outweigh the cost of using the technology? We can observe the value of 360-degree image exploration with applications like Google Street View. Unlike video, Street View images are accessed spatially instead of temporally. Could a similar concept be applied to MR recordings?

The MR system Loki provides an interface that allows an asynchronous user to "scrub" through a replay of both implicit and explicit interactions between objects, holograms, and collaborators from a collaboration session (Thoravi Kumaravel et al., 2019). Rich information about the session and the space is captured and displayed; information like user position, gaze, gestures, objects, holograms, and audio can be observed from any timestamp and any perspective using a holographic control system like those found in a typical video player application. The information provided by this interface is very dense; fine details of a session may be observed multiple times from different perspectives. This makes for a comprehensive session analysis tool, especially for instructor/learner sessions involving a single workspace. The level of detail with fine motor tasks involving objects, such as playing guitar, sculpting, and swinging a baseball bat have all been demonstrated as suitable use cases for this sort of system. From an annotation perspective, Loki allows a user to scroll through a history of many annotations within a space; all information is captured and converted into a consumable visualization.

Within this context, Loki can implicitly generate annotations from all information sources from a recorded MR session, such that the session may be simulated and experienced in the future.

Due to technological limitations, one of the biggest issues with the Loki system is its complexity, which ultimately detracts from its usability. To run it locally, a workspace must be set up with a Microsoft Kinect and the HTC Vive HMD, as well as the controllers and receiver (Thoravi Kumaravel et al., 2019). These devices must also be physically wired to a computer installed with the Loki application. When doing a remote session, this setup must also be duplicated at the remote workspace. While the authors suggest that these technological limitations will eventually disappear, it may be worth considering other approaches to MR session production and consumption in the meantime. Wireless, MR-capable devices like smartphones and the HoloLens can record MR session information like video, audio, gaze, and position. If we can produce this data over time, is there a simple, yet meaningful way to access it within a physical space?

Also consider the practical applications of MR recording tools. The Mixed Reality Analytics Toolkit (MRAT) is another system that supports the placement of MR annotations that visualize recorded MR data (Nebeling et al., 2020). The produced information is visualized within the space in a “raw” format. This offers developers a platform to analyze their recorded MR data in-situ. Keep in mind that information like head position, eye gaze, and hand gestures can only be easily visualized for consumption

in a 3D environment, be it virtual or physical. This means that an MR solution like the MRAT is ideal for visualizing MR data. With the Loki system, the raw session information is visualized for the purpose of training a user in a physical task (Thoravi Kumaravel et al., 2019). Like the MRAT, Loki records MR session data, then reconstructs it for consumption. The goal for this system was to record and simulate a real environment for an immersive training session, which could serve as a richer alternative to a recorded video session. However, it is important to remember the costs of using an MR recording technology relative to a simple video recording for learning tasks like those demonstrated in the Loki study. Because this system is limited to the fixed hardware in the workspace, this system would be insufficient for macro-navigation tasks on the coarse end of the spectrum.

The use cases for Loki were focused on training the user in an object-related task in a relatively empty workspace (Thoravi Kumaravel et al., 2019) . The value of this system seems to be in the granular spatial exploration of the task subject; the user may experience a baseball bat swing from many angles, or learn to build a product with remote assistance at a virtually shared worktable. In these scenarios, the task, the objects used, and the interactions between expert and learner are the focus. Loki assumes a controlled environment where a session about a particular task may be taught in relative isolation. The session is a chronological iteration of the information that was recorded, to be replayed and explored within a suitable controlled space. For Loki, the annotations are not

just spatially distributed; they are also distributed temporally. The user must observe a sequence of frames, like a holographic video. Like 360-degree video, users must explore both the spatial and temporal dimensions of the data. There is more depth and to this form of visualization, which increases the need for a user to explore and interpret the information they seek. For a task-based information seeker, this medium seems to provide sufficient information, but could search time be reduced further? Furthermore, could we simplify the process of recording and replaying MR session data?

Correia et al. have proposed a system, "HyperMem", which displays recorded information as spatially-distributed virtual annotations that contain text, images, videos, or objects that are relevant to particular museum exhibits (Correia et al., 2005). HyperMem displays persistent annotations in key locations within the physical environment. Instead of a workspace-wide temporal iteration of recorded session data, these annotations act as static, informational access points for each exhibit. From each, information can be accessed as desired by the user; text descriptions, photos, video clips relevant to the exhibit may be viewed or ignored, depending on task requirements. For a museum learning task scenario, this would mean that an information seeker only needs to approach annotations to interact with task-relevant data; they may choose to skip annotations they are not interested in learning about. This persistent visualization environment contrasts with Loki's visualizations, which change over time.

Consider a user who requires knowledge of a task to perform within an unfamiliar or frequently changing space; environment-based tasks involving building maintenance, inventory management, or firefighting exercises could potentially benefit from the placement of persistent, virtual annotations within a space to provide in-situ context to users' tasks. Google Maps AR is a macro-navigation focused system that uses location data to place such annotations using a mobile device camera. While it helps with location-based guidance, it does not provide deep, task-related information that a session-based MR recording system like Loki can (Thoravi Kumaravel et al., 2019). With HyperMem, persistent hypermedia annotations reduce the search complexity for information seekers, but the explicit authoring of such annotations takes time (Correia et al., 2005). Explicit authoring of annotations may not seem so bad in a relatively unchanging environment like a museum, where annotations may only need to be updated when there is a change to an exhibit. However, for workspaces that could evolve daily, hourly, or even every second, MR session recording system interactions would need to be lightweight, like the semantic annotations recordings offered by the SAGE system (Moghadam et al., 2016).

3 System Design

3.1 Summary of Previous Work

Standards for MR collaboration have yet to be established, but video game designers have already laid the groundwork by creating effective virtual 3D guidance cues. For asynchronous MR groupware, information can be left behind as virtual annotations within a workspace. These annotations can be placed in meaningful locations based on user position or gaze, and can contain information like POV video clips, images, or text that reflect the author's explicit actions or implicit behaviours. MR groupware systems have already been shown to be effective for task-learning. When considering physical work tasks, annotations are essential tools to help guide users through their tasks when separated by collaborators by time or distance. Our first system design requirement will be to create an MR groupware system to support this type of collaboration that is comparable in efficiency to the status quo for similar non-MR asynchronous groupware.

Laptops, tablets, and phones offer more mobility than traditional desktop computers, but these tools are insufficient for workplace tasks that require more than one hand. Furthermore, conventional 2D displays require users to shift their focus away from their task subject, whereas immersive displays offer spatial data that allows us to communicate about workspace-specific tasks more effectively. Because of this head-mounted augmented reality displays offer an ideal solution because they allow

information to displayed alongside physical work and allow both hands to be free. Together these characteristics allow people to better focus on their primary tasks and place less focus on systems to support collaboration. Our second system design requirement is to leverage desirable MR-specific features, such as hands-free interaction and in-situ display.

MR groupware systems have the potential to support features that offer rich media for efficient workspace-centric communication and collaboration. While groupware systems already support collaboration in modern workspaces, these are often designed to support planned, formal and heavyweight forms of collaboration. However, loosely coupled collaboration is often unplanned and informal in nature and favours lightweight and easy to use tools. Our third system design requirement is to provide a simple, low-cost way to produce and consume asynchronous MR annotations (collocated or remote) such that the system is as easy to use as status quo approaches.

Based on the review of related work, in this thesis we seek to create a groupware system specifically designed to support loosely-couple collaborations in physical scenarios (i.e., scenarios involving work on physical objects) that satisfies the following requirements (labelled R1-R3, for future reference).

R1: Facilitate the production and consumption of asynchronous MR annotations to support effective task guidance relative to status quo approaches.

R2: Support hands-free interactions to enable people completing physical tasks to interact with the system and receive guidance while using their hands to complete the primary work task.

R3: Provide a low-effort interface for authoring and viewing annotations to. To support loosely-coupled collaboration, collaborators need to have a minimal effort means to share and access information, otherwise they will opt to forgo the collaboration and choose to work in a less optimal scenario.

3.2 Conception

The original inspiration for HoloVision was derived from the 3D guidance cues typically used in video games. Game developers implement annotations and traces within the virtual space to guide the player through the space in a way they intend (Thoravi Kumaravel et al., 2019). These cues can blend seamlessly into the environment or be emphasized under the pretense of a player character's supernatural abilities. Footprints, scents, or ghosts may be temporarily visualized within the space to encourage the player to move, look around, or perform a specific action. Typically, "enhanced vision modes" can be toggled by the player when they want to know more about their immediate environment in the game. Countless hours of experience with the helpfulness of such systems, together with access to the HCI Lab's HoloLens and HoloLens 2 are what originally motivated our approach to beginning a project to create a system that satisfies requirements R1-R3 stated above. Below, the features are of the system are described and are related back to the requirements extracted from previous work (see 0).

For MR applications, data is visualized within a 3D space, so it seems natural to attempt to design a system that implements some of the core features of typical enhanced vision modes from video games. With an MR device, such as an optical see-through (OST) HMD like the HoloLens (see Figure 3.1), virtual objects can be visualized in and integrated with the user's real space. Where traditional groupware is restricted to 2D display (be it handheld or desktop), MR has the potential to offer more spatially aware,

hands-free, in-situ interactions that were previously not possible (requirement R1). Essentially, a MR application on a HoloLens could act as an enhanced vision mode, provided that the necessary data was stored. With a tool like this, remote or collocated workers could access spatially relevant task information, even if a task expert is not available, while still maintaining use of their hands to complete the task. Hence, we had an idea for a MR application, “HoloVision”; a system that could facilitate the production and consumption of holographic guidance cues within a workspace (R2).

Figure 3.1: A user wearing a HoloLens optical see through, head-mounted display. The wearer can see and interact with virtual objects placed in the physical world. Image from https://www.flickr.com/photos/sndrv/38564811131/in/photostream/ included under CC BY 2.0

While some MR groupware for session recording and playback already exist, the focus has been creating a robust, comprehensive system that supports each aspect of the

CSCW matrix (see Figure 2.1). In projects like Loki, there has been an earnest attempt to capture and display as much MR session information as feasible, to evaluate the technology's potential in simulating an "ideal" immersive groupware experience (Thoravi Kumaravel et al., 2019). Systems like this showcase the information produced using MR (requirement R2), but as described earlier, the system presents considerable overhead costs when it comes to setting up the MR recording workspace. While it was found that Loki was helpful in teaching tasks, it can also be argued that the information produced by the system is unnecessarily dense for the user to explore. Consider 360-degree videos versus typical 2D videos; 360 offers richer information than its predecessor, yet users may find that more time is required to explore the video to consume information they seek. With rich media like this, the user can explore spatially in addition to typical time-scrubbing exploration. Overall, systems like Loki demonstrate the value of MR session recording, but it also suggests a need to reduce interface complexity. This system satisfies R1 and R2, but not R3.

On the other hand, the SAGE system offers a relatively low complexity session recording solution (Moghadam et al., 2016). By recording voice, user position, and spatial map data during a session, it has been shown that speech-to-text annotations can be mapped to a 3D environment. However, the SAGE system does not use a MR HMD, so sessions are not visualized in-situ (R2). This system offers a low-complexity interface design that supports the recording of 3D annotations, user position, and spatial maps, to

support macro-navigation. Unlike Loki, this system does not allow users to view information-rich animations, nor does it allow the user to explore annotations in-situ. This system satisfies R1 and R3, but not R2.

For HoloVision, we wanted to create a user experience that simplified the recording and the replaying of an MR session that covered the spectrum of guidance cue purpose, from room-to-room ‘macro-navigation’ to desk-level ‘micro-navigation’. Virtual annotations should be information-dense but should also still be easy to create and explore. Semantic authoring offers a low-cost method of session recording; using a HoloLens, spatial maps, voice, video, position, gaze, and gesture can be passively recorded during a session. Using that data, we could then try to generate information-rich annotations in appropriate locations. Instead of just using position, we could experiment with gaze direction or gestures to enrich the annotations with contextual information that help us understand the actions of the author. Instead of just text, we could use video clips that are placed in the physical location where they occurred. Annotation sizing and placement could offer macro-navigation cues like the locations of rooms or fixtures. The rich annotation content like speech-to-text and video clips could offer rich micro-navigation cues, like the orientation of a tool and how to use it. The goal in creating this system was to satisfy all requirements (i.e., R1, R2, and R3) by making an easy-to-use application to record MR data to be visualized later (asynchronously) with the purpose of guiding a future user in the completion of a workspace-related task.

3.3 System Overview

HoloVision is designed as groupware that supports asymmetric, asynchronous collaboration in both collocated and remote environments by allowing users to produce and consume rich, in-situ annotations (R1 and R2). It was developed using the Unity game engine, and runs on Microsoft HoloLens, HoloLens 2, and PCs running Windows 10. It has two modes of operation that each focus on one aspect of asynchronous communications, “Record Mode” (for sharing information) and “Replay Mode” (for receiving and using shared information). The HoloLens version of the app supports both Record Mode and Replay mode for collocated sessions. The PC version supports Replay Mode for remote navigation of a virtual version of the session space. Besides the need to import a spatial map model of a workspace (.obj) and the session data (.json and .mp4) from the HoloLens to the PC for the remote desktop version, the application requires no additional hardware, software, or configuration to function once installed on the device (R3). See Figure 3.2 for a comprehensive data diagram of HoloVision.

3.3.1 Main Menu

Upon application launch, the user will be brought to HoloVision’s main menu, where they may select either Record Mode or Replay Mode by selecting one of the respective buttons using the HoloLens “air-tap” interaction. On the desktop version, only Replay Mode is available, and buttons are selected using a mouse.

3.3.2 Record Mode

Record Mode allows a user to record MR sessions within the real space. Record Mode facilitates the production of MR session data. During a session, the user may move freely, look around, and make verbal comments about the environment. In addition to a first-person video of the session, information about the user's position, gaze, speech, gestures, and space are all recorded at a half-second interval for later use. The half-second recording interval was chosen as a sensible balance between precise data capture and application performance. In Record Mode, multiple sessions may be separately saved or overwritten as desired. Besides the video, the session data is stored in a .json file, so the information can be easily edited if necessary, like if the speech to text interpretation was inaccurate.

HoloVision Data Diagram

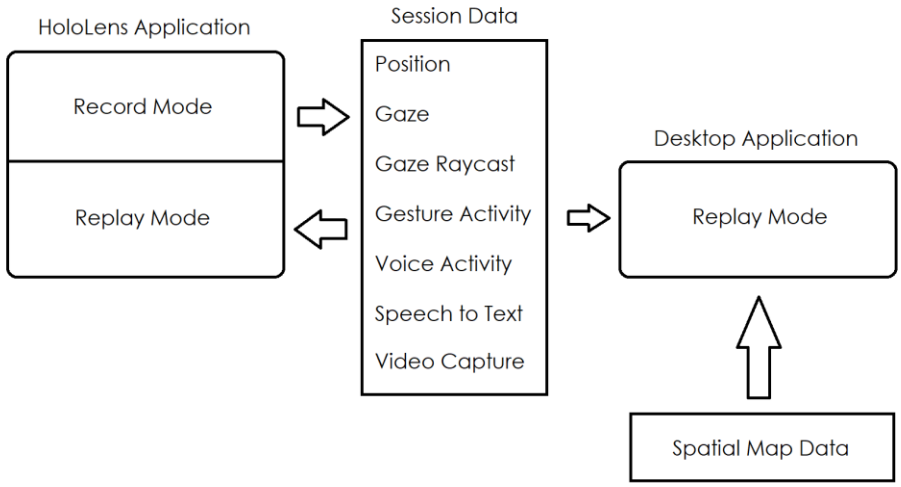

Figure 3.2. HoloVision Data Diagram

HoloVision session data can be produced and consumed on the HoloLens. The data may also be replayed for consumption on the PC version, provided that the session data and appropriate spatial map have been imported (see Figure 3.2).

The Record Mode interface consists of “Stop” and “Start” buttons, as well as “Recording ID” selector buttons to increment/decrement the current integer ID to be associated with the recording session. There is also a “Back” button that allows the user to return to the main menu. After selecting a recording ID, the user may select the Start button to begin their recording session. During each half-second of the recording, the position of the HMD relative to the origin point (the HMD position when Record Mode was opened) is stored as a Vector3 (a 3-dimentional vector object containing x, y, and z coordinates). Similarly, the user’s gaze is also stored as a Vector3, representing the directional vector that matches the direction that the user’s head is facing. Gaze raycasts are Vector3s that result from the projection of the user’s gaze vector onto the HoloLens-generated spatial map mesh object. These 3D positions represent approximately where the HMD user’s gaze fell on the physical environment throughout the recording session. Gesture activity is a per-interval boolean value determined by the presence of the user’s hands within view of the HMD cameras. If the user is making a gesture, like pointing, waving, or reaching, this value will be true. For gesture, I found that the system would return many false positives, as it would often detect hand movement as gestures when

walking, so gesture data was not used to inform annotation placement. Voice activity is a boolean value that returns true when user speech is detected. Speech-to-text data is the string data generated by the DictationRecognizer class. Although it is not always completely accurate, the result provides a decent string representation of the words that were spoken by the user during the recording session (these are stored as plain text and can easily be corrected if desired, post-recording). An empty string is returned when the user is not speaking. Meanwhile, an .mp4 video of the entire session from the user's POV is recorded (using an HMD camera) and stored when the user selects Stop.

Aside from the video file, the session data for each recording is stored sequentially in a .json file. For each half-second interval of recording, there is a datum available for each data type. For instance, if a user made a thirty-second recording, the .json file would contain sixty objects, each containing data for position, gaze, gaze raycast, gesture activity, voice activity, and speech-to-text. Each object essentially contains a snapshot of the recording user's behaviour for each half-second of the recording. Note, the speech-to-text data is inaccurate in this regard, as strings are only recorded after an utterance has been processed by the DictationRecognizer, typically a second or so after the user has finished speaking. To account for this, the voice activity (a boolean value) is also recorded at the true time of an utterance, to help map the speech-to-text data on the session timeline (see Figure 2.1).

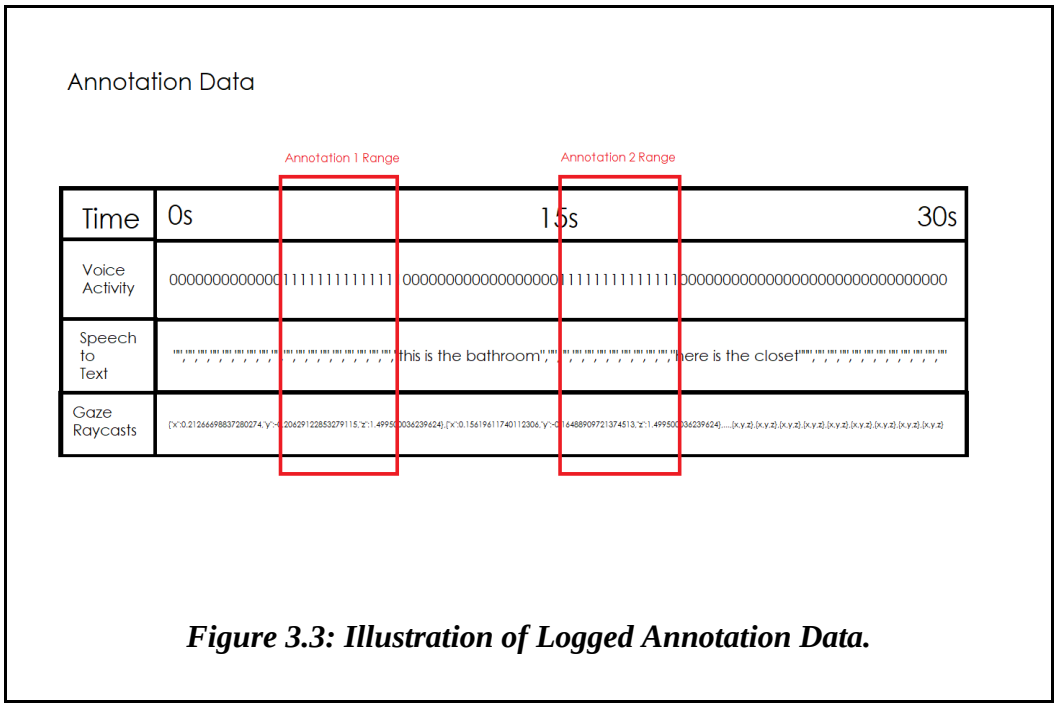

Figure 3.3: Illustration of Logged Annotation Data.

3.3.3 Replay Mode

In Replay Mode, a user may consume the information that was produced during a chosen Record Mode session. In HoloVision's initial design, Replay Mode was limited to the visualization of MR position data. Like video game scent trails (Dillman et al., 2018), position data is visualized as semi-transparent, colored particle clouds. For 3D position data, the points are relative to the origin point of the virtual space. Since the data was recorded at the HMD's position, the raw data points contain the positions that the Record Mode user's head was, assuming that the origin point of the current space matches that of the past space. After an iteration of system testing, it became clear that the clouds should be adjusted slightly downward to prevent occlusion by giving the user a better viewing

angle. Other visualizations such as heatmaps, line traces, and discrete 3D objects were also explored tested in the test, but particle clouds seemed to be the most preferred form of guidance; they show a continuous path, the cloud is both easy to spot in terms of location and size, and its transparency reduces occlusion of both real and virtual objects. It was tempting to use floor-level "footprint" traces, which are both intuitive and typically used in video games as emphasized guidance cues. However, due to the limited FOV of the HoloLens, it is not possible to see nearby floor-level traces without looking directly down; an action that is both uncomfortable and potentially hazardous (due to unseen obstacles) while wearing the HoloLens.

After position traces, Replay Mode was expanded to include rich, semantically authored annotations. Position traces offer guidance in coarse macro-navigation (e.g. finding a room in a building) tasks, but fail to cover finer micro-navigation (e.g. finding an object on a desk) task guidance. First, annotations needed to be placed in a suitable location. Like the SAGE system (Moghadam et al., 2016), speech was used as the key indicator of annotation placement. During a recording session, a verbal utterance from the user suggests that attention should be brought to that specific place in space and time in the recording. Within that range of voice recording time, it seemed reasonable to assert that the focus of a user's gaze could always be the subject of their speech. Since a verbal utterance is ideally an explicit interaction, this could serve as a user's cue to the system to make generate contextual annotations in a way that feels natural. By calculating an

average position from the gaze raycasts stored within the voice recording time range, we are given a single 3D point representing the speaker's approximate focal point during the time of their utterance. Given the available data, this method of annotation positioning seemed intuitive; the annotation is placed where the recording user was looking when they created the annotation, so it helps ensure a reasonable viewing angle and can provide spatial context to the utterance (R2). This placement method also reduces the need for the viewer to determine context by observing the recording user's behaviour. For example, in a real-time collaboration, an information seeker would typically need to follow the teacher's gaze or gestures toward a subject to extract context, which may leave room for confusion. With gaze-based semantic authoring, the annotation should be centered on the subject. This has the potential to eliminate potential context miscommunication at the cost of imposing an additional usage requirement on the user.

Voice activity data is the key indicator used to create annotations. The average of the gaze raycast Vector3s from these annotation ranges are used to determine annotation placement within the space. The speech-to-text string that immediately follows is used as the annotation label.

Centered on the average gaze raycast positions, speech-to-text results are placed as holographic text labels. During a recording session, if the user said "testing, one, two, three" while gazing at a wall, the ideal resulting annotation would be a text label that says "testing, one, two, three", centered where the user had gazed. To capture speech-to-text

data, I used the *DictationRecognizer* class from *UnityEngine.Windows.Speech*. The speech data occasionally fails to detect some words properly, but the resulting annotations are otherwise intelligible and correctly placed on the speech subject. During a Replay Mode session, a persistent text label is generated for each distinct verbal utterance. This distinction is made based on arbitrary (two second) time gaps in speech data. Essentially, this two second gap in the speech recording data suggests that the user has finished speaking about something. When configuring this time delay, I found that two seconds was short enough that I did not need to wait too long before the next utterance, but long enough to not accidentally fragment a sentence. There is more work that could be done to implement a less naive approach to determining more accurate speech-to-text labels in HoloVision.

With speech-to-text annotations placed on gaze raycast points, the system now supported semantically authored text annotations within the space. However, the text content is not always accurate, and doesn't provide rich, visual context like the hologlyphs offered by the Loki system (Thoravi Kumaravel et al., 2019). In video games, annotations often represent interactable points that trigger animations or dialogue, like the hypermedia access points discussed in HyperMem (Correia et al., 2005). To accompany each annotation's text, I implemented a VideoPlayer object (from UnityEngine.Video) to display a video clip taken from the relevant range of time in the stored video recording. These video annotations are represented by static "thumbnail" panels behind each text

label. Each thumbnail panel is textured with an image of the first frame of a video clip that, when selected, shows a brief POV video of the recording user's behaviour at the time of their utterance. Referring to the previous example, the clip would show what the HoloLens camera recorded around the time when the user said "test, one, two, three". The video clip ranges are arbitrarily extended by one second at both the beginning and the end to ensure that the entire utterance is captured in the clip.

3.4 Interactions

3.4.1 Record Mode Interactions

HoloVision can be used to record and virtually display information within physical space. It acts as a platform to record and replay MR sessions for later consumption. Similar to using a video camera, an expert may begin recording the session by pressing a button using the HoloLens "air-tap" gesture (for all HoloVision button presses, the user must gaze at the button and perform a simple tap gesture with their finger). This interaction could also be performed with voice commands, if a 100% hands-free system is essential. During a recording, the expert is expected to make verbal comments as they move and look around within the space, with the purpose of informing a future information-seeker. Until the recording user selects the stop button, voice, video, position, gaze, and other

information is passively recorded (see Table 3.1). Like other recording systems, certain conditions must be met to ensure an ideal recording environment:

- The lighting in the workspace is sufficient; the HoloLens has difficulty creating spatial map data in dark environments. Without sufficient spatial map data, gaze raycasts (and the resulting annotation positions) will not be accurate.
- The user only speaks when they wish to leave an annotation behind; speech communication from the user to the system must be explicit, as it is the key indicator to the system that an annotation should be made.
- The user speaks and enunciates clearly to reduce errors in speech-to-text data.
- The user briefly focuses their gaze on the subject of their verbal commentary within the space while speaking. Annotation placement relies on gaze, so the user should focus on localizing their gaze near a location that they would like their annotation to be placed.
- The user briefly refrains from speech for at least two seconds after completing their commentary for an annotation. Annotations are segmented by two-second gaps in speech data, so a pause ensures that annotations will be segmented properly. Conversely, if the user would like to create a long annotation, they must avoid two-second-or-longer pauses for the duration of their commentary.

Table 3.1. Information produced during a recording session

Information	Format	Current usage
Position	Vector3	Movement trace position
Gaze direction	Vector3	Unused
Gaze raycast onto spatial map	Vector3	Annotation positioning
Gaze raycast normal	Vector3	Unused
Speech to text	string	Annotation text, voice commands
Speech activity (is the user talking?)	bool	Annotation segmentation
Gesture activity (is the user gesturing?)	bool	Unused
Video capture	mp4 (30 fps)	Video clips, thumbnails

3.4.2 Replay Mode Interactions

Like Record Mode, Replay Mode requires that the user selects the start and stop buttons for a chosen recording ID (also selected with buttons). The replay session generates persistent annotations using the recording data. Purple clouds represent traces of user movement, red clouds highlight key focus points of user gaze. Annotations are placed using a combination of speech and gaze data (see Table 3.2). On each annotation, speech-to-text data and video thumbnails are placed. These show approximately what was spoken and what was seen at the time and place where and when a verbal utterance occurred during the recording session. Like buttons, these annotations may also be selected; doing so triggers a brief video clip (a calculated range of frames from the video data) that shows what happened during that part of the recording. Like the gaze range calculations for annotation positioning, the video clip time ranges are calculated relative

to the voice activity data, such that the video clip encompasses the entire verbal utterance for that annotation. Simple video navigation buttons, such as Play, Pause, Stop, and Replay, were also implemented to appear below the video annotations. This feature was planned (along with a red dot cursor) for use in a final evaluation of the system, in reaction to usability issues in pilot testing, but they do not appear in the screenshots from earlier iterations of the app.

Since HoloVision's position data is mapped relative to the origin point of the recording user, this means that the replay user must start in the same position that the recording user started, in either the virtual or physical space (desktop or HMD). This means that the spatial map must be manually positioned on the desktop version. On the HMD version, this means that the replay user must ensure the HMD is positioned where the recording user had the HMD when the recording began. To overcome this limitation, I ensured that I always used the same "start position" for all recordings. For future work, this limitation could perhaps be overcome using spatial anchoring and spatial understanding solutions that allow the system to "remember" the space relative to the device between sessions.

HoloVision also supports proof-of-concept voice command features. Buttons (like Back, Start, and Stop) can be selected by speaking the command word, "command", followed by the word shown on the button. Annotations can similarly be selected by speaking the command word, followed by the annotation label id (e.g. "L1"), or the words

shown on the label. Holograms like thumbnails, labels, and traces could also be enabled and disabled with voice commands, using the command word followed by “labels off”, “thumbnails on”, etc.. During task demonstration recordings and the mock pilot study, these features were not stable enough to ensure consistency throughout sessions, so the features were disabled.

3.5 Design Considerations

Gaze direction is unused because the focus of a “lingered” gaze (gaze raycast onto the spatial map) is already an improvement on the coarser gaze direction information, which is has previously been used in synchronous MR collaboration (Thoravi Kumaravel et al., 2019). In a typical face-to-face collaboration, gaze information can inform the user about the focus of a collaborator’s speech; if the speaker is gazing at something while speaking, they may be providing implicit context to their words. Similarly, the words spoken, and gestures made can also provide context by bringing the listener’s attention to the area of focus within the space. With mixed reality, gaze raycasts and virtual objects can be used to emphasize precisely where the speaker was looking, eliminating the need for the listener to extrapolate eye gaze or ask for clarification. For this reason, this feature serves as low-cost way to leverage spatial and gaze data to convey meaningful information to the user (R2 and R3). Raycast normal data was included in the recording in case surface-detection was needed for label orientation, but wall-oriented annotations are not always

visible to the user. Annotations are easier to see if they are actively updated to face the user, so this method of orientation was chosen instead (R3).

Gesture information is currently unused because the recorded data was not as informative as speech when used to bring attention to an area in space. When using the HoloLens, gesture activity occurs when the HMD user's hands are within view of the cameras. Gesture activity is often a byproduct of common movements like walking or reaching. False positives like this are not helpful when trying to extract user intention towards areas within space. For this information to be helpful, the HMD user should refrain from allowing their hands from entering the view of the cameras unless they wish to signal annotation intent to the system. However, this condition makes recording sessions interactions feel rigid and unnatural, which defeats the purpose of semantic authoring. Given that speech was already working well as an annotation authoring indicator, there did not seem to be a need for the gesture data (R3).

When using speech as the sole indicator for annotation authoring, the user must remain silent unless they would like to signal their intent to place an annotation. To make recordings more tolerant of mistakes, it could be helpful to aggregate the speech data, with something else, like gestures. This would remove the "stay silent" condition but add a "must gesture while speaking" condition. Again, this begins to make interactions more conditional and less intuitive for the user. Some people do not gesture while speaking, and others may end up accidentally authoring annotations if their hand is in view. HoloVision

recording sessions are intended to be initiated explicitly, like video recordings. Because of this, it makes sense for users to follow a certain protocol (like staying silent) when making a recording. However, if similar technology were to be developed further, it would introduce a need for a more complex method of extracting user intent from their recorded behaviours; perhaps involving a weighted analysis of gestures, gaze, speech content, or even computer-vision analysis of video frames. That said, these suggested features could ultimately increase annotation production cost, so these features were placed out of our system requirement scope (R3).

Overall, our system design considerations had a strong focus on reducing the interaction costs that have previously burdened the user (R3). Because MR systems have so many new features and capabilities, it can be tempting to design a system that takes full advantage of the hardware. However, in doing so, the systems can become cumbersome for users, which makes them infeasible as improved groupware alternatives. Looking back to the printer example in the introduction, if a complex system like LOKI seems more difficult to set up than the actual printer; so, why would anyone choose to use it over the status quo? Information seekers prefer loosely coupled interactions; they need information where and when it is convenient.

Throughout the development of HoloVision, we sought to support convenient interactions for both the production and consumption of annotations, such that authoring and viewing the information is comparable in complexity to writing a note or taking a

video with a smartphone. While the tasks described in this thesis are relatively trivial, one can interpolate how more complex scenarios like instructions for industrial machine maintenance or operation procedures could be improved using a similar system. Consider the effort that would be required to produce a useful engine repair video: How would you hold the camera while repairing the engine? How long would the video be? Would you need to edit the video before sending it? With traditional systems, conveying this information would be more convenient for the producer with an in-person tutorial, but in-person meetings are not always feasible.

Table 3.2. Information consumed by a future user

Annotation	Description	Opacity Priority
Position traces	Transparent purple particle clouds are placed below the position data points, such that they are approximately waist-level.	*Low*: Faded unless markers and labels are off.
Gaze markers	Transparent red highlights that emphasize points of interest. These points are calculated using the average of sufficiently-close gaze-raycast points that were recorded during a period of speech activity. The positions of these markers are also used for placement of labels and thumbnails.	*Medium*: Faded if labels are off.
Text labels	Text generated by the HoloLens DictatonRecognizer during a recorded session is segmented into separate labels based on gaps in speech data. If the expert speaks for several seconds without pausing, that speech text will be grouped together as a single label. Labels are scaled based on the recording user's relative distance from the point of interest at the time of recording, such that more distant annotations will be large and nearby points of interest will be smaller. Label scaling was added to reduce overlap of micro-navigation task annotations and increase the locatability of macro-navigation task annotations. Labels are constantly updated to face the direction of the replay user. Clustered labels are rearranged in a grid formation to prevent overlap.	*High*: Never faded.
Thumbnail images	A snapshot of the recording user's POV at the time of recording is placed slightly behind the text label of an annotation. Like the text labels, they are scaled based on relative distance between the recording user and the point of interest. Thumbnails are constantly updated to face the direction of the replay user. Clustered thumbnails are rearranged in a grid formation to prevent overlap.	*High*: Never faded.
Video player	Labels can be selected (by voice or gesture) to trigger a video player that shows a clip from the expert-POV video capture of the recording session. The range of frames for the clip are selected based on the range of time that speech occurred for that annotation. The player faces the user and is placed between the user and the selected label.	*Not applicable*, visible and fully opaque when triggered.

During a session replay, traces, markers, and labels may be activated or deactivated via voice commands (or keyboard shortcuts on desktop) as needed. All can be displayed at once, and all but the most dominant visualization will be faded (more transparent) to prevent occlusion and visual clutter.

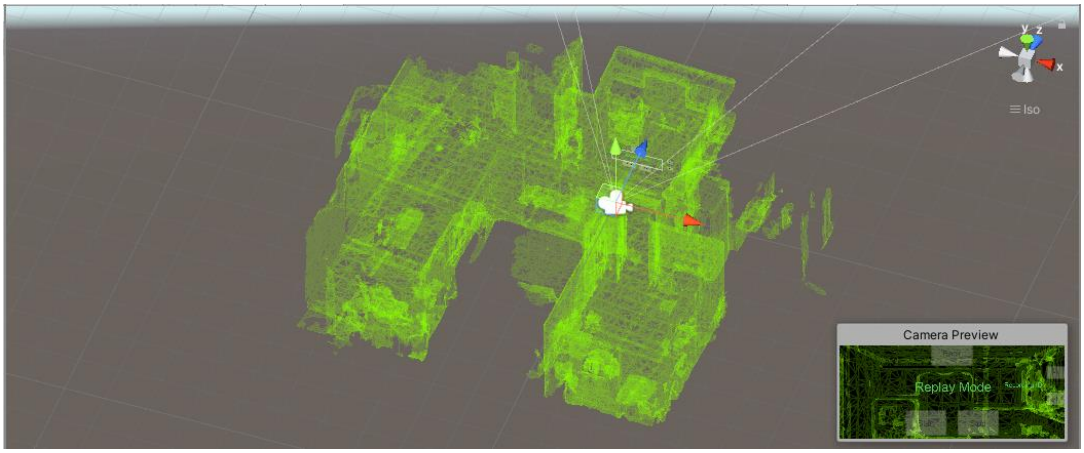

Figure 3.4: Apartment Spatial Map

Above is a spatial map of the apartment used for the task scenario recordings (see Figure 3.4). This screenshot was taken in the Scene view of the Unity editor. In the bottom-right corner, a preview of the in-situ user perspective of HoloVision's desktop Replay Mode is shown.

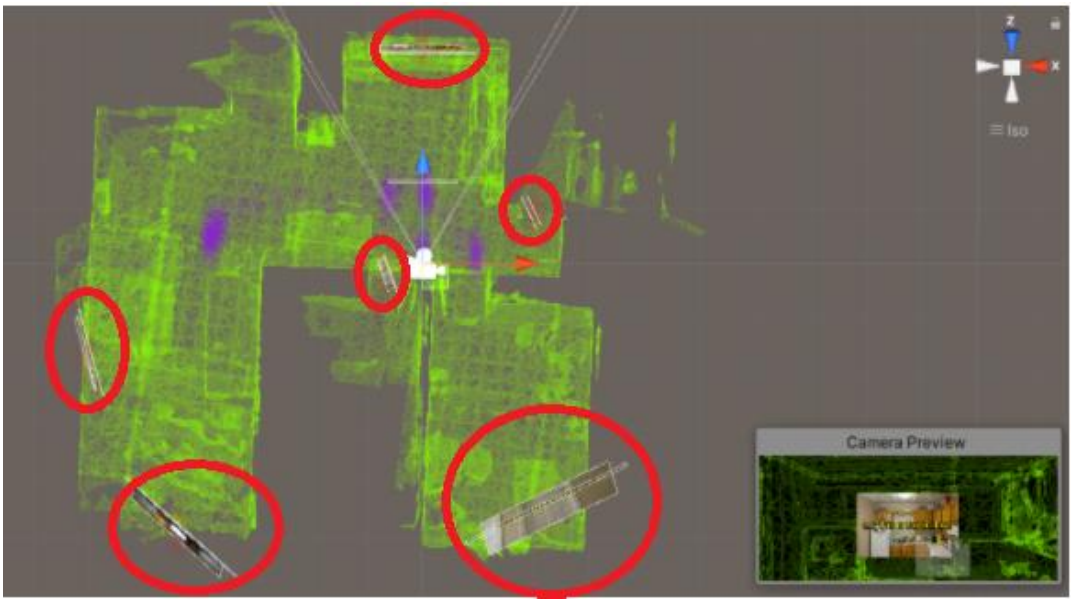

Figure 3.5: Annotated Apartment Spatial Map

Above is a spatial map of the apartment used for the task scenario recordings (see Figure 3.5). In this image, annotations from the "Apartment Tour" task scenario are circled in red. The purple traces indicate where the recording user travelled within the space. As shown in the bottom-right corner, one of the annotations is currently visible in the preview of the user's perspective.

4 Task Demonstrations

The following is a collection of task scenarios that demonstrate the features and limitations of HoloVision. In each scenario, we demonstrate how to complete a task during a Record Mode session. Next, we explore the recording in Replay Mode to demonstrate how the system placed the annotations. Together, these demonstrations serve to provide a few tangible examples of how HoloVision can be used to quickly produce consume rich information about different tasks within a workspace and be used later in time by a collaborator to receive assistance in completing the task.

4.1 Scenario 1: Making Cereal

In this scenario, an asynchronous, collocated learner will be shown how to acquire the necessary ingredients and tools to make a bowl of cereal in an unfamiliar kitchen. All instructions occur within a relatively small area. Instructions are fairly trivial but reduce the need for the user to search the drawers, cupboards, and the fridge for the locations of the necessary components (e.g., milk, bowl, spoon, cereal box, etc.), thereby reducing effort and the need to receive support from a collaborator. This task serves to demonstrate how MR can be used to improve common search-and-retrieve tasks without an expert being present within a workspace.

4.1.1 Record Mode Session: Cereal Making Example

Figure 4.1: Scenario 1, HoloVision Menu (Record Session)

1. The user stands in the starting position and opens HoloVision (see Figure 4.1).

Figure 4.2: Scenario 1, Record Mode Menu

2. The user selects Record Mode, sets Recording ID to 1 and selects the Start button (see Figure 4.2).

Figure 4.3: Scenario 1, Bowls

3. The user gazes at and approaches the cupboard, and says, “the bowls are in the cupboard” as he opens the cupboard, grabs a bowl, then places it on the counter (see Figure 4.3).

Figure 4.4: Scenario 1, Spoons

4. The user gazes downward at the drawer and opens it while saying, “the spoons are in the drawer”, then retrieves a spoon and places it on the counter (see Figure 4.4).

Figure 4.5: Scenario 1, Cereal

5. The user crouches down and says, “the cereal is in here”, as he reaches into a cupboard to grab the cereal and places it on the counter (see Figure 4.5).

Figure 4.6: Scenario 1, Milk

6. The user stands up, turns, approaches the fridge, and says, “the milk is in the fridge”, then reaches in to retrieve the milk, then places it on the counter (see Figure 4.6).
7. The user selects the Stop button to end the data capture.

4.1.2 Replay Mode Session: Cereal Making Example

Figure 4.7: Scenario 1, HoloVision Menu (Replay Session)

1. The user stands in the starting position and opens HoloVision (see Figure 4.7).

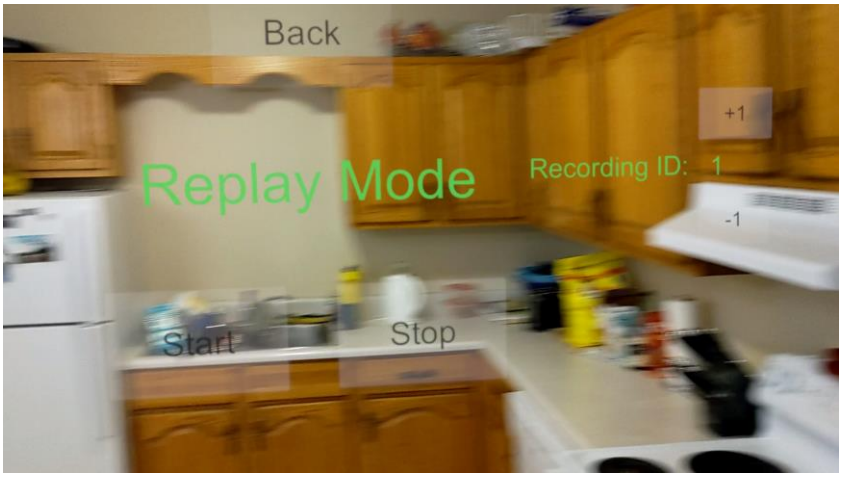

Figure 4.8: Scenario 1, Replay Mode

2. The user selects Replay Mode, selects the cereal making recording and selects the Start button (see Figure 4.8).

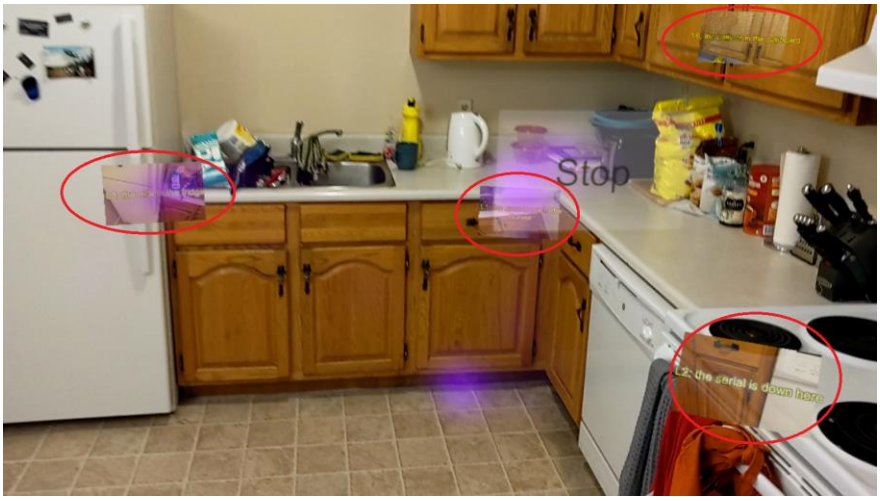

Figure 4.9: Scenario 1, Kitchen Annotations

3. Four text annotations appear in the kitchen, accompanied by video thumbnail images, and transparent purple clouds (position traces). Red circles have been added to the image to highlight the annotations (see Figure 4.9).

Figure 4.10: Scenario 1, Spoons Annotation

4. The user approaches an annotation L1. The speech-to-text for this recording shows that the previous user said, “the spoons are in the drawer”. The annotation has been placed on the drawer, and the thumbnail shows what the previous user saw at the time of recording (see Figure 4.10).

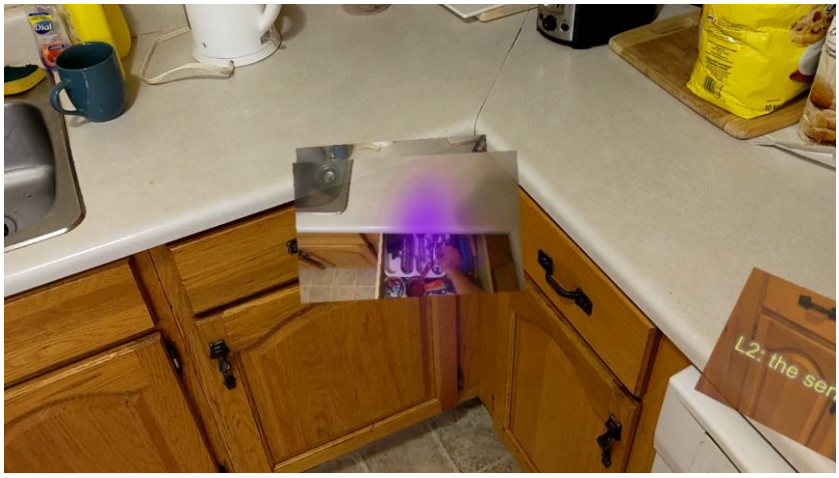

Figure 4.11: Scenario 1, Spoons Video

5. The user selects annotation L1, which triggers a video clip that shows what the previous user saw and what was said at the time of recording. As shown in the video, the user reached into the drawer for a spoon, and said, “the spoons are in the drawer”. Note the position trace slightly obscures the video clip; these traces can be disabled or enabled as needed by the user. The user then selects Stop to exit the current replay session (see Figure 4.11).

4.2 Scenario 2: Apartment Tour

In this scenario, a remote user will be shown the layout of an apartment. The task in this scenario is simply to inform the remote user about the apartment space using the

desktop version of HoloVision. The annotated spatial map emphasizes relative room locations and sizes, to teach the user what it might be like to be within the space. Similar to video games or 3D editing software, the space can be navigated by moving the camera using the mouse and keyboard. This demonstration serves to illustrate how MR could be used to familiarize remote users about a physical space before they are able to access it. This sort of task could be applied to navigation tasks, whether remote or collocated, where a user does not have access to someone familiar with that space.

4.2.1 Record Mode Session: Apartment Tour

1. The user stands in the starting position and opens HoloVision.
2. The user selects Record Mode, selects the apartment layout recording and selects the Start button.

Figure 4.12: Scenario 2, Kitchen

3. The user gazes at a central location on the other side of the kitchen and says, “this is the kitchen” (see Figure 4.12).

Figure 4.13: Scenario 2, Bathroom

4. The user turns and gazes at the bathroom door and says, “here’s the bathroom” (see Figure 4.13).

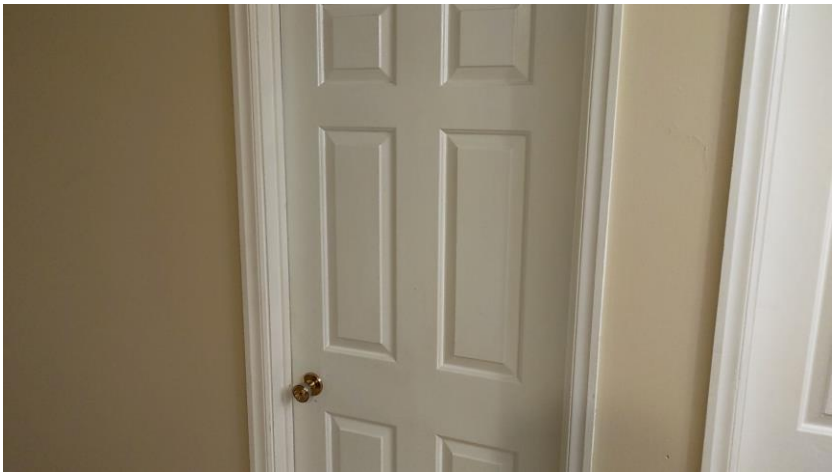

Figure 4.14: Scenario 2, Office

5. The user opens the office door and looks inside while saying, “this is the office” (see Figure 4.14).

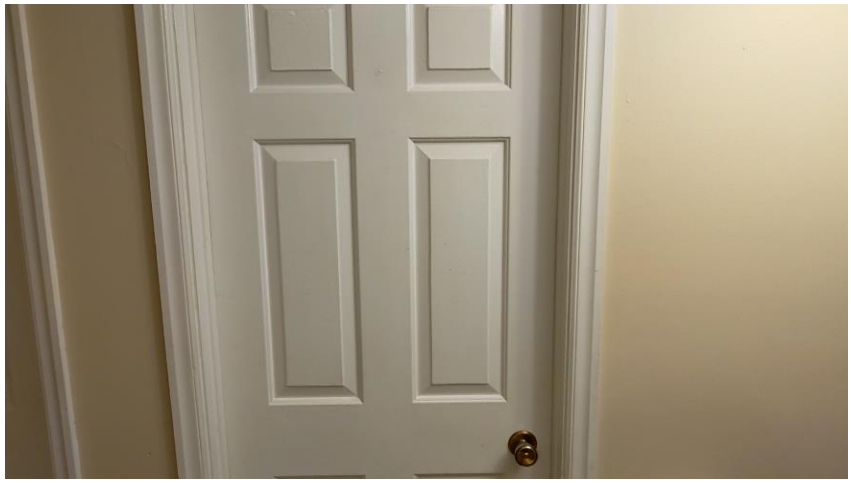

Figure 4.15: Scenario 2, Bedroom

6. The user gazes at the bedroom door and says, “this is the bedroom” (see Figure 4.15).

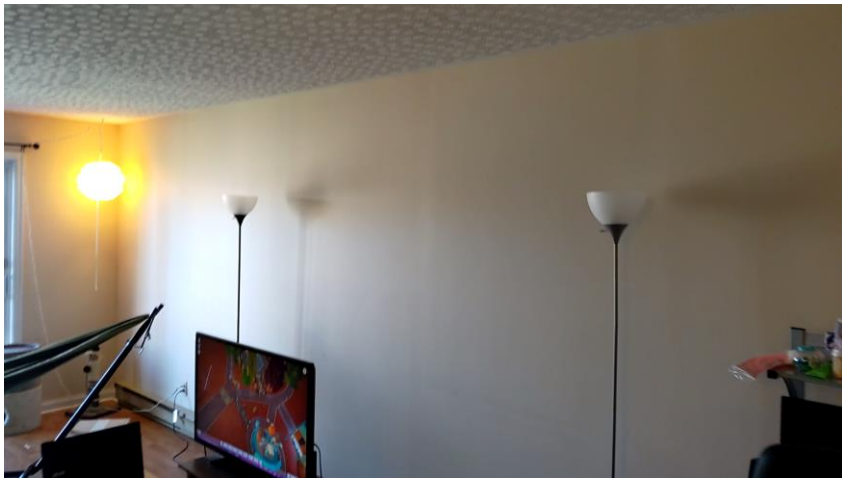

Figure 4.16: Scenario 2, Living Room

7. The user gazes into the living room and says, “here’s the living room” (see Figure 4.16).

Figure 4.17: Scenario 2, Balcony

8. The user enters the living room, gazes toward the balcony door, and says, “here’s the balcony” (see Figure 4.17).
9. The user selects the Stop button to end the data capture.

4.2.2 Replay Mode Session: Apartment Tour

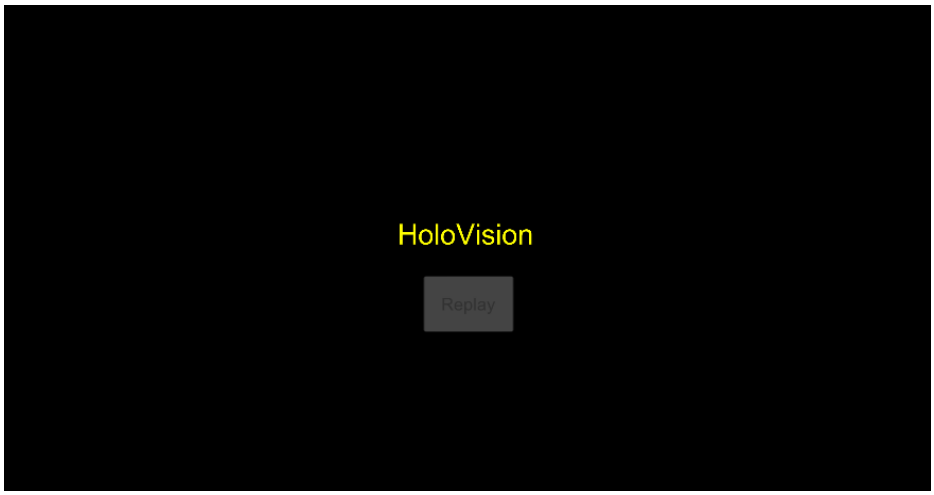

Figure 4.18: Scenario 2, HoloVision Menu (Desktop)

1. After importing the appropriate session data, the user opens HoloVision (PC version). The session data includes one .json file (containing position, gaze, speech-to-text, etc.), one .mp4 file (containing the full video of the recording session), and one .obj file (the spatial map object) (see Figure 4.18).

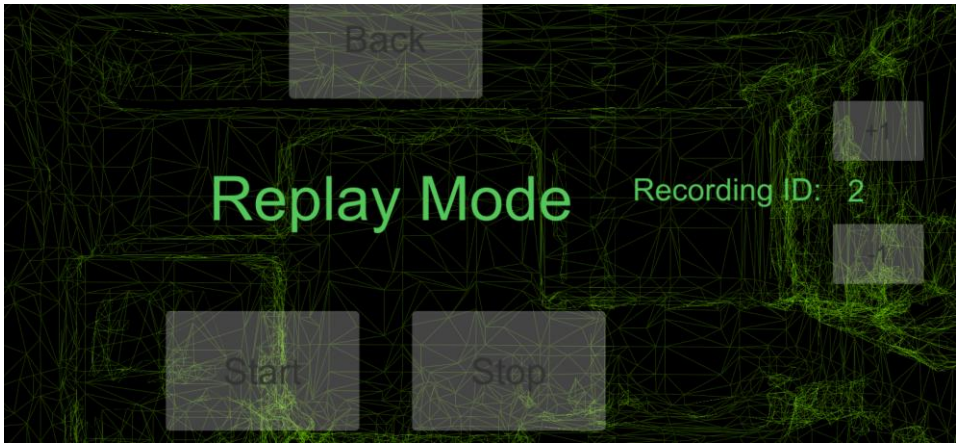

Figure 4.19: Scenario 2, Replay Mode (Desktop)

2. The user selects Replay and can now observe and traverse the virtual version of the apartment space using mouse and keyboard controls. The user then selects recording ID 2 and selects Start (see Figure 4.19).

Figure 4.20: Scenario 2, Kitchen Annotation

3. Immediately, the user can see a large annotation of the kitchen (see Figure 4.20).

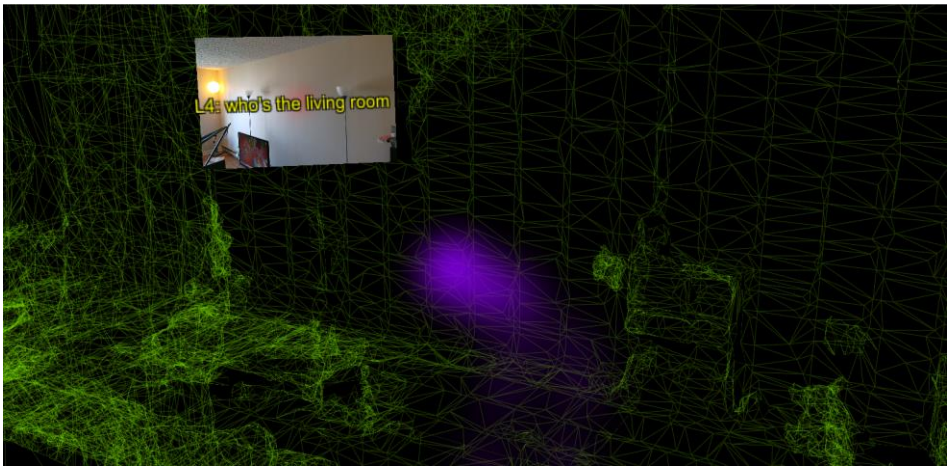

Figure 4.21: Scenario 2, Living Room Annotation

4. To the left, the user follows a purple trace to another room, which contains two more annotations. annotation L4 shows the living room (see Figure 4.21).

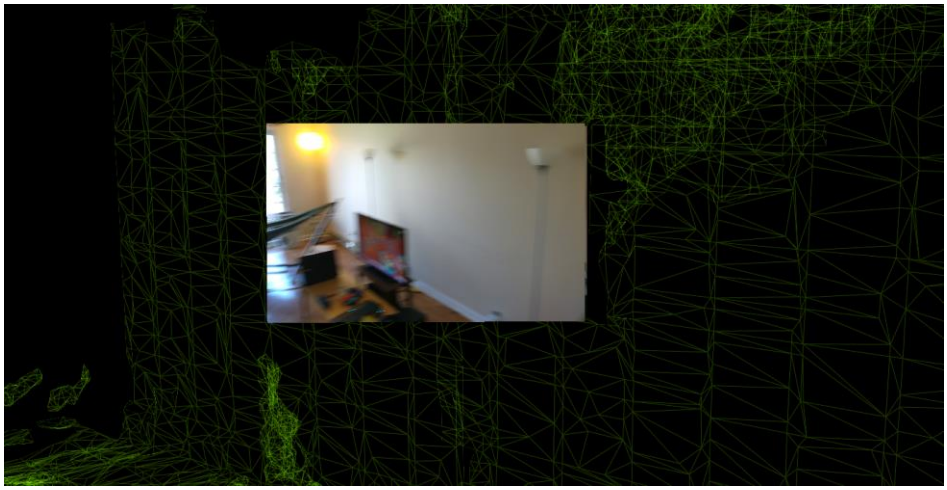

Figure 4.22: Scenario 2, Living Room Video

5. The user selects annotation L4 and watches the video clip. In the recording, the previous user says, “here’s the living room”, clarifying an error in the speech-to-text recording (see Figure 4.22).

Figure 4.23: Scenario 2, Balcony Annotation

6. At the far end of the room, the user can see annotation L5, which shows the balcony. Note that it is larger due to its distance at the time of recording; suggesting the appropriate interaction distance for this annotation. In contrast, smaller annotations encourage the replay user to approach them to interact (see Figure 4.23).

Figure 4.24: Scenario 2, Office Annotation

7. The user returns to the kitchen, noting several annotations. L1, L2, and L3 show the bathroom, office, and bedroom, respectively. The user then selects Stop to exit the replay session (see Figure 4.24).

4.3 Scenario 3: Plant Watering Instructions

In this scenario, an asynchronous, collocated learner will be tasked with watering the plants in an apartment. Instead of a written list of instructions, or an in-person walk-though, the expert has left a HoloVision recording that shows the locations of each plant in the apartment. For each plant, the expert has described specific watering instructions, which is both denoted by text annotations and fully expressed within video clips. When

the user arrives without the expert present, they will have access to the MR recording to consult as needed. This task serves as a more complicated version of the cereal task. The plants are distributed throughout multiple rooms, and the tasks are denser than a simple search-and-retrieve; certain plants may require more or less water than others. This demonstration serves to illustrate how HoloVision can be used to easily leave higher-level navigation guidance cues and information-rich contextual documentation for future users. Recordings like this could be shared between loosely-coupled collaborators to describe instructions or changes made to physical components within a workspace.

4.3.1 Record Mode Session: Plant Watering Instructions

1. The user stands in the starting position and opens HoloVision.
2. The user selects Record Mode, sets Recording ID to 3 and selects the Start button.

Figure 4.25: Scenario 3, Window Plants

3. The user enters the office, approaches the plants on the windowsill, and says, “there's a couple plants here though need to be watered everyday” (see Figure 4.25).

Figure 4.26: Scenario 3, Office Plant

4. The user gazes down at the other plant in the room and says, “this one here only needs water about once a week” (see Figure 4.26).

Figure 4.27: Scenario 3, Watering Can

5. The user enters the kitchen, gazes at the watering can in the corner and says, “we keep the watering can here” (see Figure 4.27).

Figure 4.28: Scenario 3, Living Room Plants

6. The user enters the living room, approaches the plants on the shelf, and says, “these plants also need to be watered pretty much everyday” (see Figure 4.28).

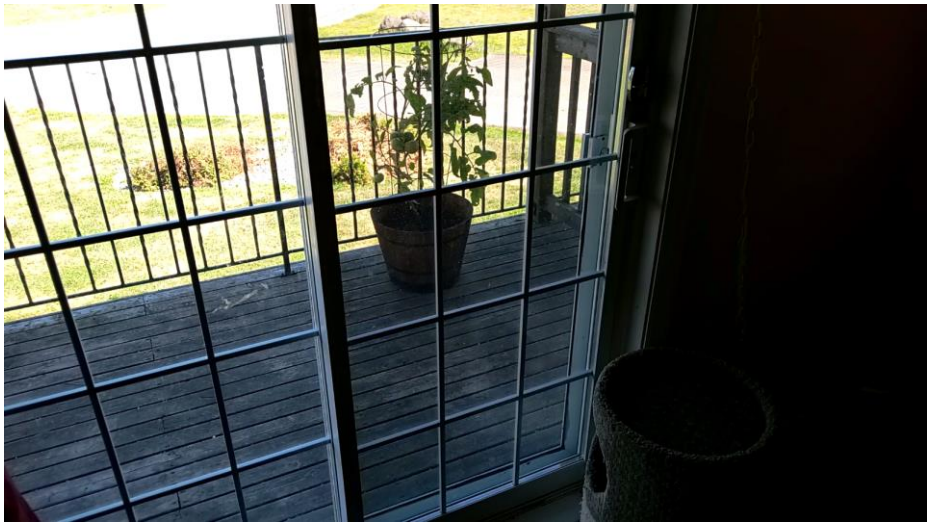

Figure 4.29: Scenario 3, Tomato Plant

7. The user gazes toward the balcony and says “and the tomato plant requires a lot of water everyday” (see Figure 4.29).
8. The user selects the Stop button.

4.3.2 Replay Mode Session: Plant Watering Instructions

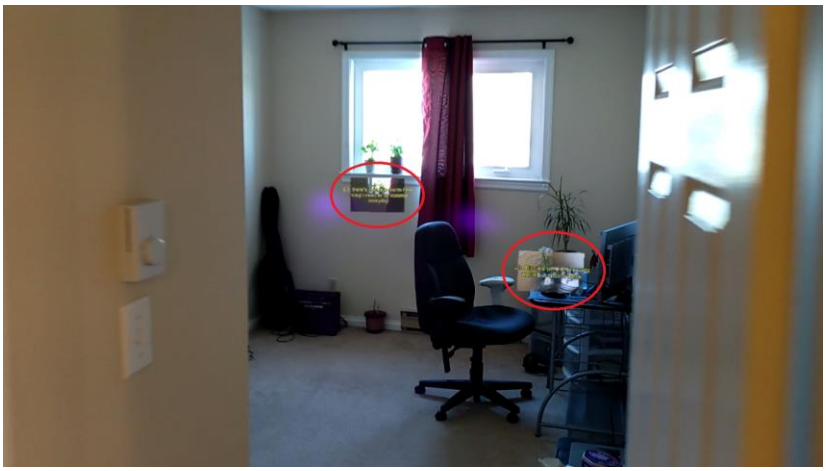

Figure 4.30: Scenario 3, All Office Plant Annotations

1. The user starts HoloVision, selects Replay, then selects recording ID 3 and selects Start. The user looks around and finds that some annotations seem to be behind a closed door (the annotations are not occluded). The user enters the room and approaches the small annotations in the office. Annotations in the image have been circled in red (see Figure 4.30).

Figure 4.31: Scenario 3, Window Plant Annotations

2. The user approaches annotation L0 which says, “there's a couple plants here though need to be watered everyday” (see Figure 4.31).

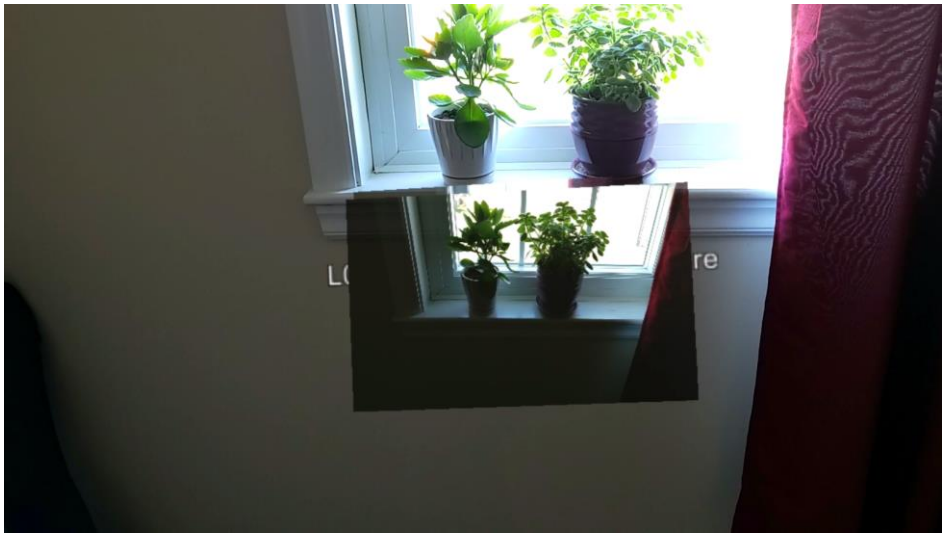

Figure 4.32: Scenario 3, Window Plants Video

3. The user selects annotation L0. During the clip, the previous user says, “there's a couple plants here that need to be watered everyday” (see Figure 4.32).

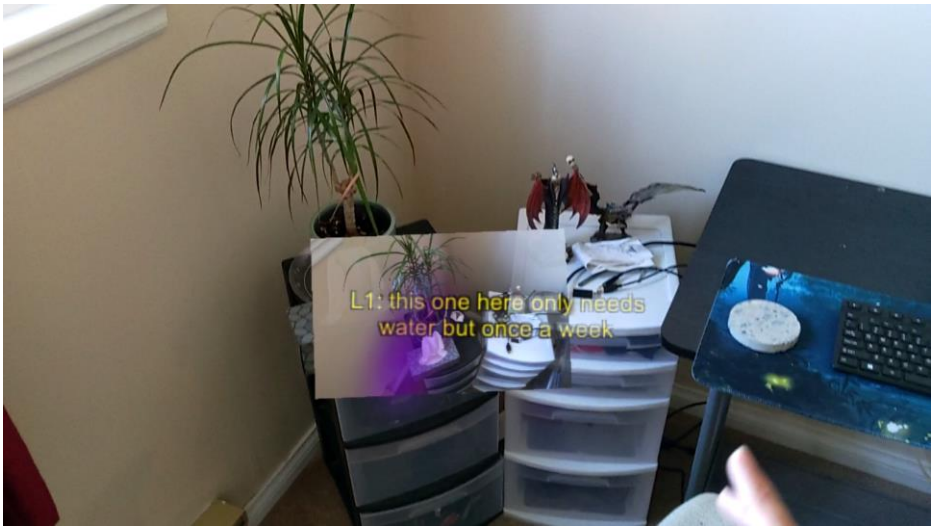

Figure 4.33: Scenario 3, Office Plant Annotation

4. The user gazes down at annotation L1 which says, “this one here only needs water but once a week” (see Figure 4.33).

Figure 4.34: Scenario 3, Living Room Plants Annotation

5. The user leaves the office and enters the living room, taking note of the other annotations as they approach annotation L3. The annotation reads, “these plants also need to be watered a pretty much everyday”. The user selects Stop to exit the replay (see Figure 4.34).

4.4 Scenario 4: Video Game System Setup

In this scenario, the expert has recorded instructions for the proper setup of a video game console. The expert shows the necessary components and how to use them to complete the setup task. This task serves to demonstrate how MR can be used for finer, desk-level assembly tasks. Since instructions are all performed at the desk, the

visualizations are essentially grouped as a list of video clips that illustrate the sequential steps to take during setup. Overall, this scenario does not leverage the spatial distribution of annotations, but it demonstrates that the information density of the annotations allows for more than just object-searching or room-to-room navigation.

4.4.1 Record Mode Session: Video Game System Setup

1. The user stands in the starting position and opens HoloVision.
2. The user selects Record Mode, sets Recording ID to 4 and selects the Start button.

Figure 4.35: Scenario 4, Setup Tutorial

3. The user sits at the table, gazes at the video game console components, and says, “So first you’ll need to collect all these things, here’s the playstation power cable,

the playstation, the HDMI cable, the TV, and the controller, we also have a power bar over here" (see Figure 4.35).

Figure 4.36: Scenario 4, HDMI to TV

4. The user gazes at the HDMI cable and says, "First thing you'll need to do is connect the HDMI cable to the back of the TV like this." The user plugs the HDMI cable into the back of the TV (see Figure 4.36).

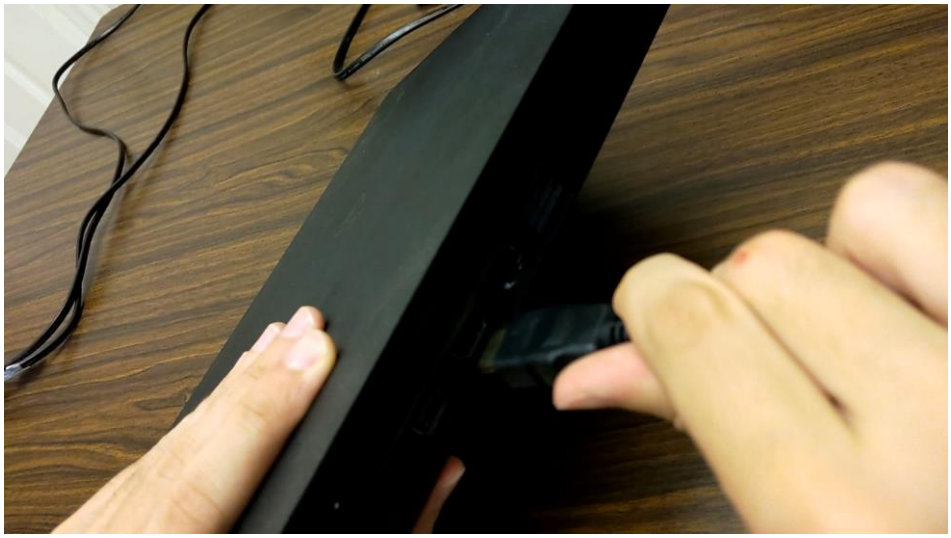

Figure 4.37: Scenario 4, HDMI to Console

5. The user gazes at the HDMI cable and says, “Then you’ll need to connect the HDMI cable to the back of the Playstation here.” The user connects the HDMI cable to the back of the console (see Figure 4.37).

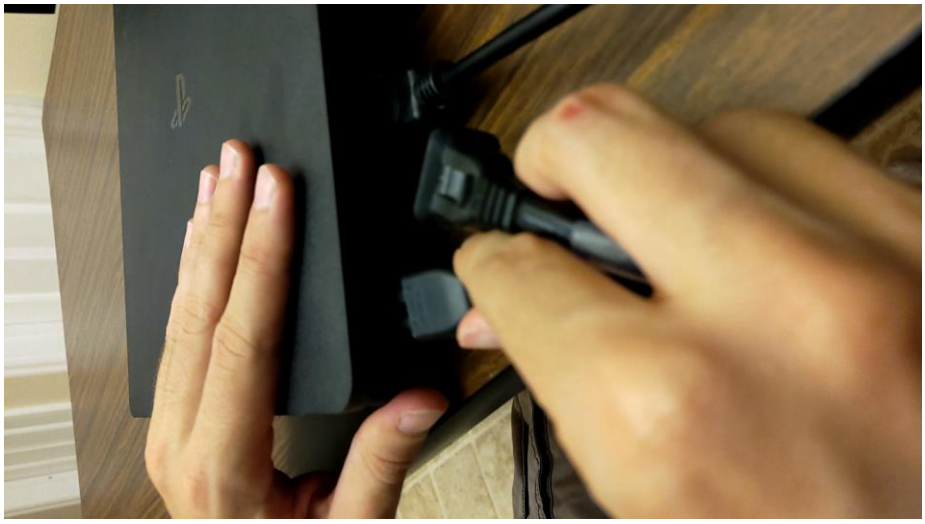

Figure 4.38: Scenario 4, Power to Console

6. The user gazes at the power cable and says, “Next you’ll need to connect the power cable to the back of the PlayStation right here”. The user connects the cable as described (see Figure 4.38).

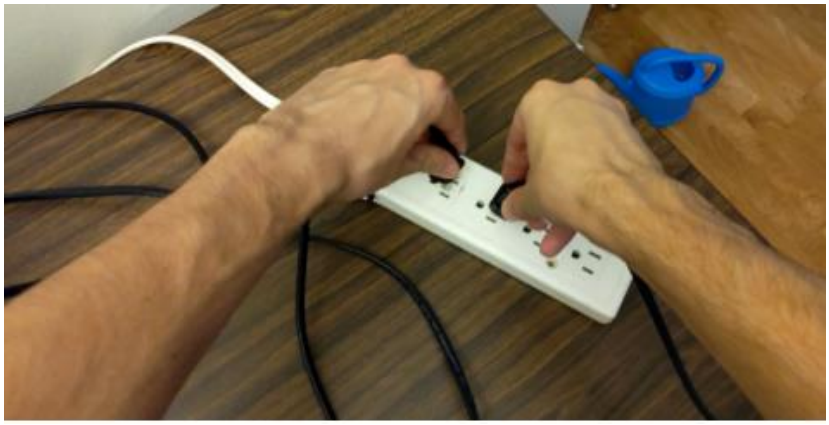

Figure 4.39: Scenario 4, Power Bar

7. The user gazes at the power cables and says, “And, finally you’ll need to connect the TV and the PlayStation power cables to the power bar here”. The user connects the power cables to the power bar (see Figure 4.39).

Figure 4.40: Scenario 4, Controller

8. The user gazes at the controller and says, “Finally then you can use the controller, press the button to turn on the system”. The user turns the system on by pressing a button on the controller (see Figure 4.40).
9. The user selects the Stop button in HoloVision to end the recording.

4.4.2 Replay Mode Session: Video Game System Setup

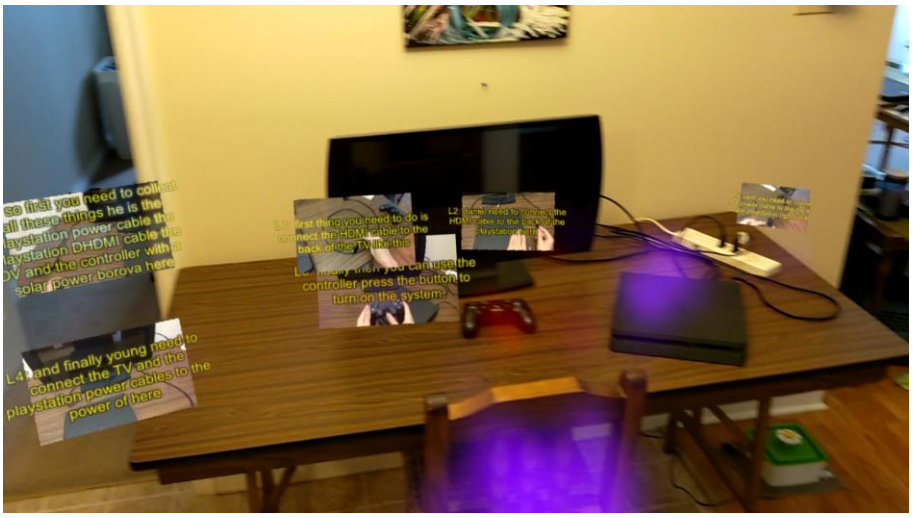

Figure 4.41: Scenario 4, Setup Tutorial Annotations

The user starts HoloVision, selects Replay, then selects the PlayStation setup recording and selects Start. The user approaches the console setup workstation and observes a group of small annotations positioned there, suggesting the previous user was close when interactions took place. A dense purple cloud near the chair suggests the previous user lingered there for a while. In this example, the recording user has assumed that the task objects will remain where they were when the recording took place. In a real-world scenario, objects may not always be found in the same locations. HoloVision does not address this limitation, but the design of similar systems in the future could integrate

a way to recognize objects via computer vision or wireless signals (like Bluetooth, RFID, or NFC). Such a system would require additional costs to set up; computer vision solutions would require training, or wireless beacons would need to be installed in each object. This would greatly increase the system complexity for developers and users alike, which is beyond the scope our system requirement, R3. Transparent red clouds (gaze markers) indicate points of interest at the workstation. These gaze markers are typically obscured by annotations but are revealed when annotations are hidden or relocated due to overlap. The rearrangement of overlapping annotations uses a relatively naïve algorithm to display annotations in a wide rectangular grid. While the annotation order is denoted in the labels, future designs could benefit from an additional visual indicator, like arrows or lines emphasize the ordering of sequential tasks (see Figure 4.41).

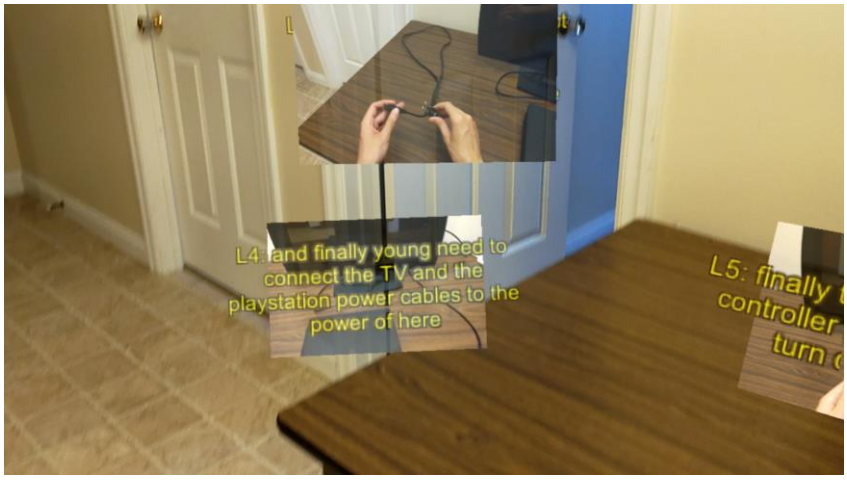

Figure 4.42: Scenario 4, Components Overview Video

1. The user sits in the chair and selects Annotation L0. The video clip describes each of the components at the station (see Figure 4.42).

Figure 4.43: Scenario 4, Power Bar Annotation

2. The user reads Annotation L3 and sees that the power cables need to be connected to the power bar. The purple cloud indicates that the previous user leaned over to access this area of the workspace. At the workstation, these components have already been plugged in (see Figure 4.43).

Figure 4.44: Scenario 4, Controller Annotation

3. The user reads Annotation L5, which shows that a button on the controller may be pressed to turn on the system. Since the system already appears to be set up, the user skips to this final step (see Figure 4.44).

4.5 Summary

To summarize, each of these example scenarios have demonstrated how HoloVision can be used to record and replay MR sessions to assist someone unfamiliar with a physical environment and/or task with how it can be completed. Due to access restrictions to the University research lab, the chosen tasks were common tasks that could be performed within an apartment, instead of an actual workspace. Although these tasks

do not necessarily reflect those that could be performed on a job site, the examples demonstrate how Holovision can capture and display representative interactions (like room-to-room navigation) object search-and-retrieval, verbal instructions, and desk-level activities. I also demonstrate how HoloVision can be used to explore the MR session in-situ or remotely (using the HoloLens PC application). Ultimately, this chapter demonstrates the system's features and limitations, and serves to accompany the system design section to provide a complete overview of the current system.

5 Planned Study

To receive user feedback about the system and inform future work, a study was planned. However, due to the COVID-19 pandemic, proceeding with the study was not feasible. The study was initially planned for a lab setting, then re-planned for an apartment setting (the home of the primary researcher). Ultimately, the study was not able to be conducted, but the author was still able to personally perform the planned experiment's task sets for demonstration purposes. The lack of a formal, empirical study is a significant limitation of this work.

5.1 Study Design

This iteration of the study design was planned to be conducted in the primary researcher's apartment in Fredericton, NB. Before the study, each participant was to be asked to fill out a brief "Pre-Experiment" questionnaire asking about gender, age, and their experience with MR and task guidance. Each participant would be told that they were to complete a set of three tasks using HoloVision's Replay Mode (collocated Replay) as guidance. Half of the participants were to use HoloVision on the HoloLens 2 HMD to complete their tasks in-situ. The other half would use HoloVision on a desktop PC (Remote Replay) to learn about their tasks and the workspace before completing their tasks. After completing the set of tasks, each participant would be asked to complete a

"Post Condition" questionnaire (based on NASA-TLX) about their experience with the guidance tool. Next, participants were to be asked to create a set of three tasks using HoloVision's Record Mode on the HoloLens 2 HMD. Upon completion, the participants would then be asked to complete another Post Condition questionnaire, followed by a "Post Experiment" questionnaire that asked about their overall experience during the experiment.

To avoid learning about the space beforehand, the "Remote Replay" participants would be asked to close their eyes upon entering the experiment space and would be brought to a separate room before starting.

5.2 Study Part 1: Replay Mode

In preparation for this part of the study, the primary researcher used HoloVision's Record Mode to record a set of three tasks within the apartment:

- *Task 1: Search in an Unfamiliar Space.* Participants will be asked to locate and retrieve a pair of scissors. The scissors were completely hidden from view underneath a tissue box on an end table next to the couch. This task was chosen as a representative task for hidden object search and retrieval.
- *Task 2: Simple Task.* Type "5728" into a calculator and press the "=" button. The calculator was placed in view amongst other objects on a coffee table. This task

was chosen as a representative “simple” task that could be accomplished based solely on text instruction.

- *Task 2: Complex Task.* Draw a picture of a happy face using markers of specific colors. The markers and paper were placed in view on the kitchen table. This task was chosen as a representative “complex” task that would require image or video analysis, in addition to the text instruction.

After creating the recording, it was verified that the annotation locations and information appeared as intended in Replay Mode. This is the recording that was used to be used by both the collocated and remote users in the study.

5.3 Collocated Replay Participants

Collocated users were to be asked to put on the HoloLens 2 and were shown how to use the air-tap gesture. The participants would have a few minutes to watch a pre-recorded tutorial video and familiarize themselves with the device, the software, and to calibrate the display if necessary.

Next, the user would be asked to launch the application and navigate to Replay Mode. From there, they would be asked to select recording ID -1 (the ID I used for the task recording) and confirm when they had done so. Once the appropriate ID was selected, they were to be informed that they would be performing a set of tasks as specified in the

virtual annotations that would appear once they select the start button. The participants were to be told that they were to locate and complete these tasks as fast as they could do safely while avoiding errors. They would be informed that annotations may be air-tapped to view a brief video clip if further information was required. If the participant had no further questions, they would be told that they may begin when ready. During this session, the primary researcher would take note of total session duration and specific task errors that were made.

5.4 Remote Replay Participants

Remote participants were to be brought (with their eyes covered) to a room that was not included within the task workspace. From here, participants would be seated at a desktop computer that had been installed with HoloVision. Participants were to be told that they were going to be performing a set of tasks within an unknown workspace, and that they may use HoloVision to virtually navigate that space using the recorded session in Replay Mode. The primary researcher would explain the mouse and keyboard controls for 3D camera movement and would instruct the participants that they must locate their tasks within the virtual space, and that they may click annotations for video containing more information. Participants were to be provided a pencil and paper to take notes to help them recall their tasks later on. Once the participants were ready, they were to be brought

to the start position and asked to complete their tasks. Like the collocated participants, total session duration and specific task errors would be noted.

5.5 Study Part 2: Record Mode

After completing the Replay Mode tasks and the associated questionnaire, all participants were to be asked to stand in the start position, put on the HoloLens 2, launch HoloVision, and navigate to Record Mode. For remote replay participants, this would be their first time using the HoloLens, so like before, they would be given some time to familiarize themselves with the device. Since task recording performance does not rely on active air-tap interactions besides starting and stopping the recording, the learning effect should not have significantly affected results. For this task, participants would be presented with three small pieces of paper labelled, “A”, “B”, and “C”. Participants were to be asked to deliver these “tickets” to three different within the task workspace. They would be given time to decide where to put them and would be told that they must speak only when placing the tickets once the recording had begun. The ideal recording conditions for a Record Mode session would be summarized to help them create successful recordings (avoid dark areas, only speak when it is relevant, ensure a minimum of two second gaps between utterances, etc.). When ready, the participant would be asked to select the start button, perform their recording, then select the stop button when complete.

5.6 Results

This section would typically report on results, observations, and feedback received by the participants. Since the study was not performed, I will offer some of my personal experience performing the experiment tasks using the system as the primary researcher in a mock study. This insight could be used to inform further studies with HoloVision, or other systems of similar design. I will also provide some images to

5.6.1 Part 1 (Replay Mode)

For Part 1 of the study, participants were to be asked to perform a set of tasks with guidance (either remote or collocated) from a pre-recorded session in HoloVision. In this section, I provide screenshots, mock session data, and observations from my personal "mock" completion of these tasks.

5.6.1.1 Collocated Replay

For collocated replay, participants were to be told that the task annotations could be found within the apartment space. From the start position, all three tasks were relatively easy to locate. If not immediately visible, I found that turning around and walking shifted my perspective enough to emphasize the annotations in the space, as they were not occluded by walls or other physical barriers. All tasks were initially located beyond the FOV of the HoloLens' display. I felt that the tasks could be located within just a few

seconds of visual search. Once they had entered my FOV, the annotations (floating images with yellow text) became obvious. I found that the purple traces occluded a task annotation, which made it difficult to read, but it sometimes seemed almost too transparent to be useful as a guide. This was likely because the position data was denser in areas where recording user was stationary, and sparse in areas where the user was moving. I speculate that the annotations were relatively easy to locate due to the limited size of the search area. Within the small space, even small annotations like the ones generated for the study were easy to spot from the start position. In larger workspaces, or in environments with extra virtual or physical clutter, it could be more difficult to find relevant annotations. Because I found that the purple traces were not particularly helpful, I suggest that future work on similar projects should support annotation better macro-scale guidance features that indicate where to look and walk to find the annotations.

I think that participants would have struggled with the air-tap gesture. Despite my experience with the HoloLens, I sometimes had difficulty centering my focus on buttons or annotations to select them, especially when trying to be quick with task completion. Selectable objects do change color on hover, but HoloVision did not have a visual cursor at the time I performed the mock study. This issue was later amended by implementing a visual cursor.

During the mock study, I realized there was an error in the wording of task 2, which specified a non-existent “enter” button on the calculator. The referenced button (the

"=" button) was faded from use, and slightly out of frame in the video clip. The ambiguity of this task was unintentional but serves as a good example of how miscommunication can manifest in a workspace, even with rich, in-situ annotations. The annotations were only as good as the information I provided. While some participants may have been able to interpret the solution, there was still room for error in the recording itself. For future recordings with HoloVision (or similar systems), it will be important to review recordings to ensure their accuracy by editing the json files as necessary.

During the mock study, I found that it would have been convenient to interact with the annotations while my hands were occupied by task-related objects, like when I was drawing the picture. Although it was not stable enough for use during the mock study, voice activation for video clips is a working feature of HoloVision; it requires the user to say "command", followed by the annotation label id (e.g. "L1"). I believe improving this aspect of the system would leverage one of the HMD's key benefits.

5.6.1.2 Remote Replay

For the other task set, participants were to sit at a desktop computer that had HoloVision installed and the appropriate recording session imported. Using mouse and keyboard controls, they would navigate the space to learn the where and how to perform the tasks. Since I was already aware of task locations and instructions, I simply used the system's features as a participant would for the sake of the mock study.

As someone who is experienced with navigating 3D spaces in video games, and as the creator of the system, I had little difficulty using desktop Replay Mode to find the annotations within the space. However, if participants were unfamiliar with typical mouse and keyboard controls for video games, they may have encountered some difficulty. The spatial map for the apartment space was imperfect, and would likely be more difficult to interpret for participants with no prior knowledge of the space. As a resident of the space, it still took some effort to interpret the 3D wireframe reconstruction of the space relative to my mental model of the actual room layout. Despite this hurdle, I believe that participants would still be capable of discerning between rooms using the spatial map.

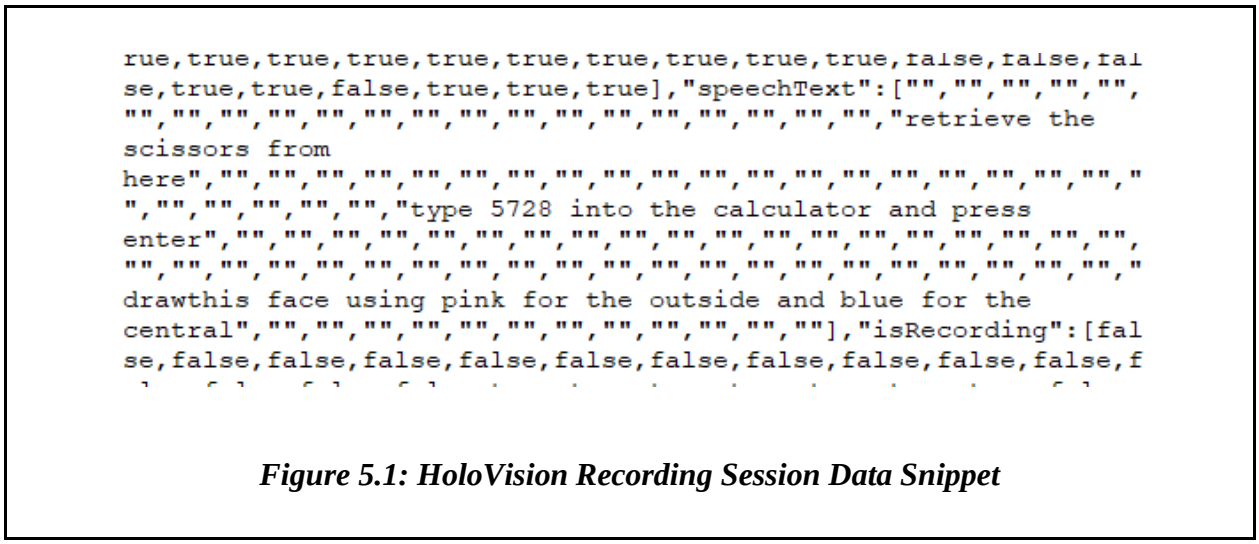

Figure 5.1: HoloVision Recording Session Data Snippet

Above (see Figure 5.1) is a code snippet from the .json file used to store the HoloVision session recording used for the pilot study. Note that the text in the "speechText" portion of this file could be replaced if minor changes to the instructions

were required, instead of recording an entirely new session. Note, the video clips will still contain the original utterance, so this action should be typically reserved for accuracy errors caused by the dictation recognizer, not the recording user.

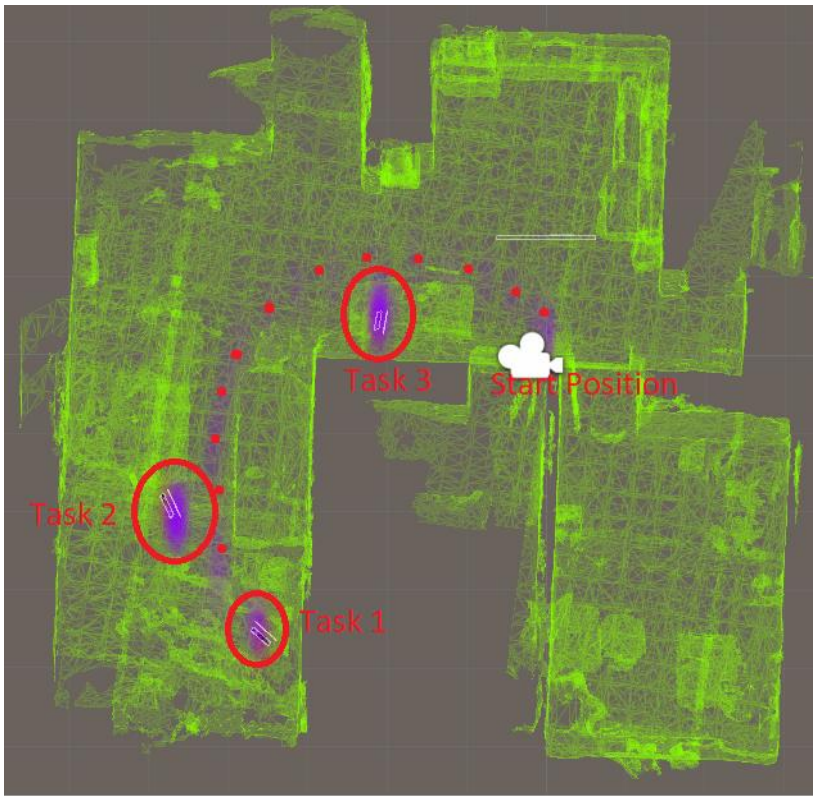

Figure 5.2: Pilot Study, Task Set

Above (see Figure 5.2) is a top-down screenshot of the visualizations generated by HoloVision's desktop Replay Mode. Shown are the annotations generated by the recording session containing the set of three tasks. The camera symbol indicates the start position, located in the kitchen. The three task annotations are circled and labelled. The transparent purple traces are a bit difficult to see with the grey background at this angle and distance, so they have been emphasized with red dots.

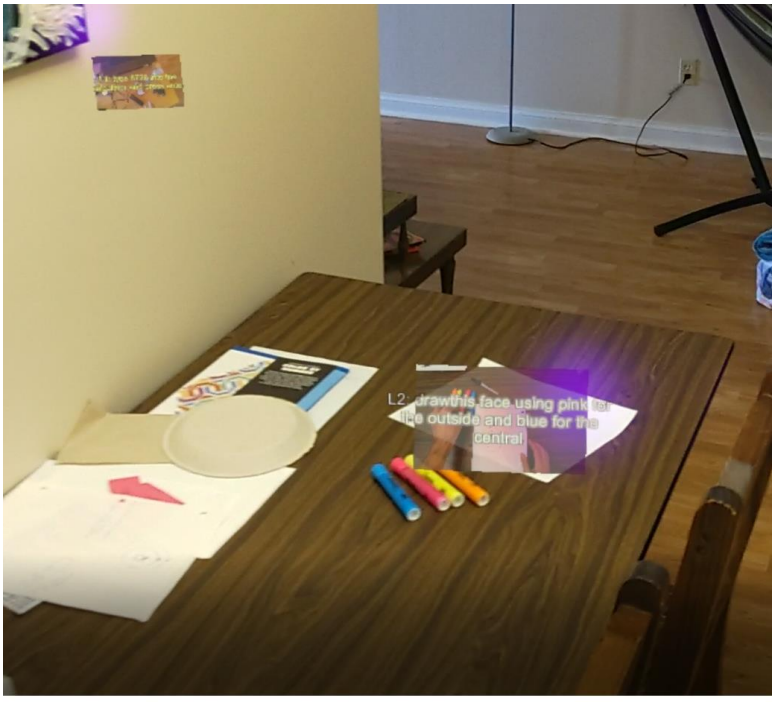

Figure 5.3: Pilot Study, Drawing Task Annotation

Above, (see Figure 5.3) is the annotation for task 3 (the drawing task) is shown on the table. In the background, the annotation for task 2 (the calculator task) can be seen through the wall.

Figure 5.4: Pilot Study, Calculator Task Annotation

Above (see Figure 5.4) is the annotation for task 2 (the calculator task) is shown on the coffee table.

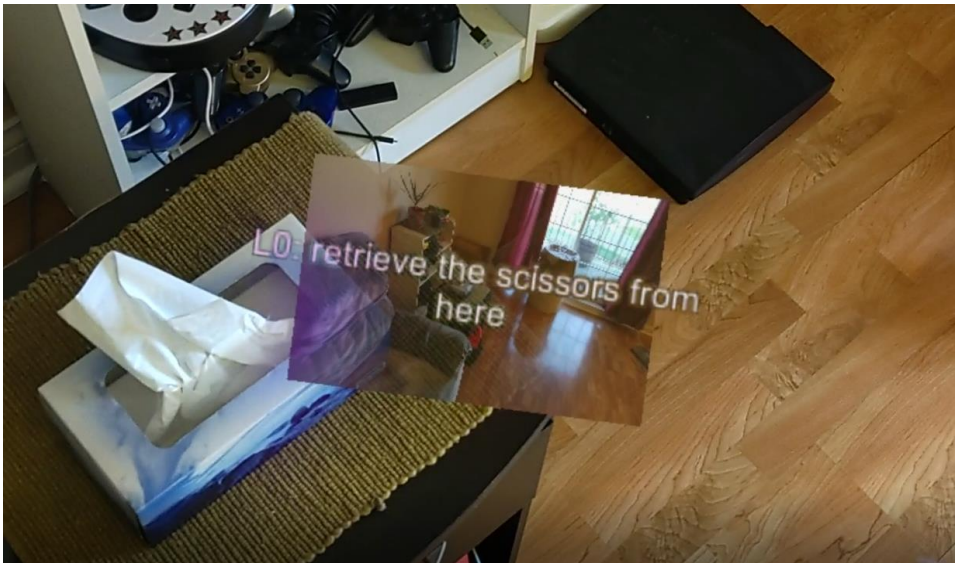

Figure 5.5: Pilot Study, Scissors Task Annotation

Above (see Figure 5.5) is the annotation for task 3 (the scissors task) is shown on the end table.

Figure 5.6: Pilot Study, Scissors Task Video

Above (see Figure 5.6) is the video for task 3 (the scissors task) is played, revealing that the scissors can be found beneath the tissue box.

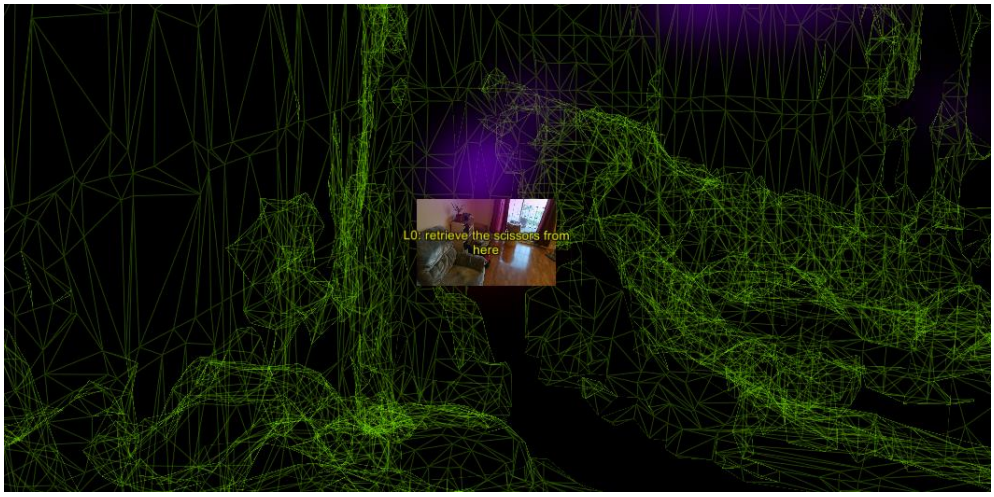

Figure 5.7: Pilot Study, Scissors Task Annotation (Desktop)

Above (see Figure 5.7) is the annotation for task 1 (the scissors task) is shown at the far end of the room.

Figure 5.8: Pilot Study, Calculator Task Video (Desktop)

Above (see Figure 5.8) the user plays the video clip for task 2 (the calculator task) showing the calculator on the coffee table.

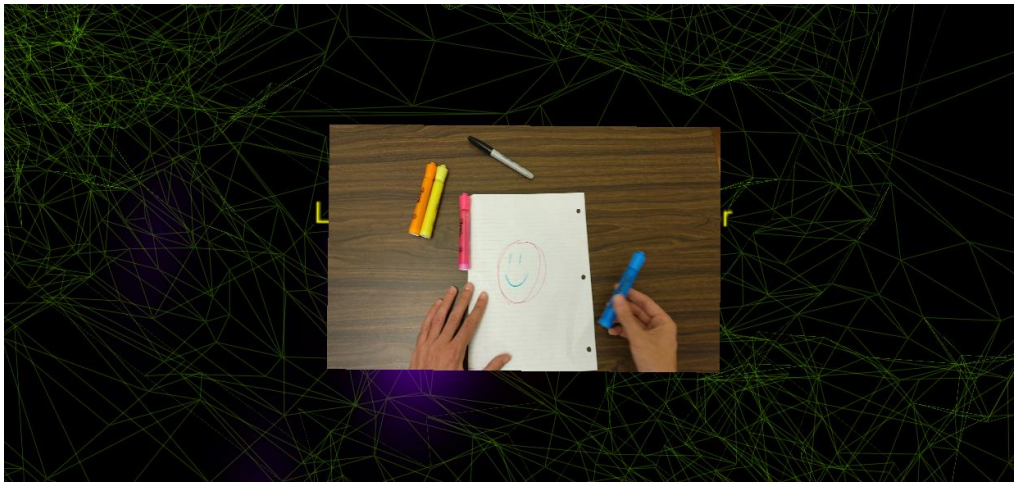

Figure 5.9: Pilot Study, Drawing Task Video

Above (see Figure 5.9) the user plays the video clip for task 3 (the drawing task) showing the complete drawing and the markers used.

5.6.2 Part 2 (Record Mode)

For Part 2, participants were to be asked to use HoloVision's Record Mode to record a set of three tasks. From the start position, the task would be explained task and participants could begin when ready. During my mock recordings, some of the attempted sessions were unsuccessful due to poor lighting and fast movement, which cause the HoloLens to lose spatial tracking and disrupt the recording. This resulted in incomplete session data; annotations were not generated, and it was not possible to verify the accuracy

of these sessions. Some inaccuracies also arose from speaking too quietly. Since recordings require a full-length video capture of the session, in addition to use of the voice recognition features, Record Mode is relatively unstable compared to Replay Mode. In my personal experience using the system leading up to the mock study, task demos and pilot study recordings sometimes required that I make several attempts to record a viable session when preparing annotations for the original task sets. It would not be surprising to observe that participants would also have some degree of difficulty generating a successful recording on their first attempt if they were to complete the study. Because I was moving relatively quickly for the mock study, I did not always hold my gaze on the task subject while speaking, so the resulting annotation appeared on a distant wall that the that I approached after my utterance. This suggests that the gaze linger tolerance could be improved further. Despite these recording errors, I believe a novice user would be able to create a viable recording within a relatively short period of time with minimal instruction, as the entire process of recording and verifying the annotations takes only a slightly longer than the length of the recording time.

We created Record mode to rapidly produce rich, in-situ annotations in a low-cost way. While recordings do sometimes fail, it is easy to notice when and how it happened. It is also easy to make the recording again if an error occurred. As noted above, there is also room for further improvement of the system to reduce error. Because of this, I believe

the system offers a viable method to author asynchronous MR annotations for collocated or remote environments.

5.7 Summary

The original goal of the pilot study was to receive user feedback to inform design decisions in future work with HoloVision and other similar systems. I believe that participants would have been able to achieve satisfying results using each of the modes in HoloVision. The errors encountered during the mock study emphasized a need for better selection techniques and more error tolerance for recordings. Ultimately, the mock study focused on performing a comprehensive test run of HoloVision's features. As discussed in the next section, a proper evaluation would include a group of participants and it would include control conditions (like videos, maps, notes, or human guidance) to compare the system to traditional or ideal guidance solutions. Ideally, a future evaluation would also take place in a more appropriate environment.

6 Discussion

6.1 Is HoloVision successful?

In Chapter 3, the goal was to establish how HoloVision can be used to rapidly produce and consume asynchronous MR annotations by providing a set of sample use cases. The use cases provided are relatively trivial tasks, but they demonstrate how the system meets our requirements (R1-R3; see page 27). For Record Mode, it is demonstrated how HoloVision is used to rapidly author annotations, simply by pressing the start button, performing the task with verbal commentary, then pressing stop. In each scenario, the recording followed that procedure. If annotation placement was inaccurate, it could be just as easily re-recorded with the same process, or if just the speech to text labels were not quite right, the plaintext in the json file could be quickly edited. For Replay Mode, it is demonstrated how the annotations are generated for and used by future users (i.e., loosely-coupled collaborators). Ultimately, we hoped to show that HoloVision's annotation production and consumption process was easy, and that the resulting annotations were valuable.

In Scenarios 1, 3, and 4 (of Chapter 3), we demonstrated different co-located use cases, whereas Scenario 2 represented how the annotations can also be used remotely with a virtual reconstruction of the workspace. Scenario 1 demonstrated same-room search and

retrieval of objects. The value to focus on with this example is not just the annotation content, but the location of those annotations.

In addition to demonstrating the remote usage of HoloVision, Scenario 2 focused on room-to-room navigation. In this scenario (a mock apartment tour) we envisioned using HoloVision as a tool that someone could use to learn about a physical space before they were able to enter it. One can imagine how this could be valuable for workers like firefighters or soldiers, where understanding the layout of a physical space could reduce navigation time and errors in a situation where every second matters. Similar to annotations produced using spatial and verbal indicators with SAGE (Moghadam et al., 2016), these locations leveraged speech, gaze, and spatial mapping data from the device to place an instruction, an image, and a relevant video clip, in each location where a task was described. Unlike SAGE, however, these annotations provide much more than just text labels.

Scenario 3 (plant watering) provides co-located instructions that are placed in various rooms. Similar to Scenario 1, the annotations simply show the user where to go and what to do. Unlike Scenario 1 objects in this scenario are not hidden, but they aren't all in the same room, and the instructions are a bit more complex than mixing cereal. This example was created to reinforce the value of the annotations in supporting room-to-room navigation and the explanation of slightly more detailed tasks. While the purple particle trails were not very easy to see, the annotations (unoccluded by walls) were still relatively

easy to locate. Where Loki (Thoravi Kumaravel et al., 2019) provides a complex and cohesive (albeit costly) task reconstruction, HoloVision provides a low-cost breakdown of the tasks. With both systems, the completion of the task must be demonstrated (or at least mocked) by the expert.

Finally, Scenario 4 demonstrated how more complicated, multi-step tasks could be distributed in a way that provides hands-free, labelled image and video documentation about a task at a workstation. We found that this scenario did not leverage spatial information in a valuable way. Furthermore, since the recording involved a relatively sequence of instructions that were not spatially or temporally spaced, we found that maintaining verbal pacing and subject gaze felt unnatural, making the recording process relatively difficult. When consuming these annotations, the sequential nature of that task and the spatial arrangement of the annotations made the experience feel a bit overwhelming; the area was cluttered with persistent annotations with similar thumbnail images.

Together, the scenarios described demonstrate the value and the success of the new approach, as exemplified through HoloVision, in supporting common collaborative scenarios.

6.2 When would HoloVision be most beneficial?

In Chapter 4 the goal was to demonstrate how to evaluate the effectiveness of HoloVision as a task guidance tool. Acknowledging what we had learned from the scenarios in Chapter 3, the task set involved a series of temporally independent tasks that were spatially distributed. While we were unable to complete a true study, we were able to show how HoloVision (or other asynchronous MR groupware) could best be tested to establish ease of use and value; by focusing on location-dependent, time-independent tasks.

Loosely-coupled collaboration scenarios are those that involve ongoing, infrequent, and opportunistic access to information as needed by the information seekers in the collaboration. A core requirement of loosely-coupled collaborations is that collaborations be as easy and lightweight as possible. With HoloVision we provide a way to easily create rich task documentation. Since annotations are spatially distributed, they will exist where they are needed, making them easy to access and use. Since they are persistent, they will be accessible when they are needed. HoloVision provides a feature set that lends it ideally to support loosely-coupled collaborations.

An important limitation here is that in order for these to be truly easy for people to use, they must be easily viewable. With conventional HMDs, they are still relatively bulky, battery life is realtively short and they are not likely to be always worn. This means that there is some fundamental overhead to engage with a system like Holovision.

However, in the future as the technology improves, it is reasonable to expect that the technology will be less cumbersome (in form factor such as eyeglasses), faster and easier to startup and access, and/or be viewable in the environment without a HUD (e.g., through projections in the environment). In this seemingly likely future, it is expected that a solution like HoloVision could be a common and even more desirable tool.

An additional limitation in our current system is the lack of user feedback to the expert; whatever the expert recorded is what the worker gets. With HoloVision, the user must trust that the information in the recording is accurate, and the expert needs to ensure that the information they provide is accurate. During the calculator task in the mock study, we noted a failure to properly communicate the "=" button. HoloVision provides video recordings, which could help solidify a worker's trust in the instructions, but it does not enforce that the instructions are actually accurate. Effective collaboration using HoloVision relies on the assumption that the expert provides sufficient information with their recording. However, it is reasonable to believe that missing some information in a HoloVision recording would only occur infrequently, would likely not always lead to failure, and that the expert would improve their recordings with some practice. Further, future versions of the system could incorporate features to facilitate accessing the original author to augment their recordings with missing details when needed.

7 Limitations and Future Work

Due to COVID-19, more extensive testing of the system was not feasible. In addition to the apartment pilot study (which became the mock pilot study) a larger, and more formal lab-based evaluation of HoloVision was planned, with a focus on the collocated Replay Mode part of the system. The experiment would compare Replay Mode as a task-learning tool versus traditional guidance. Replay Mode tasks would involve several pre-recorded sets of tasks like the set demonstrated in the mock pilot study. Traditional guidance would involve paper instructions accompanied by supplementary verbal instructions and gestures as needed. For the experiment, 12+ participants were expected to complete several sets of tasks involving the location and assembly of Lego blocks (instead of the improvised tasks from the mock pilot study). To avoid the learning effect, the participants would be split randomly into two, equally-sized groups; one would perform several sets of tasks with Replay Mode guidance first, then use traditional guidance, and vice-versa for the second group. For analysis, task completion time, task errors, questionnaire responses would be recorded, like what was planned for the pilot study. The goal of this experiment would be to evaluate the use of an asynchronous, collocated MR system as an effective alternative to task guidance when a human guide is not available.

HoloVision's Record Mode acts as a semantic authoring tool for asynchronous MR annotations in remote and collocated environments (R1). It allows the user to produce

holographic, labelled video clips and automatically place them in a contextually relevant location (R2). While this part of the system is not as stable as its counterpart (Replay Mode), it demonstrates the relative ease at which this rich media can be produced (R3). Note that this part of the system was not originally the primary focus of the work; it was designed to reduce overhead in creating experimental task sets for the evaluation of Replay Mode. Ultimately, Record Mode became a significant part of the system contribution. I believe that future work in this area should focus not just on the benefits of worker guidance, but also on reducing the expert's barriers to the authoring of such guidance material.

It should also be noted that HoloVision is currently only supported on the HoloLens and HoloLens 2 devices, or on PC using spatial maps generated by the HoloLens. Other MR devices may not have the same capabilities as the HoloLens, so attempts to port functionality may be limited in this regard. It is my hope that HoloVision will inspire future work in this area by demonstrating some of the functionality that is now possible with the information that these devices can produce and consume.

8 Conclusion

In this thesis, from a review of related work, I have identified 3 main requirements for a system that can best support loosely-coupled collaboration in physical tasks. Below I revisit the requirements identified in section 0, identifying how HoloVision supports each requirement.

R1: HoloVision facilitates the production and consumption of asynchronous MR annotations to support effective task guidance relative to status quo approaches. It does this with two modes of operation, Record and Replay to respectively produce and consume the annotations. Although HoloVision has not been empirically proven as an effective guidance tool, I believe the evidence provided by the task demonstrations and mock pilot study establish that there is strong potential for such systems in this regard.

R2: HoloVision inherently leverages MR features (like position, gaze, and spatial mapping) to support hands-free, interactions (like in-situ, contextual guidance cues) that cannot be supported in traditional groupware.

R3: Minimizes the effort of creating and consuming information for loosely-coupled collaborators. HoloVision does this by providing an easy-to-use interface based on authoring annotation through speech-based demonstrations

and by enabling annotation access as close to the physical location where they are needed as possible.

While this work would benefit from an objective evaluation of the system, the justification for and demonstration of HoloVision illustrates how the system, through its low-cost production and consumption of contextual traces, annotations, and videos, can make a substantial contribution for the field MR groupware systems. No previous MR system has specifically addressed loosely coupled collaborations, which is best supported by low-effort tools and techniques. HoloVision demonstrates a new way for designers to approach how information can be collected and shared.

Bibliography

Baecker, R. M. (Ed.). (1995). *Readings in human-computer interaction: Toward the year 2000* (2nd ed). Morgan Kaufmann Publishers.

Bateman, S. (2012). *Social Feedback: Social Learning from Interaction History to Support Information Seeking on the Web*. 250.

Bateman, S. S., Gutwin, C. A., & McCalla, G. I. (2013). Social navigation for loosely-coupled information seeking in tightly-knit groups using webwear. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 955–966. https://doi.org/10.1145/2441776.2441885

Billinghurst, M., & Kato, H. (1999). Collaborative Mixed Reality. In Y. Ohta & H. Tamura (Eds.), *Mixed Reality* (pp. 261–284). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-87512-0_15

Correia, N., Alves, L., Sa, R., Santiago, J., & Romero, L. (2005). HyperMem: A system to store and replay experiences in mixed reality worlds. *2005 International Conference on Cyberworlds (CW'05)*, 8 pp. – 92. https://doi.org/10.1109/CW.2005.52

Dillman, K. R., Mok, T. T. H., Tang, A., Oehlberg, L., & Mitchell, A. (2018). A Visual Interaction Cue Framework from Video Game Environments for Augmented Reality. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–12. https://doi.org/10.1145/3173574.3173714

Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., & Billinghurst, M. (2019). Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, *131*, 81–98. https://doi.org/10.1016/j.ijhcs.2019.05.011

Gutwin, C., & Greenberg, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work (CSCW)*, *11*(3–4), 411–446. https://doi.org/10.1023/A:1021271517844

Hansen, F. A. (2006). Ubiquitous annotation systems: Technologies and challenges. *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, 121–132. https://doi.org/10.1145/1149941.1149967

Irlitti, A., Von Itzstein, S., Alem, L., & Thomas, B. (2013). Tangible interaction techniques to support asynchronous collaboration. *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 1–6. https://doi.org/10.1109/ISMAR.2013.6671840

Katz, B. F. G., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., & Jouffrais, C. (2012). NAVIG: Augmented reality guidance system for the visually impaired. *Virtual Reality*, *16*(4), 253–269. https://doi.org/10.1007/s10055-012-0213-6

Kim, S., Lee, G., Huang, W., Kim, H., Woo, W., & Billinghurst, M. (2019). Evaluating the Combination of Visual Communication Cues for HMD-based Mixed

Reality Remote Collaboration. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3290605.3300403

Kraut, R. E., Fish, R. S., Root, R. W., & Chalfonte, B. L. (2002). *Informal Communication in Organizations: Form, Function, and Technology*. 55.

Kwon, J. U., Hwang, J.-I., Park, J., & Ahn, S. C. (2019). Fully Asymmetric Remote Collaboration System. *IEEE Access*, *7*, 54155–54166. https://doi.org/10.1109/ACCESS.2019.2912397

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). Augmented reality: A class of displays on the reality-virtuality continuum. *Other Conferences*. https://doi.org/10.1117/12.197321

Moghadam, P., Evans, B., & Duff, E. (2016). SAGE: Semantic Annotation of Georeferenced Environments. *Journal of Intelligent & Robotic Systems*, *83*(3), 635–648. https://doi.org/10.1007/s10846-015-0302-3

Nebeling, M., Speicher, M., Wang, X., Rajaram, S., Hall, B. D., Xie, Z., Raistrick, A. R. E., Aebersold, M., Happ, E. G., Wang, J., Sun, Y., Zhang, L., Ramsier, L. E., & Kulkarni, R. (2020). *MRAT: The Mixed Reality Analytics Toolkit*. 12.

Palmarini, R., Erkoyuncu, J. A., Roy, R., & Torabmostaedi, H. (2018). A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, *49*, 215–228. https://doi.org/10.1016/j.rcim.2017.06.002

Pinelle, D. (2004). *Designing CSCW Applications to Support Loose Coupling in Organizations and Groups*. 38.

Piumsomboon, T., Day, A., Ens, B., Lee, Y., Lee, G., & Billinghurst, M. (2017). Exploring Enhancements for Remote Mixed Reality Collaboration. *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, 16:1-16:5. https://doi.org/10.1145/3132787.3139200

Piumsomboon, T., Lee, Y., Lee, G., & Billinghurst, M. (2017). CoVAR: A Collaborative Virtual and Augmented Reality System for Remote Collaboration. *SIGGRAPH Asia 2017 Emerging Technologies*, 3:1-3:2. https://doi.org/10.1145/3132818.3132822

Poupyrev, I., Tan, D., Billinghurst, M., Kato, H., Regenbrecht, H., & Tetsutani, N. (n.d.). *Tiles: A Mixed Reality Authoring Interface*. 8.

Ranjan, A., Birnholtz, J. P., & Balakrishnan, R. (2007). Dynamic shared visual spaces: Experimenting with automatic camera control in a remote repair task. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1177–1186. https://doi.org/10.1145/1240624.1240802

Sun, H., Liu, Y., Zhang, Z., Liu, X., & Wang, Y. (2018). Employing Different Viewpoints for Remote Guidance in a Collaborative Augmented Environment. *Proceedings of the Sixth International Symposium of Chinese CHI*, 64–70. https://doi.org/10.1145/3202667.3202676

Thoravi Kumaravel, B., Anderson, F., Fitzmaurice, G., Hartmann, B., & Grossman, T. (2019). Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, 161–174. https://doi.org/10.1145/3332165.3347872

Toups, Z. O., Hammer, J., Hamilton, W. A., Jarrah, A., Graves, W., & Garretson, O. (2014). A framework for cooperative communication game mechanics from grounded theory. *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play*, 257–266. https://doi.org/10.1145/2658537.2658681

Tseng, P. Y., Haraldsson, H., & Belongie, S. J. (2019). *Annotate All! A Perspective Preserved Asynchronous Annotation System for Collaborative Augmented Reality*. /paper/Annotate-All-!-A-Perspective-Preserved-Asynchronous-Tseng-Haraldsson/b5bbd0d0080df8d98c0bb86afcb1d6b0b1665c28

Vayanou, M., Katifori, A., Karvounis, M., Kourtis, V., Kyriakidi, M., Roussou, M., Tsangaris, M., Ioannidis, Y., Balet, O., Prados, T., Keil, J., Engelke, T., & Pujol, L. (2014). Authoring Personalized Interactive Museum Stories. In A. Mitchell, C. Fernández-Vara, & D. Thue (Eds.), *Interactive Storytelling* (pp. 37–48). Springer International Publishing. https://doi.org/10.1007/978-3-319-12337-0_4

Wuertz, J., Alharthi, S. A., Hamilton, W. A., Bateman, S., Gutwin, C., Tang, A., Toups, Z., & Hammer, J. (2018). A Design Framework for Awareness Cues in Distributed

Multiplayer Games. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3173574.3173817

CURRICULUM VITAE

Candidate's full name: Alexander William Kienzle

Universities attended: University of New Brunswick (BCS, 2018)

Publications: None

Conference Presentations: None