Department of Mathematics and Informatics
University of Bremen

Social Panoramas: Sharing Experiences Using a Head Mounted Display. The Effect of Presence and Awareness in a Remote Collaborative Environment

by
Carolin Reichherzer

Supervisor:
Dr. Mark Billinghurst; *The Human Interface Technology Laboratory New Zealand, University of Canterbury*

Reviewers:
Dr. Gabriel Zachmann; *University of Bremen*
Dr. Andreas Breiter; *University of Bremen*

Dissertation submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Department of
Mathematics and Informatics at the University of Bremen

2014

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

....

date (signature)

Acknowledgments

I would like to express utmost gratitude to my supervisor Prof. Mark Billinghurst for his unprecedented support and supervision despite tight time schedules, and his guidance has been of the greatest importance during the entire course of the project. Second, I would like to thank Alaeddin Nassani for being most helpful in providing help with software development and Dr. Adrian Clark for reviewing my work. Additionally, many thanks go to Dr. Gun Lee, who has greatly endured all my nagging questions on statistics and the wonderful people and students that work in the Human Interface Technology Laboratory New Zealand for a warm and supportive environment. Lastly, I would like to thank my home supervisor Dr. Gabriel Zachmann for his support and advice from the other side of the world.

Abstract

Applications which allow for the creation and sharing of panoramic images on mobile devices have gained in popularity due to the better user experience made possible by improved hardware and software. The possibility of sharing the users' full surroundings, instead of just the limited field of view glimpse provided by traditional cameras, introduces an intriguing new way of sharing immersive experiences. However, the current state of panoramic image sharing has limited interaction modalities, which rarely move beyond viewing and sharing a still image.

This thesis introduces the concept of Social Panoramas. Social Panoramas leverage technologies such as panoramic imagery, collaborative interaction and natural communication to provide a new way to connect people over a distance through sharing their immediate and personal spaces. It begins with an exploration of how established collaborative interaction methods in other fields of study can improve user experience, how different modalities affect the users' sense of social presence (i.e. the feeling of being with another), and which visual cues provide the best awareness of the other person's actions in this particular environment. It also examines how panoramic images, Mixed Reality and wearable computers can be used to support remote collaboration.

To show how these ideas can be implemented, this research presents a working prototype that allows panorama images to be shared between a Google Glass user and a remote tablet user. An example use case was created which simulates the Glass user capturing a panorama of his or her surrounding indoor space and sharing it with a remote collaborator . The users were able to interact within a panoramic scene by using pointing or drawing gestures. A variety of cues were developed to support awareness and enable both users to share pointing and drawing. Two user studies were conducted to explore if these cues could enhance awareness and increase Social Presence. The results show that visual cues can significantly enhance awareness compared to audio only, that Glass users prefer pointing to drawing for shared interaction and usability problems decrease the sense of Social Presence. Based on these findings directions for future research are presented.

Content

List of Tables

List of Figures

1 Introduction

This research explores how to use panorama images, and real time awareness cues and gesture interaction to increase the feeling of Social Presence between remote people. In recent years, mobile devices such as smartphones and handheld tablets have become increasingly popular. The ubiquity of smartphones with integrated high-quality cameras has created a social phenomenon of taking pictures at any time and place, quickly enabling users' to share a part of what they are experiencing [1]. In a similar way, wearable computers (Figure 1) that combine head mounted displays and cameras provide new opportunities to share experiences and spaces by taking pictures or video and sharing them with a remote collaborator. However, one limitation of these systems is that they typically only share the portion of the users' surrounding space that they are currently looking at. The purpose of this thesis is to investigate how wearable and mobile devices can use panorama imagery to provide a more immersive view of a user's surroundings and overcome this limitation.

FIGURE 1: Example of a wearable computer as a head mounted display

Panorama applications can provide an easy way for a user to capture and share his or her surroundings in an omni-directional view in one single image. However they are traditionally limited to asynchronous sharing and viewing, not used in real time collaboration. However, modern technology enables communication and interaction over distance, so it is possible to create new methods for acquiring and interacting with a user's surroundings in real time. Research in remote collaboration has shown a variety of ways that people in different locations can interact efficiently, using techniques such as pointing and sketching [3, 4]. In this thesis, pointing and drawing methods are explored to determine how they can support collaboration between a head mounted display (HMD) and handheld device in an immersive environment. Thus, the aim of this research is to examine real time collaboration with panoramas and interaction methods that support the medium the best.

A major issue covered in this thesis is the concept of Social Presence, and how collaboration and awareness cues within a personal panoramic image can increase such presence. Social Presence is the feeling of being with another in the same communication space [38]. A good collaboration based on mutual understanding is connected with the concept of presence and shared awareness of the other user's actions. In the field of Human Computer Interaction (HCI), studies have shown that Social Presence is affected by interface types [5] and interactivity [6]. This project explores several visual cues to provide proper awareness, and examines how interaction patterns have an impact on the sense of presence. Furthermore, it investigates if better usability can produce higher Social Presence.

Figure 2 demonstrates the concept of the Social Panorama. A person is wearing a head mounted computer and by panning their head around they can capture the surrounding space and share it with a remote collaborator as an immersive panorama. Both the wearable user and the remote collaborator can view the space independently while talking and collaborating by naturally pointing and drawing on the image as if they are with one another.

FIGURE 2: Idea of Social Panoramas

The motivation of this project lies in the personal experience of the author not being able to fully communicate the sensation of traveling to a new place or visiting an exciting event. While panoramic images are an improvement over normal still images, the inability to directly refer to topics of interest makes remote communication tiring and cumbersome. For example, trying to refer to a specific object in a rich outdoor environment without using gestures may involve lengthy descriptions until the second person understands what the first person is referring to. Being able to use pointing and drawing gestural tools can provide a useful and more natural way of showing a remote user around a virtual environment.

During the entire project a user-centered design approach was applied. Starting on a conceptual level, we undertook several design iterations beginning with acquiring the user's needs, followed by design studies and the development of the prototype, with a gradual refinement in fidelity, concluded with conducting user experiments.

In conclusion, this thesis aims to investigate how awareness cues, and established interaction techniques for remote collaboration can be used in new wearable application that uses panorama images to increase Social Presence.

1.1 Outline of the Thesis

This thesis draws on existing research from a number of different fields such as remote collaboration, panoramic imagery, presence and wearable computing to motivate the work, guide the execution of two user experiments on awareness and presence, interpret the results and define the contribution of this work.

The chapter two, *Related Works,* presents an introduction to previous work in panoramic imagery, panorama creation, and use of panoramas in combination with wearable devices. Interaction patterns commonly found in remote collaboration applications are also described.

Chapter three, *Collaboration and Presence in Virtual Environments*, follows on from the related works by explaining the theoretical background. Several related concepts are explained, beginning with g*rounded communication*, which builds the basis for mutual understanding and thus successful collaboration, followed by *awareness,* which builds upon grounded communication, providing cues about the action of each user, and finally *social presence*, or the feeling of "being with another".

In chapter four, *Process*, the methods used to develop the successive prototypes are presented. This chapter covers the initial idea of interaction on several devices and accompanying interfaces, use cases, and a description of the decisions that ultimately lead to the working implementation. The process is based on a User-Centred-Design methodology, and as such follows an iterative approach. Pilot studies were conducted to evaluate each iteration, and the results and needed changes are discussed afterwards.

Development of the Prototype in Chapter five describes the hardware that was used and the software developed in this research. Currently available technology is reviewed, a final platform chosen, and the developed software is described in detail.

In chapter six, *User Evaluation*, the methods used to conduct user studies on a wearable computer and tablet are described. It summarises methods and results of several studies that have been conducted during this research. The demographics of the participants are presented as well as an overview of the pilot testing, followed by the two main experiments, measuring the effect of interface type on awareness, and interaction method on social presence.

Results follows User Evaluation with the results of the user studies and conclusions that can be drawn based on the results. Chapter eight, *Discussion and Limitations*, discusses the findings in detail and which effects the tested variables had, as well as limitations of the study and the impact on generalisation of the results.

Finally, in *Conclusions and Future Work* conclusions are drawn and suggestions for future research are made. The thesis will conclude with a summary of the entire project.

2 Related Work

This thesis expands upon existing research from several fields, including panorama imagery, wearable computing and collaborative systems. This chapter presents each of these fields individually, and concludes with a description of research opportunities and how this research is novel compared to these earlier works.

2.1 Panoramic Imagery

Work on panoramic imagery has been around since the 1990's [23], and with increasing technological advancements panorama applications have become popular in recent years. Such applications as Google Street View [7] (Figure 3), which has been downloaded over 500,000 times [75], or Microsoft Streetside [8] offer 360-degree-imagery of street scenes. These can be used to explore remote locations in a similar manner to that of a virtual tour. However, these immersive panoramic scenes require special hardware and well-calibrated cameras to capture the panoramic image, which comes with a high cost and thus are not accessible for the end-user.

FIGURE 3: Screenshot of Google Street View

An alternative method of creating panoramas uses image processing algorithms on consumer devices such as mobile phones and digital cameras [9]. Advances in hardware for mobile devices and smartphone platforms has resulted in widespread use of portable devices equipped with high-quality cameras. There has been significant interest in the ability to quickly and easily create panorama imagery at anytime and anywhere [10]. After years of research in the field of computer vision towards algorithms that can quickly and easily create panoramas, it is now a popular feature on mobile phones to offer functionalities to combine or stitch collections of images together [11]. One such example is Ztitch [10], an application for

Windows Phone, which lets users create, modify and upload panoramic imagery onto an online portal to share with other users. After capture, users can look around the image naturally, with the phones accelerometer rotating the image with the phone (Figure 4).

FIGURE 4: Panorama acquisition with Ztitch (left) and Photosynth (right)

Other existing panorama applications include Photo Sphere [12], TourWrist [13], AutoStitch [14] and Photosynth [15]. However, interaction with these platforms is limited to online viewing and there is no support for real-time collaboration. There are few, if any, handheld interfaces that combine real-time remote collaboration or annotation together within an immersive panoramic application, although prototype systems such as that of Langlotz support asynchronous collaboration and annotation [17].

Besides applications that simply create and share Panoramas, there has been ongoing research that uses panoramas in more functional ways. The authors of PanoInserts [9] provide a system for teleconferencing using smartphones to create a static panorama of a remote place, which is then augmented with live video inserts. This allows users to explore the environment of the remote person in real-time. Several mobile devices are required; one to create the panoramic imagery, and one for each participant to record their video stream. Each user receives the video streams from all other participants, presented with a representation of the meeting space. A similar application is Indirect Augmented Reality [18], which simulates an outdoor AR environment by replacing the live image with a panoramic image to reduce the perceptual tracking error on a mobile device. Similar to the approach in this thesis they use a lens interaction metaphor – by physically turning the device it is possible to explore the corresponding parts of the panorama.

There has been some research in using panoramic imagery on mobile devices to provide a more interactive TV experience. In FascinatE [19] a high-resolution video panorama of a live event from a given viewpoint is broadcast to a mobile device, panoramic display or TV. During events such as a concert or a sporting event, professional panorama cameras at the location capture a live panoramic image, while users at home can access these panoramas with their mobile devices and browse the environment. FascinatE offers gesture-based interaction to zoom in or choose a region of interest for the user to follow, for example tracking a player in a sports event. A similar concept, described in [20], provides a real-time panoramic video streaming service for devices such as smartphones, tablets or smart TVs. Here, the interface also offers social media functionalities and collects social media content from their respective network. In these examples the panorama itself was not created on a mobile phone, but it was used to display interactive panoramic imagery within a smaller social context.

FIGURE 5: Panorama view with social networks

A more consumer friendly and cost-effective option is CamBlend. This is a video collaboration tool that uses three cameras that can be installed in any space by the user, and aims to facilitate interaction and referencing in a panorama view [21]. Pointing is used as the primary interaction tool, and a 180° wide field of view allows for a better understanding of the users surroundings in a face-to-face scenario. The user has a focus windows at their disposal, which allows zooming into the panorama or taking of a screenshot to record the state of the system. In addition, it is possible to see a focus window with the video stream of another user in order to identify details of the scene.

2.2 Panorama and Head Mounted Displays

There is on-going research on how to view and interact with panorama imagery inside a Head Mounted Display (HMD). In this section we review some of these efforts, and draw out lessons relevant for our research.

FlyVIZ [16], is a head mounted display device specifically designed to acquire panoramic images. Similar to previously presented desktop and mobile application, this device takes the form of a helmet which provides 360° vision in real-time, however it does not provide any interaction.

The Touring Machine [22] (Figure 6) is a famous example of a very early version of an outdoor AR system using a HMD to provide a virtual tour guide. In some applications shown on the Touring Machine the user could view panorama imagery of a historic view of a real world location. The panoramic image would turn with the user's head movement, while additional data such as video clips is provided on a handheld display. This offered some basic interaction to access information of the environment, but offered no support for remote collaboration.

FIGURE 6: Touring Machine

2.3 Wearable Computing and Sharing Experiences

Since the early days of wearable computing there has been research on capturing and sharing experiences. Mann's WearCam and WearComp systems [23] allowed a user to stream live video to a remote collaborator, and to capture and stitch together panoramas of their surroundings. Unfortunately, these systems had limited support for remote awareness cues or for shared annotations. Unfortunately, these systems had limited support for remote awareness cues or for shared annotations. Cheng [24] developed a system for collaboration that used image mosaics (Figure 7) from a wearable computer and allowed a remote user on a PC to place annotations in the wearable user's field of view. However, their work focused on image tracking and they didn't evaluate the usability of the system.

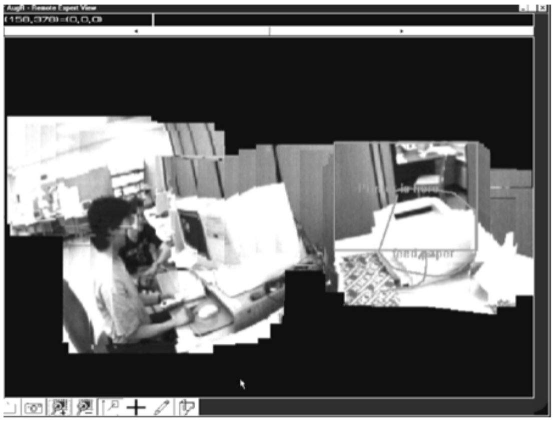


FIGURE 7: Image Mosaic

Research from Kraut [25] and others explored how head mounted cameras, HMDs and annotations on live video and still images can be used to enhance remote collaboration. They completed a number of user studies that showed how remote pointing and drawing can assist in establishing common ground and understanding. However, in most of these systems the remote user's view is limited to the live camera feed from the head mounted camera, causing reduced awareness of the user's surroundings. TAC [26] addresses this problem by using wider field of view cameras to support greater situational awareness. Panoramic imagery can provide an even more immersive view of the current environment.

2.4 Interaction in Remote Collaboration

One of the earliest wearable systems using pointing gestures is The Wearable Active Camera/Laser-Pointer (WACL) [28] (Figure 8). This system featured a head mounted display and camera that the worker would wear, while the remote expert would receive a live stream and would be able to set his or her own independent view point with a remote controlled camera. With a laser spot/pen he or she was able to point at locations and objects.

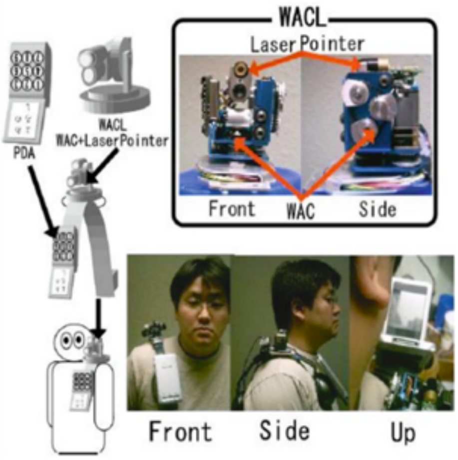


FIGURE 8: WACL system

Similarly, GestureMan [29] allows a remote person to control a robot at the local site that can execute pointing instructions with a laser pointer. However, end-users perceived it as too difficult to control the robot. In the Drawing over Video Environment (DOVE) [30], hand gestures are shown inside the video stream of the remote helper and displayed on a display to the local worker to facilitate remote gesturing. Similarly, [31, 32] built a system where the hands of the helper were sent via a video stream to the local worker and projected onto his desk. Poelman et. al [33] describe a system to support collaboration on a crime scene where investigators at the scene are remotely supported by experts. The colleagues at the scene wear a head mounted display with stereo cameras, which maps the entire scene in 3D and gives the expert user the ability to have an independent viewpoint and provide a shared augmented space. Additionally, the hand gestures of the local user are tracked and streamed to the remote user. The lack of presence however resulted in difficulties to effectively communicate and users ended up often interrupting each other. HandsOnVideo [34] is a project that further explores the richness of hand gestures. Here, the worker wears a display close to his eye and can see the representation of the helper's hands. The remote collaborator can see the viewpoint of the worker on a big touch-enabled display. Unfortunately the size of the display used in this system meant it was too big to be portable.

Most of the existing research in remote collaboration features an asymmetrical setup, where the equipment used and interaction patterns involved are not equal for both participants. However, there are collaborative sketching systems that make use of a symmetrical system using video projection, such as VideoWhiteboard [35] and VideoArms [36], which can be used to create shared drawings in a collaborative activity [3]. However, these were made for a work environment and not for casual communication.

2.5 Research Opportunity

There is a trend in ongoing research for technological improvement in terms of better algorithms and applications using panoramas for a work environment. In particular, optimisation of efficiency for a specific task in a workflow is of major importance, while neglecting a more general enjoyable user experience. To the author's knowledge, there are no systems which connect users on a social basis sharing their entire environment in real-time and supporting multiple interaction possibilities.

While the basic panorama viewing examples offer simple functionality, they do not go beyond sharing and commenting on a still image. The applications reviewed which had an entertainment purpose did offer some more advanced interaction possibilities, however the main goal was not collaboration and sharing experiences with another user. Additionally, the panorama pictures provided were not created by the user, but rather provided by a special and expensive set of hardware implemented directly at the scene of event.

Previous research has put a lot of effort into improving remote collaboration for specific work tasks, as well as capturing and sharing the captured panorama. These fields have been happening in parallel, however there has been no common thread to take advantage of each fields benefits and no common use case that could potentially further connect users in a world where personal interaction is already digitised. Social Panorama is means to connect the dots by providing immersive imagery in a collaborative environment with the use of natural interaction tools.

Based on these research opportunities this research makes the following novel contributions:

- explores the use of real time panorama sharing for remote collaboration
- presents some of the first formal user studies on pointing and drawing interaction methods in shared panoramas
- studies the effects of presence and awareness inside a panoramic image
- explores presence and usability on HMDs and a tablet

3 Collaboration and Presence in Virtual Environments

Successful collaboration requires a common ground between participants that can be achieved by mutual understanding of each other's actions. The feeling of connectedness or being present in the same space, is defined as *Presence.* Both collaboration and presence have been extensively researched and shown to improve efficiency in mediated interactive environments. This chapter provides an introduction to both principles, and the connection between the two. Furthermore, methods of evaluating Presence are presented and their use in this thesis is defined.

3.1 Collaboration

Successful remote collaboration can be challenging, as the participants have to communicate via a restricted medium and cannot rely on all the senses they would normally use in face-to-face communication. Remote collaboration has been studied extensively [34, 37], with typical use cases involving a worker (a local person needing guidance), a helper (a remote expert), and a task that involves physically manipulating an object. Example scenarios include a local technician needing help with a remote expert guiding him through maintenance, a specialist doctor helping a non-specialist doctor, or investigation of a crime scene. As the collaborating people cannot see each other, non-verbal communication cues common in face-to-face communication are not possible, and ways have to be found to translate these cues in a restricted medium to support natural interaction patterns. In remote collaboration it is important to share the same visual space, as well as provide gesturing abilities [3].

Collaboration requires coordination between the participants and the achievement of mutual understanding. Collaborative interfaces aim to provide an environment where users can establish a common mindset, which is commonly referred as "grounding" [39]. This refers to the communication process between people and their understanding of the situation, and the subsequently build up of a common ground [39]. To achieve mutual understanding, people must be aware of the activities of the other user [6], a process known as shared awareness. When people collaborating are situated in remote locations they cannot rely on cues that are typically used in face-to-face communication, such as facial expression or gestures. In order to establish successful grounding it is necessary to provide situational awareness to aid the interaction. Shared awareness can therefore be defined as "an understanding of the activities of others, which provides a context of your own activity" [40].

Wisneski defines shared awareness as “the state of knowing about the environment in which you exist; about your surroundings, and the presence and activities of others” [41]. Awareness provides information about what is happening in the environment, of the task and the partner’s activities. The user can use this information in order to make a decision relevant to his or her task and can help reach a specific goal. For example, an icon indicating the online status of another user in a text messaging service provides awareness. Shared visual spaces, in the sense of seeing the same as the other user, have been proven to be helpful in establishing grounding and increasing efficiency for completion of tasks [27].

Other studies suggest that grounding can be established through using gestural interactions to reference objects and verbal expressions. An example of this process would be the use of a referential expression to select an object, followed by instructions on how to manipulate it, and finally confirmation that the outcome was what was desired. Besides speech, gestures are used in order to reference objects and have been shown to increase task performance [27]. In collaborative tasks, it is common to identify the object in question, describe what the person is about to do and confirm once the action has been done. A user can feel helpless in a virtual setting without having the means of directly identifying an object or environment [42].

In the field of remote collaboration there has been research on the impact the use of representations of gestures on collaboration tasks. Representational gestures attempt to represent the task in a natural way. Examples for representational gestures are pointing and drawing. Pointing is often used to reference and locate objects, while drawing gestures such as hand movements or shapes are typically used more figuratively, as in demonstrating the form of an object or the nature of tasks, and carry a higher amount of semantic meaning [27, 36]. Pointing gestures have been shown to be of high importance in grounding deictic references and are easy to interpret [39], while it has been suggested that more complex gestures are beneficial to increase performance [27] and increase the effectiveness of collaboration [3]. The *surrogate* approach expresses the intention by images such as cursor pointers or drawing tools instead of showing the hands directly [27].

With surrogate gestures providing the ability to achieve mutual understanding quickly, as well as offering a way of introducing patterns that are common in human communication, it is important that Social Panoramas provide tools for both awareness and also conveying gestural cues.

FIGURE 9: Gesture types.
The hand as a representational gesture (left), and drawing as a gesture surrogate (right).

3.2 Presence

The concept of presence has various meanings and definitions, and can take the form of Spatial Presence, Telepresence, co-presence, Virtual Presence and Social Presence among many others [38]. The term Presence in a virtual environment is generally referred as the feeling of “being there”, and can be found in almost all mediated environments [44]. This refers to a sensation of feeling physically present, and can also have the terms “physical presence”, “telepresence” or “mediated presence” [43]. Co-Presence goes further and includes mutual awareness of another person [5]. Social Presence was introduced by Short et. al [45] as “the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships”, and emerged in HCI by researching efficiency and satisfaction in different telecommunication media [46].

Social Presence has been used quite extensively in the study of remote collaboration, often in combination with additional sensory information such as haptic feedback [6, 38]. In the area of remote collaboration, providing users with the sense of being together has shown to contribute to the perceived effectiveness of new technologies. It is of importance for the grounding phase and thus for the success of collaboration. Underlining the grounding theory is the need for two users to be aware of each other in a shared virtual space, so it can be assumed that mutual awareness is the essence of social presence [5].

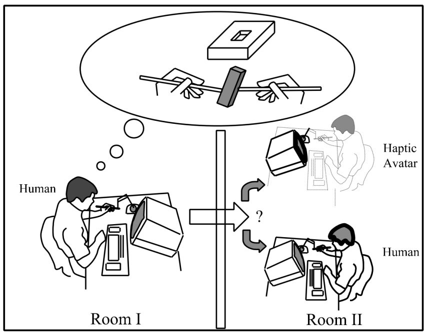


FIGURE 10: Study on presence using haptic feedback

Although the definitions of presence share a common ground, the different terms can be confusing. It has been suggested that the concept of presence is a continuum that can vary through the course of interaction, where other participants can be more or less present [5].

A similar concept to Social Presence is that of Media Richness [47], defined as the verbal and nonverbal information that a medium is able to transmit in a given time interval. Based on this definition, the highest form of richness provides by face-to-face communication, followed by telephone, voice mail and lastly written messages [48]. The assumption that higher richness provided more Social Presence is not guaranteed [47], for example, text messaging would score low on media richness, but can have high Social Presence depending on the content of the message and how well the receiver knows the sender.

In the scope of this thesis, we refer to presence as the ability of a medium to connect people, otherwise also known as "the sense of being with another" [5].

3.2.1 Evaluating Presence

The concept of presence has been used by researchers in different fields, resulting in varying definitions and different approaches in measuring it. Although the idea of presence goes as far back as the 1990s, a common definition is yet to be found [49]. Universal definitions and means of measurement have been suggested, but a consensus is still lacking [5]. Currently, there are several types of Presence measurements that can be of either subjective or objective nature.

Presence has been argued to be mainly a subjective sensation, to be assessed with subjective methods [49]. These subjective measurements include the personal impression of the participant in relation to the mediated environment. The biggest group of subjective measurements is comprised of questionnaires, ranging from measuring one dimension to multidimensional assessment (e.g., spatial presence, quality of the interface and emotional involvement). The design of these questionnaires depends highly on the definition of Presence used. Questionnaires can have from only two items [50] to up to 103 items [51], and their high validity and easy administration makes them popular for practical use. Further ways of assessment are qualitative methods and subjective corroborative measures (such as perceived task performance) [49]. In this thesis, we base our subjective measures of presence on a questionnaire first created by Basdogan et. al. with eight questions that define Social Presence as a "*Sense of Togetherness*" [71].

Objective measurements comprise of bodily responses or other means of measuring objective criteria. Objective measurements can range from physiological measures (e.g., recording of facial electromyography via electrodes), to behavioural (e.g., face expressions, posture) and task performance. The physical responses in a displayed environment become similar to real world responses to the same objects or events. For example, cardiovascular activity or skin conductance are connected to emotional experience and can indicate how immersed the user is in the virtual environment. However, physical measurements can be ambiguous depending on what is actually being measured, and behavioural observations can underly the bias of the observer who interprets the results. While a high task performance may indicate a higher sense of presence, there is no evidence showing a causal link [49].

This thesis uses a mixture of both subjective and objective approaches, with the main element being a post-test questionnaire on a multidimensional level with added observations and measured performance, to acquire a deeper understanding of what the user is experiencing.

4 Process

The aim of this research is to create a system for remote collaboration that provides a high feeling of presence and a good user experience. All of these requirements focus on the user, and as such the process employed during development of the system was User-Centred Design. This chapter explains the User-Centred Design process in detail, and how this process was used to brainstorm initial broad ideas, conduct pilot tests, and draw conclusions for further development of the prototype.

4.1 User-Centred Design

In order to provide a good user experience for mobile and wearable devices alike, a User-Centred Design approach was followed. This focuses on understanding the user's needs during the entire process. The idea of putting the user's needs first for systems that involve interaction was first described in 1985 by Gould and Lewis [55], who established three key principles that are still in use today: (1) An early focus on users, (2) Iterative Development, and (3) Empirical Measures. Following these principles requires including target users early in development with interviews and discussions. During the iterative process, it is common to create several prototypes and test each stage with user evaluations, increasing fidelity as the process goes on. The results of these user evaluations are used to improve the prototype, and the cycle repeats [56]. This refinement goes on until the system reaches a state that is considered final. The empirical measures are used to test the product in quantitative and/or qualitative ways, such as performance and task completion time for the former, and opinions, feedback, attitude for the latter.

The typical process goes as follows: (1) The early prototypes are created fast and simple, often starting out with a pen and paper approach. (2) Using an iterative approach, these prototypes undergo a user evaluation, with the results being used to improve the prototype. The prototype will increase in fidelity as the development goes on. Short iteration steps will improve the outcome of the prototype.

4.2 Concept

The initial concepts were approached in a top-to-bottom manner. The first ideas explored how the final application would function at a high level. The first concept included targeting not only wearable and tablet devices, but also a version for the Desktop, and annotation methods that were not implemented in the final prototype. Based on this holistic view, the concept was separated into what was considered the most important steps for the first set of experiments, and carried out with a pilot study. In this section, the initial studies and ideas are presented, while the execution is presented in Chapter 5, Development of the Prototype.

4.2.1 The Big Picture

The general idea of the concept aims at exploring the possibilities of panoramas in a way that remote collaborators would feel as if they were in the same place, e.g. traveling together or having a look in an apartment together. Figure 11 demonstrates the idea of a user in an outdoor environment who is having a conversation with a remote tablet user. The system should feel as intuitive as possible with the tools provided. The two main goals were to create new experiences by sharing the same spaces together, and create the feeling of being together in the same space.

FIGURE 11 : Illustration of the general idea

There are some aspects that had to be considered. Different devices require different ways of interaction and displaying information. Wearable devices typically offer a small display with low resolution, so consideration must be made during the design stages such that visual elements are not a distracting size and are always readable. In addition, this scenario targets a private environment where communication and interaction tools should be the same for all users involved. Users should feel comfortable using the application.

4.2.2 Requirements

Panoramas can be taken in a variety of places, indoors and outdoors. To determine where users most often capture panoramas, around four hundred panoramas created by the public using the Photosynth platform [15] were analysed. Photosynth offers several viewing options such as “Most Recent”, “Top Favorites” and “Most Viewed”. Two hundred pictures from the categories “Most Recent” and “Most Viewed” were examined, with the rationale that the “Most Recent” category gives a broader idea about where users like to take panoramas, and the “Most Viewed” would give an insight on the panoramas which other people are most interested in. Table 1 below shows the results. As can be seen, there were slightly more indoor panoramas created than outdoor in the most recent category, but people viewed more outdoor panoramas.

	Outdoor	**Indoor**
Most Recent Category	86 *(43%)*	114 *(57%)*
Most Viewed Category	135 *(67.5%)*	65 *(32.5%)*

TABLE 1: Panorama analysis on Photosynth

While this cannot be seen as representative, it shows there is user interest in capturing both indoor and outdoor panoramas. Further investigation provided some insight on locations where people feel the desire to take panoramic images. Indoor panoramas were usually offices, theatres, restaurants and cafés, conference venues and rooms in a house, while outdoor panoramas were often taken during a hike, overseeing a city, or the beach.

Following this investigation, an initial brainstorming session was conducted, focusing on what could be offered in an application that could potentially increase the sense of being with another person. Potential users were interviewed, and usability issues of existing systems were discussed.

From the brainstorm, the following key ideas were uncovered:

- Overlaid information, such as the exact location of remote user
- In addition to mobile platforms, an application for desktop users
- Toggle between panorama views
- Spatial information (audio, distances in panorama image)
- Cues that would indicate emotional state on a happiness scale
- Ease of use

To future explore the concept of an application where a user can share his or her environment, people were observed having a conversation. Two pairs of users were asked to sit once in a room, and once outside a building, and to talk about the environment for a few minutes. Besides looking around, one commonly found pattern of behavior observed was pointing at an object or location when referencing it in the verbal communication. While it is not possible to explicitly recreate this type of interaction in a panoramic application, it is possible to indicate shapes and outlines with the help of hand gestures. Although gestures can still be open to the interpretation of the other person, the use of technology can make the meaning become more clear and accessible to both users. Due to this result, adding drawing functionality was considered as well.

From this observation, the following key elements were chosen for implementation:

- Display a panorama image and looking around the image independently
- Offer awareness of the other user's actions and viewpoint
- Pointing abilities
- Annotations such as drawings or uploading images and audio files
- Switching modalities between the interaction
- Audio communication

4.2.3 Use Case

The typical user of the proposed system is imagined as someone who has taken a panorama of his or her environment and would like to discuss and show things to another participant. In this scenario, the user who has taken the panorama would share his image with the other user. Both users are able to look around the panorama independently. The user will be aware of their own viewpoint and know where the other user is looking. Should either user decide to draw or point inside the image, it will be shown on both displays.

The previous considerations were taken into account to imagine a larger scenario. User's that are traveling, looking for a new apartment or visiting a new indoor venue often take pictures of their surroundings and share them on social media. This could be further enhanced by taking a full panorama and sharing the experience with a friend. Potential interactions include pointing to a specific spot in the environment as if the friend is standing beside the user, or drawing something directly on the image to communicate creative ideas. Creative ideas can be shared immediately, for example a person moving into a new apartment can discuss with a helping friend the interior design and furniture arrangement while he or she is picking the items in a store, or standing inside the apartment and deciding of what could fit. Another use case could be students looking for a spot for an independent film project that could share their impressions without having the whole team traveling, while allowing the full staff to be involved in drawing and sketching to communicate ideas back to the remote user. In contrast to a work environment, a system that puts its focus on user experience has a lesser need on efficiency – while it is necessary to have a smoothly running system, the task performance itself is of less importance.

4.3 Hardware

This work presents an application which showcases a panorama image on a hand held device and a Head Mounted Display. Subsequently, suitable devices had to be found that would not just be able to run the application smoothly, but also fulfill some requirements.

Some aspects of the wearable device for viewing the panorama had to be considered:

- It should be lightweight and small
- It should not obstruct the user's movement
- It should blend in with the real world for the panorama capturer

There is a range of commercially available tablets, and to lesser extent wearable devices, of which several were available at the time of prototype development. The following section reviews available options and explains the final selections.

4.3.1 Head Mounted Displays

Three HMDs were available: VuzixWrap 1200 VR, Brother AirScouter and Google Glass (Figure 12). The VuzixWrap 1200 VR did not provide a see-through display. In our use case we wanted the HMD user to still be able to see the real world around him, and additionally be able to wear the device casually on an everyday basis. This leaves the two remaining devices.

FIGURE 12: VuzixWrap 1200 VR (left) Google Glass (middle) and Brother AirScouter (right)

The Google Glass and Brother AirScouter share similarities such as being lightweight, a similar form factor, and a monocular, optical see-through display, with the latter providing the possibility of attaching the display to either eye. While the Brother AirScouter offers a better resolution than the Google Glass of 800x600 pixels (compared to 640x360 pixels), it is quite big in comparison to Google Glass, making it less suitable for everyday and prolonged use. Also, it offers no input device, and a separate one would be need. While the AirScouter would serve the purpose of building a prototype, the smaller and more lightweight Google Glass was more appealing as a realistic target device.

Google Glass is the wearable computer developed by Google. It is worn as a pair of glasses, equipped with a display in front of one eye of the user. In addition to the display, it has a camera, microphone, and bone conducting transducer in a form factor that fits on the head. The display has a resolution of 640x360 pixels with a 15 degree field of view. The camera features a wide angle field of view with 5 Megapixel image resolution and 720p HD video recording. Interaction is conducted using a touchpad mounted on the right side of the device, which can be navigated using swiping or tapping gestures. Furthermore, the interface can be controlled by natural voice commands. The device pairs with a mobile phone to offer functions such as text messaging and video conferencing using Google+. The device can be attached to a computer via Micro USB, and applications can be built with the Glass Development Kit (GDK), an add-on to the Android SDK.

Overall, the Google Glass was more favourable than the Brother AirScouter as it was more compact and lighter. Additionally, as Google Glass applications run on the Android API, this would make the application more portable between the HMD and Tablet interfaces.

4.3.2 Tablet

With the performance and portability of modern tablet devices, and the lack of specific requirements for the tablet application, the choice of the hand held device was straight forward. In this project the application was developed to run on the Google Nexus 7 tablet. The Nexus 7 offers a resolution of 1920x1000 pixels on a 7 inch screen, runs on a 1.51 GHz quad-core Qualcomm Snapdragon processor and features 2GB of RAM.

4.4 Look and Feel

Several design studies were conducted in order to determine the look and feel requirements. The main hardware targets were a tablet PC and a wearable device, as these were seen as realistic devices in real world situations. Even though handheld displays have a small screen they can still provide a level of immersion in virtual environments similar to that of larger displays [57]. A desktop client was considered early in the user-centered design process, so some of the earlier concepts were also explored on this platform.

The concept interaction flow is shown in Figure 13. The experience begins with a Title Screen, followed by the option of creating a new panorama or loading an existing one. The user can either create a new panorama and share it with another user, or they can load a pre-captured panorama and start a connection with the remote partner by triggering the sharing option. Although shown in the overall flow, the creation of the panorama is not further considered in the scope of the thesis, as the focus was on presence and awareness, rather than image processing algorithms. When sharing the panorama, both collaborators can point, annotate or add further data such as images or audio files.

Two things were important for the interface: The support of surrogate gestures such as pointing and drawing tools, and an interface that provides awareness of the actions of the collaborator.

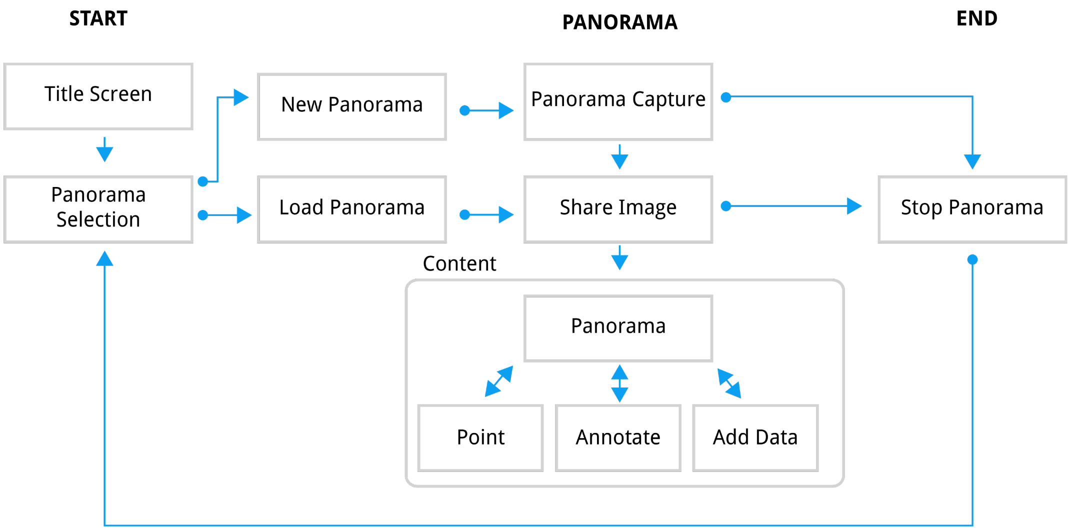
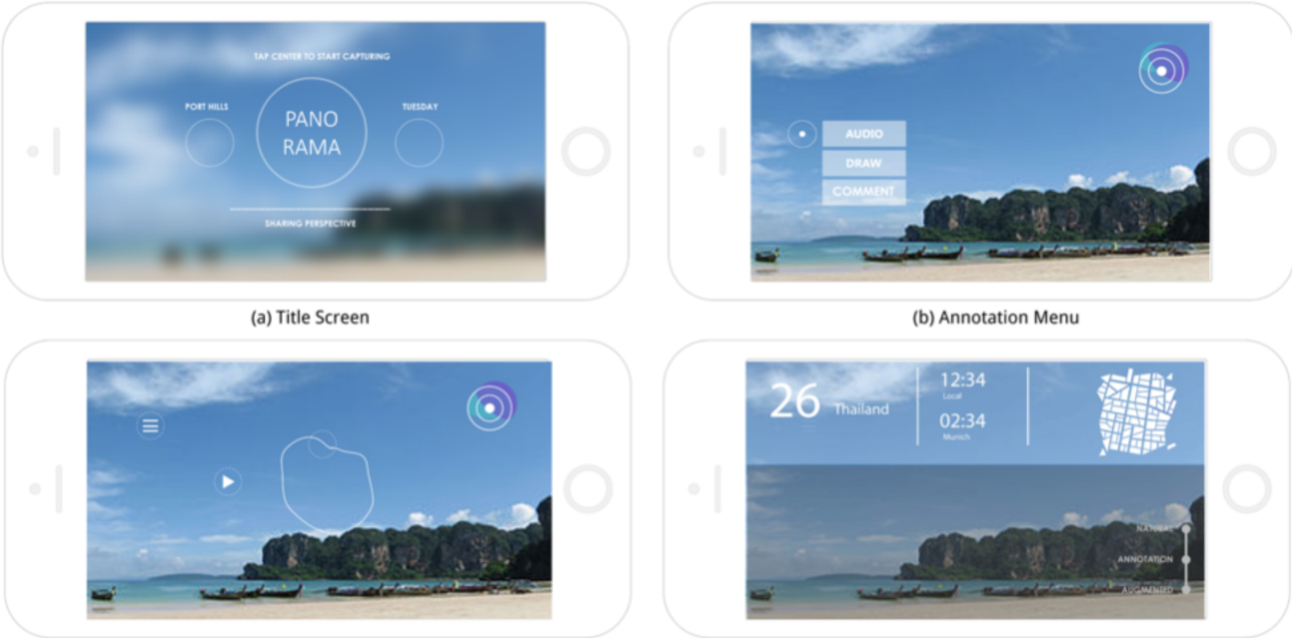


FIGURE 13: Interaction flow

On Mobile Devices

Figure 14 shows some examples for a mobile device. Annotation takes place by touching the surface of the device, which then opens up a menu. Once the annotation has been made it will leave either the drawing itself on the screen, or an icon indicating there is more to explore to not clutter the screen. The radar on the upper right shows the viewpoint of the current and remote user. Additionally there is the concept of creating “layers” to switch between further information such as location and maps or switching back to the naked panorama.



(a) Title Screen

(b) Annotation Menu

(c) Drawing and Icons indicating more annotations made visible by touch

(d) Additional Layer presenting information about the location of the user

FIGURE 14: Smartphone screens

For tablets there are some additional options (See Figure 15). The panorama can be viewed as the full image or using the normal camera view to just view a portion of it. The rectangles represent the users own viewpoint on the full view as well as the viewpoint of the collaborator. The circles in the centre represent the centre of each users view.

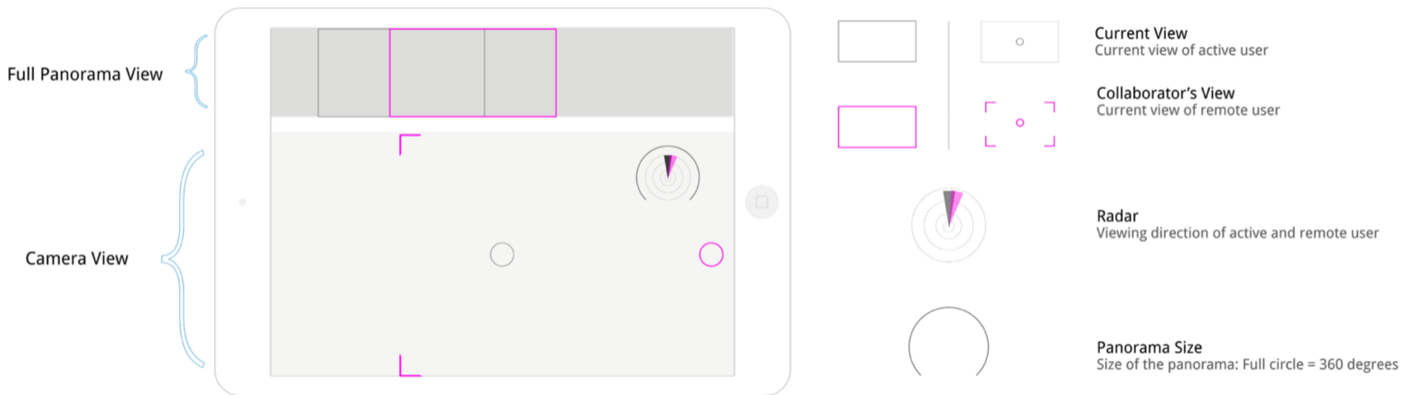
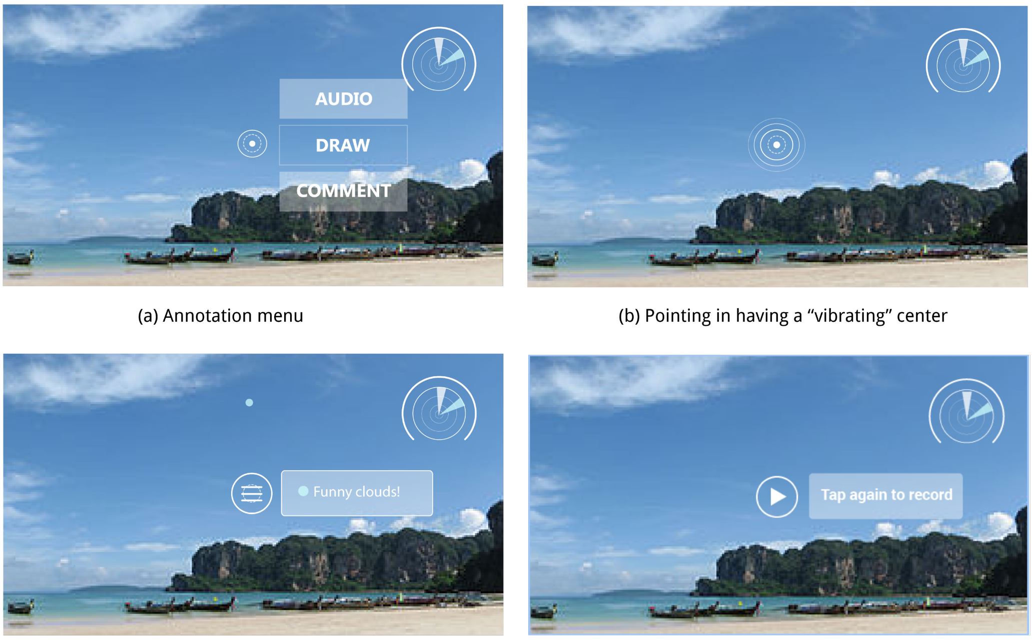


FIGURE 15: Elements of an UI for tablet

Wearable Device

Figure 16 presents the concept interface design for a wearable device, such as Google Glass. Assuming the display sits on front of one eye and has a small display, the menu has been made much larger to accommodate for the smaller screen. The design follows the same principles as on the mobile, providing annotations via small icons; however, as movement is presumed to be difficult on a surface for a wearable device, the head movement is the main "cursor". The centre point acts as a cursor that can be translated by moving the head. Navigating the menu occurs by swiping on the touchpad up and down. Tapping invokes an action, such as placing a comment. Pointing can be simply done by tapping once, and the centre point will turn into a vibrating animation, visible to the other user as well.



(a) Annotation menu

(b) Pointing in having a "vibrating" center

(c) Example for a comment

(d) Example of activating an icon

FIGURE 16: Display view on a HMD

Desktop

While the idea of the Desktop interface was later dropped, there were some initial concepts that will be shown for the sake of completion. The display on a desktop computer provides much more space, and this can be used without implementing any further interaction layers while not cluttering the screen with unneeded information. The "extra" layer of additional information can be found at the bottom of the screen, while the side panel shows previously taken panoramas or simply a history of these. Interaction is performed using the mouse. The two circles indicate the possibility to toggle between a 360° view that allows the user to see the whole panorama at once or a normal view.

Due to the large display space, the desktop application offers a lot of possibilities in what additional information could be added to the display and what could be shared. This could include uploading and transforming personal images into the panorama, displaying the GPS location included with a map, or showing time and weather information of the remote person to offer an immersive experience.

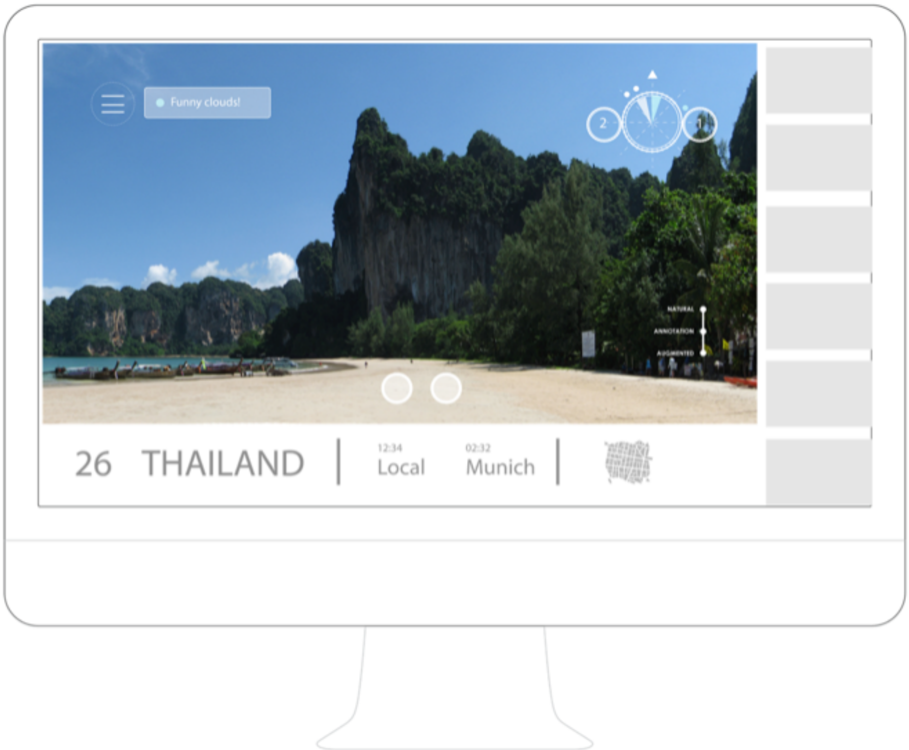


FIGURE 17: Desktop version

As the project developed the scope focused on the use of different interaction modalities and interfaces for awareness between the participants, and so the interface was restricted to pointing and drawing as annotation methods.

During the design process the idea of having a menu to switch between annotation possibilities turned out to not work on the wearable device: The menu options were too small to be easily readable, and scaled elements would take up too much space. Additionally, the menu would create another layer of interaction, and there are methods of accessing the functions faster. Thus, there was the idea of accessing the modalities by the number of fingers on the screen or touchpad. The idea of the head movement as a cursor was disregarded as well. The constant movement of the head proved to be too tiresome and candidates lost interest quickly.

Other points that were considered relevant for the first steps of implementation were considered: Independent viewpoints and support of awareness through the use of visual cues and interaction methods such as pointing or drawing.

4.5 User Interface

To determine which user interface features were important for a good user experience, several existing collaborative applications were examined. Previous research has suggested that non-verbal communication cues are important for collaborative tasks. Collaborative tasks require more information than just receiving feedback and interacting individually. It is important that each user has an understanding of what other users are doing [58]. A panorama viewer provides an independent viewpoint for each user, so it is necessary for both users to be aware where the other user is looking, and what he or she is doing.

There are two possible reference frames representing the mental representation of a user: Exocentric and Egocentric [59]. The exocentric approach is a top-down view and is a commonly used approach in AR applications. One common exocentric metaphor used in many AR applications is the Radar, first proposed in 2004 [60], which indicates the users viewpoint in relation to points of interest (POI) in the immediate surroundings. Examples can be found in applications like Wikitude [61] or AntarcticAR [62]. Figure 18 demonstrates the use of a radar in AntarcticAR. The radar is usually located at a corner of the screen giving an overview of the environment, and shows the viewing orientation using a triangular shape that moves in a circular motion according to the orientation of the device being held. Usually these cues are designed for only one user's point of view.

FIGURE 18: Radar example on the lower left

Another often used exocentric element is the Context Compass [63] (Figure 19). It usually appears as a box on a line on top of the screen, where the box represents the users current view, the line represents the full 360 degree view, and the box moves side to side according to the head orientation. The Context Compass was designed to give an overview of the real world through head mounted displays. It has some benefits over the radar view as it clutters less of the screen.

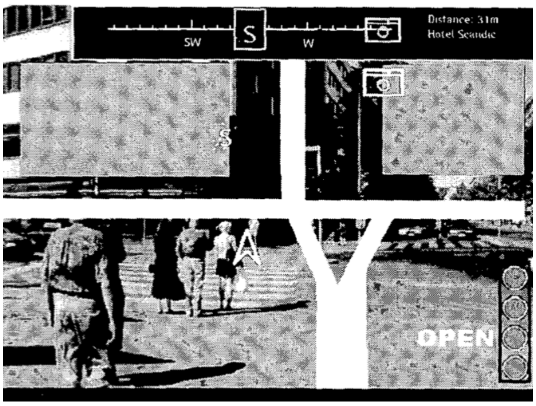


FIGURE 19: The original Context Compass as a bar on top

In contrary to the exocentric view, where the user is aware of everything going on around them, the egocentric view is concerned with what is happening in front of the user. To provide more spatial awareness, affordances must be made to alert users there are things happening off-screen. For example, the Halo UI [64] (Figure 20) features circles that "reach" into the display to inform there are objects of interest outside the screen. Another common method is the use of arrows to hint at off-screen elements. While these metaphors are usually used for POI in geospatial applications, they can also be applied to the panorama scenario. Instead of circles, a rectangular frame was introduced which represents the field of view of the remote user. Another commonly found element in photography was introduced as an egocentric cue: The centre point, indicating the centre of the remote users field of view.

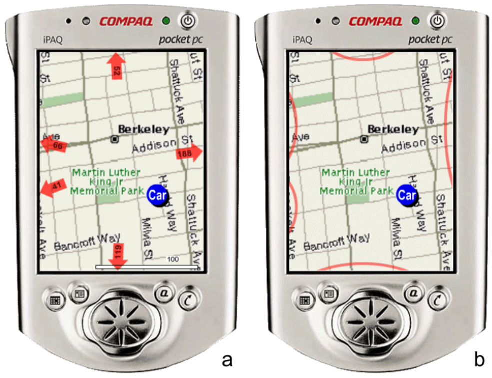


FIGURE 20: Off-screen hints as arrows (left) and Halo (right)

4.6 Pilot Study

To explore the extent that the various interface elements provide a sense of presence, an initial user study was conducted with four participants. The test conditions involved an exocentric awareness cue, egocentric awareness cue, and a condition that provided only audio with the panorama, but no further visual cues. The exocentric awareness cue used was a Radar element, providing a top down view of what each user was looking at. The radar display shows a triangle for each user that moves in a circular motion according to the orientation of the device they are using. The egocentric awareness cue was a rectangular frame and center point which would show each user's field of view and so would overlap when facing in the same direction.

All test subjects were members of the HIT Lab NZ. A panorama was pre-captured of the room where the Glass user was seated. A second user was sitting in a different room with a handheld tablet. Both users were connected via Wi-Fi and an audio connection. Pairs of subjects communicated using Glass or the Android tablet under one of the following three different conditions:

FIGURE 21: Panorama for the pilot study

- *Audio only* – Both users could view the panorama image but only talk about it.
- *Audio + Radar* – In addition to audio, a radar display was used to provide awareness.
- *Audio + View Frame + Centre Point* – The Frame is a 2D shaped box that represents the viewport of the remote user. The rectangle moves horizontally based on the remote user's orientation. If the entire rectangle is visible on the user screen then it means that both users are looking at the same part of the image. In addition, the centre point indicates the centre of the user's view. For the local user, it does not move, while the remote users will move depending on where they are looking. Aligned points show a complete overlap of both fields of view. To differentiate between the local user's and the remote user's centre points, they were drawn as an outline and filled circle respectively.

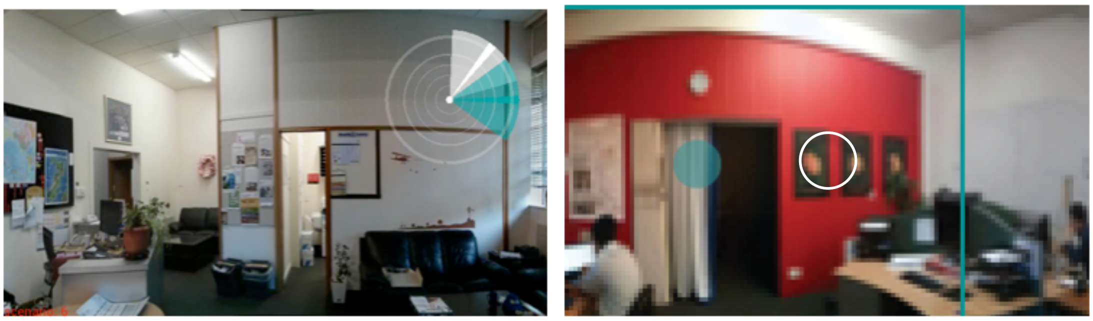


FIGURE 22: Screenshots from the exocentric cue (left) and egocentric cue (right)

The task was for each pair of subjects to discuss the room that the Glass user was in for two minutes, and answer a series of interior design questions, such as where they would place lights to best light the room. A within subjects study was chosen to have a direct comparison, so each pair experienced all three conditions with five different interior design questions. After each condition the following questions were asked to each participant to determine how well they felt they collaborated:

1. How easy was it to work with your partner?
2. How easily did your partner work with you?
3. How easy was it to be aware of what your partner was doing?
4. To what extent was this like you were in the same room with your partner?
5. To what extent did your partner seem "real"?

All questions were ranked on a scale 1 to 7, with Questions 1 to 3 ranging from *" 1 = Not Very Easy"* to *"7 = Very Easy",* Question 4 ranging from *"1 = Not like being in the same room at all"* to *"7 = A lot like being in the same room"*, and Question 5 ranging from *"1 = Not real at all"* to *"7 = Very real"*.

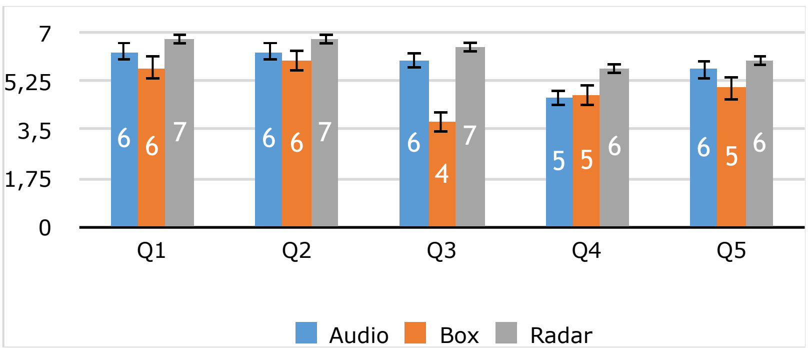


FIGURE 23: Median values for each question in the pilot study

Figure 23 shows the median values for each question. Since this was only a pilot study, there were not enough subjects for a detailed statistical analysis, but in general users felt that they had highest awareness of their partner in the Radar condition (Q3). They felt that this gave them continual awareness of where they their partner was looking and enabled them to easily align their views when needed. Subjects also commented that the view rectangle cue forced them to move around a lot to locate their partner, and was confusing. Some people felt that using audio only was better than using the view rectangle. However, the centre point was considered as useful providing a more specific insight on how to exactly align both views. In general the exocentric cue was used to have an overview, while the egocentric view was used to align in detail. Based on these results, it was decided that the UIs in the future experiments would feature both egocentric and exocentric cues. Additionally, the color of user interface elements for the local user was changed from white to red, as white was perceived as hard to see on the panorama background.

4.7 UI Concepts following Pilot Study

After analysis of commonly used cues and the pilot study, eight potential GUI concepts were created, as shown in Figure 24. There were 4 different conditions considered: Aligned view, a slight offset, almost out of view and entirely out of view. In all cases the colours of the interface were changed to a cooler hue depending on how far the centre point of the remote user was away from the centre of the local user. The egocentric and exocentric cues changed between designs, with the exception of the centre point, which was been mentioned as being very important during the pilot study by all participants.

UI Concepts 1 – 5 feature exocentric cues, with concepts 2 and 3 featuring variations of a radar metaphor, while concepts 2, 4 and 5 feature variations of a Context Compass metaphor. Concepts 6 – 8 only feature egocentric cues, with concepts 6 and 7 using a “Wi-Fi signal” style icon with increasing or decreasing “strength”, to show how close the viewpoint of the remote partner is. Concept 6 uses arrows to indicate the closest direction to the partner, and 2 shows the frame detailing how much can be seen of the other user’s point of view.

FIGURE 24: Concept interfaces after pilot study. See Section 4.5 for further information

As part of the user centric design, all of the concept UIs were presented to a group of potential users and a discussion was held. A major concern was scalability: While colours work nicely as an extra-cue between two people, it gets increasingly difficult to differentiate with more users. The same concern applies to the Wi-Fi signal and arrow hints, which work well as long as only two users are involved, but for collaborations involving more people screen clutter increases.

The final conclusion of the discussion was that variations of the Context Compass and Radar were regarded as best choices. The originally presented Radar would be reused, as well as the Centred Radar (Concept 3) and the rectangular Context Compass (Concept 4), which represented the field of view better than circular shapes.

4.8 Summary

In this Chapter, initial ideas, elaboration of decisions and development of different User Interfaces for awareness and interaction patterns were discussed. Some existing UIs from Augmented Reality applications were examined, however these were found to focus on presentation of geospatial information to one user only, and neglected collaborative elements. After a pilot study, conceptual UI brainstorming and user discussion, three UIs were chosen to be further explored: (1) A traditional Radar UI, which will be named “Classic”, (2) A centred Radar UI combined with centre point “Centred Radar”, and (3) a rectangular Context Compass.

5 Development of the Prototype

The following chapter presents the software developed. The concepts developed in the previous chapter are used to build a prototype running on a tablet and wearable device. The resulting work was evaluated in several user studies, which are presented in the following chapter.

While the prototype was not developed to the stage ready for commercial release, it was good enough to explore the research questions on how to evaluate the UIs for awareness and the effects of interaction for Social Presence.

5.1 Software

In order to explore how panoramas and wearable computing can be used to support real-time collaboration, a prototype system was developed that connected a person using Google Glass with a second user on an Android tablet. The overall goal of the research was to develop a system that will allow a user to capture a panorama of the space around them and then share it in real time with a remote collaborator. As the current iteration of the Google Glass has limited processing power, it is currently not feasible to perform real-time panorama stitching. So in the initial prototype it is assumed that the panorama has been captured and research focused on supporting remote collaboration and interaction.

The system requires the following properties for both devices:

- Establishing a network connection via Wi-Fi for data exchange
- Displaying the panorama image at the proper resolution for each device
- Looking around the panorama by rotating the head or the tablet device
- Display the UI
- Allow for touch input to support pointing or drawing
- Rendering the drawing or pointing cursor onto the panorama cylinder

The current prototype has been developed using Processing [65] with the Ketai library [66] for sensor support. Processing offers a quick way in developing prototypes, specifically for visual applications.

The application was compiled into an Android APK file, which is then pushed to run on the target device. The code was split into two separate code bases; one for Glass and one for Tablet. The code is mostly identical between the Glass version and the tablet version, except for the touch interaction and network connection components. Figure 25 illustrates the overall flow diagram of the system components.

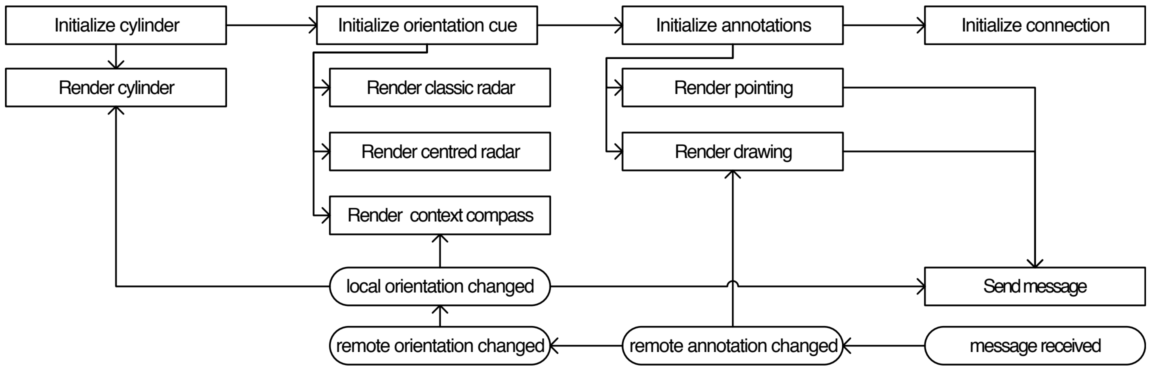


FIGURE 25: Overall flow diagram

5.1.1 Rendering the Panorama

The system uses a previously captured panorama image mapped onto a virtual cylinder to simulate a live capture system. To display the panoramic image to the user, a cylindrical 3D model is rendered with a 32 edge QUAD_STRIP shape definition from Processing. Once the cylinder is created, it is textured with the desired panorama image. The height of the cylinder is set to match the height of the image, and the width and radius were calculated to maintain the aspect ratio of the image.

The cylinder is rotated to match the orientation of the device (i.e. the user's view). On the Glass device the user can turn their head and the integrated orientation sensor is used rotate the panorama cylinder accordingly. On the tablet the tablet orientation is used to rotate the panorama. Currently only rotations around the vertical axis are supported.

5.1.2 Interface

Based on the work described in Chapter 4 and the outcome of the resulting pilot study, several cues were chosen to be implemented into interfaces for the first experiment to serve as different experimental conditions. These interfaces are labeled as “Classic”, “Centered Radar” and “Context Compass”. Most importantly, all implemented interfaces for further experiments now include exocentric cues and a centre point as an egocentric cue in all interface designs combined.

Classic

The Classic interface features the exocentric Radar metaphor commonly used in AR applications. Located at the right corner of the screen, the user’s field of view is displayed as a triangular shape on the circular radar. In addition to the radar, an egocentric box was introduced, which shows the exact field of view of the other user.

Centred Radar

The Centred Radar is an original interface also built using the Radar metaphor. Based on the assumption that a separate centre point and radar would take up too much space on the small display of the Google Glass, they were merged to form a cue that was both egocentric and exocentric.

Context Compass

The Context Compass appears as a box on a line on top of the screen. The box represents the current view of the user within the panoramic image, which is represented by the line. The remote user’s view is displayed with a different coloured box — once both boxes are aligned the users are looking at the same direction. The box moves only horizontally.

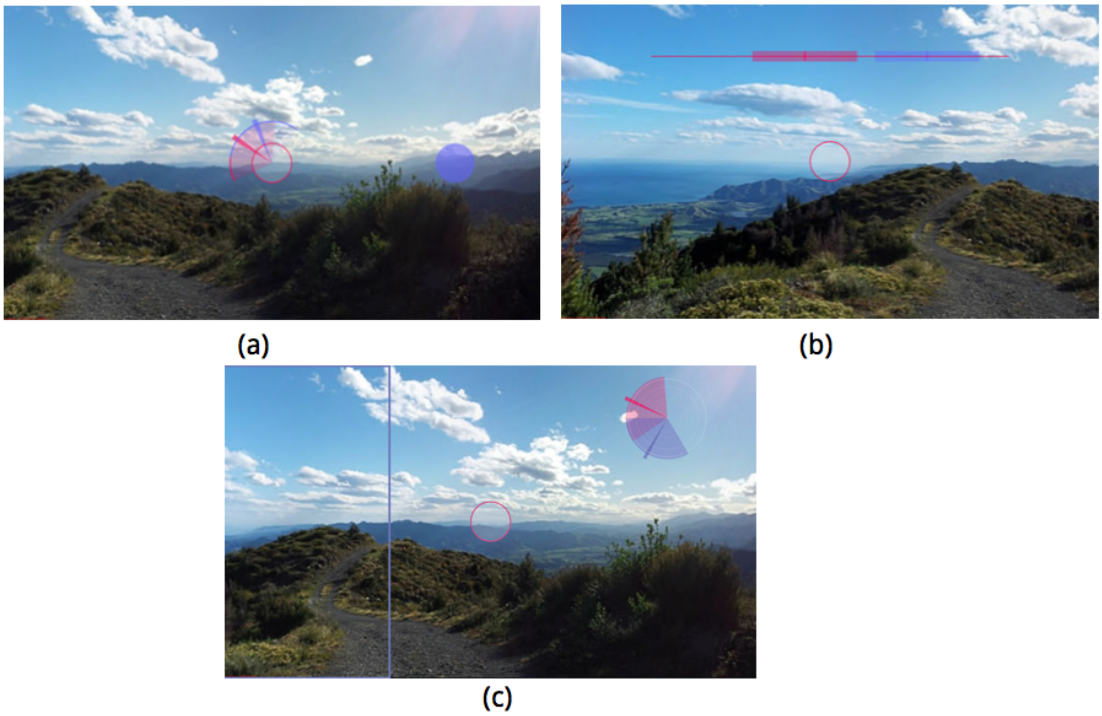


FIGURE 26: Chosen interface elements.
(a) Centred Radar with Centre Point on the right (b) Context Compass (c) Classic interface with the Radar on the right and the Frame can be seen on the left

5.1.3 Interaction and Touch Input

In remote collaboration it is important that participants are able to refer to surrounding objects and the environment, which we wanted to support in the form of pointing or drawing. The aim was for pointing to mimic using a mouse on a computer and drawing to mimic traditional sketching with paper and pencil, providing communication that feels natural. Drawing and pointing are two modes of interaction that have previously been shown to be good gesture surrogates [27].

Using Google Glass, users are able to touch the touchpad to point or draw on the panorama. The user can tap on the touchpad with two fingers to swap between pointing and drawing modes. Once selected, the pointing mode maps the x,y position of the user's finger on the touch pad to the corresponding position on the region of the panorama in view. As they move their finger on the touch pad a virtual pointer moves on the panorama. In a similar way, when they are in drawing mode, touching the touch pad will draw lines on the panorama that remain until they are erased. Any pointing or drawing points are sent wirelessly to the tablet so that the remote user can see the Glass users input. In a similar way the tablet user can touch and draw on the tablet and have their annotations appear in a different colour on the panorama.

The algorithm for calculating where a point on the panorama falls in the user's view is as follows:

1. Calculate the width and height of the part of the image that is visible on the screen (FOVw, FOVh)

```
FOVwidth = cylinder-radius * FOV
```

2. For a point (x0, y0) at the orientation (o).
3. Map point (x0, y0) of the screen to point (x, y) of the full panorama image.
4. offset x by the orientation
5. calculate point (px, py, pz)

```
px = cylinder-radius * cos(x)
py = y
pz = cylinder-radius * sin(x)
```

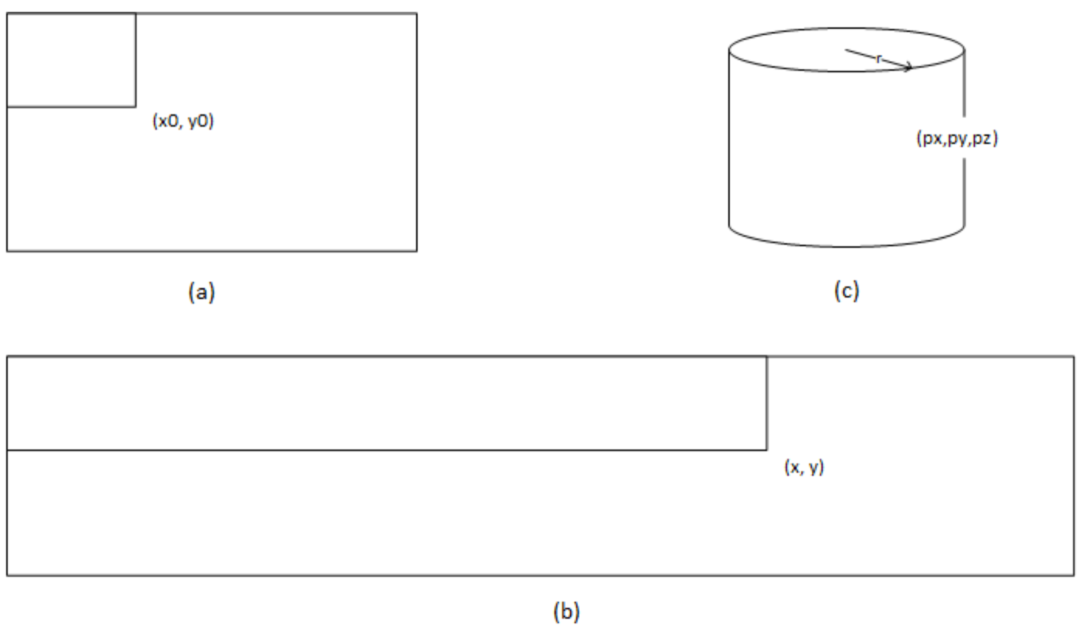


FIGURE 27: Mapping procedure:
a) represents a point (x0, y0) with respect to Field of View (FOV) width and height. (b) represents the point (x, y) with respect to the full panorama image width and height (c) represents the 3D point (px, py, pz) on the cylindrical space

The touch input was designed to help the user during pointing and drawings tasks. The initial default input is set to pointing. This helps the Glass user to begin the drawing at the right place by using the pointing to locate the start point of the drawing, then using a two-finger tap gesture to switch to Drawing mode. Once the drawing is completed, the user can switch back to the Pointing mode again by using a two-finger tap gesture. If the user wishes to clear the drawing, they can use a three-finger tap gesture to clear the existing drawing.

The application listens to user touch input on the screen (tablet) or on the touch panel (Glass) with the following functions.

```
//glass
public boolean dispatchGenericMotionEvent(MotionEvent event)

//tablet
public boolean surfaceTouchEvent(MotionEvent event)
```

The application also detects the number of fingers touched and how long the touch is applied. The Ketai touch library is used to detect a long press touch on the tablet device.

Touch input supports the following modes:

- single finger: add drawing point or update pointing location based on the current touch mode
- two fingers: toggle current touch mode (drawing or pointing)
- three fingers with long press (> 500 ms): clear drawing points

The drawing points and the pointing location are recorded in memory, and used when rendering the screen. The drawing points are represented by two arrays; one for local drawing points, and one for remote drawing points. The pointing location is represented by two variables; remote and local.

```
//drawing
ArrayList points = new ArrayList();
ArrayList remotePoints = new ArrayList();

//pointing
Point pointingPoint = new Point();
Point remotePointingPoint = new Point();
```

The location of the drawing points and the pointing are sent to the remote user via the Connection component.

FIGURE 28: Two implemented interaction modalities
Two pointers on the left with an arrow icon on the right lower corner indicating to the user to be in the pointing mode, as well as a pen icon on the right picture for the drawing mode

If the user decides to clear the drawing using the 3-fingers touch input, then the corresponding drawing points at the remote device need to be cleared as well. This is accomplished using the clear drawing command. The format of the clear command is a single string "CLEAR". Once the clear command is received from the remote device, the `remotePoints` array is cleared.

5.1.4 Connection and Networking

The tablet and Glass applications communicate over a Wi-Fi network, enabling data to be shared between them. The communication runs over TCP/IP, using the oscP5 networking library [67]. The Glass device acts as a server, listening for connections, while the tablet device attempts to connect to the Glass using its IP address and a known port number. Once the connection is established, both devices start sending their local orientation to the remote device every half second.

The format of each orientation message consists of:

- x, y, z: representing the orientation around x-axis, y-axis, and z-axis respectively

The connection component also updates the location of drawing and pointing from and to the remote device. The format of each collaboration message consists of:

- touch mode: drawing/pointing
- a, b: representing x and y of previous touch point
- x, y: representing x and y of the current touch point
- o: representing the orientation at which the touch point was recorded

5.1.5 Orientation

Once the connection is established, both devices start sending their local orientation to the remote device every 500 milliseconds. The format of each orientation message consists of: x, y, z representing the orientation around x-axis, y-axis, and z-axis respectively

5.2 Summary

This chapter describes the specific components of the developed prototypes for the Tablet and HMD platforms. It explains the requirements and choices made for the Head Mounted Display and tablet devices. The software was developed with Processing, and used the Ketai library for touch input and oscP5 library for networking. A panorama image is rendered onto a cylinder, and two participants can interact in real-time using pointing or annotation methods, while different UI features help with awareness of what the other user is doing. These UIs are used for the first experiment described in the upcoming chapter.

6 User Evaluation

In this research, two user experiments were conducted. The first experiment aimed to compare different tools that could provide a user with more awareness about their collaborator's situation, while the second experiment aimed to evaluate pointing and drawing tools and their impact on Social Presence.

As explained in Chapter 3, awareness plays a crucial part in establishing presence. For this reason, the results of the first experiment on awareness informed the design of the second experiment on Social Presence. For this reason, the environment setup and subjects stayed largely the same across the two experiments, and a single explanation of these are provided for both experiments. Following this each experiment is described individually. The quantitative and qualitative results obtained are discussed in the following chapter.

The experiments aimed to address these questions:

(1) What cues are important for awareness in Social Panoramas?

(2) Can drawing and pointing cues increase Social Presence in Social Panoramas?

The hypothesis was that visual cues do increase awareness, and that annotation and pointing interaction increase Social Presence. We assume that visual cues increase awareness by providing more information of the other users actions, and previous research has shown a positive influence of interactivity on presence [6].

6.1 Environment Setup

Both experiments were conducted in the same environment. The system was setup to simulate a real-time collaboration between a tablet user and a Glass user. The study was conducted at the HIT Lab NZ, a typical office environment, and the two participants were placed in different rooms. The choice of rooms stayed consistent during both experiments. A panorama of the office where the Glass user would be sitting was previously recorded using Android Photo Sphere [12]. The scenario that participants were given was that the Glass wearer had just captured a panorama on their device, and would like to discuss the image with the tablet user, who was remotely located. The tasks that the users had to complete depended on the experiment. Figure 29 shows the floorplans of both rooms, with the panorama user placed in the middle.

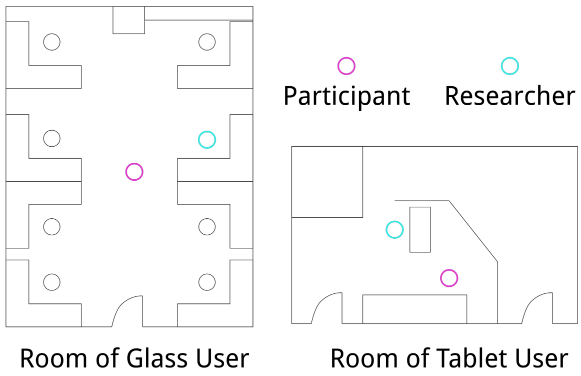


FIGURE 29: Floorplan of each room

In Experiment One, users were tasked with searching for objects in the room of the Glass user. In order to facilitate the search for objects within a repetitive office environment, several additional objects, such as sport equipment and picture frames, were placed inside the room. While the location where the tablet user was sitting was not of importance, the Glass user was seated at the exact location and approximate height of where the panorama picture was taken to maximize the overlap between the panorama and the real surroundings.

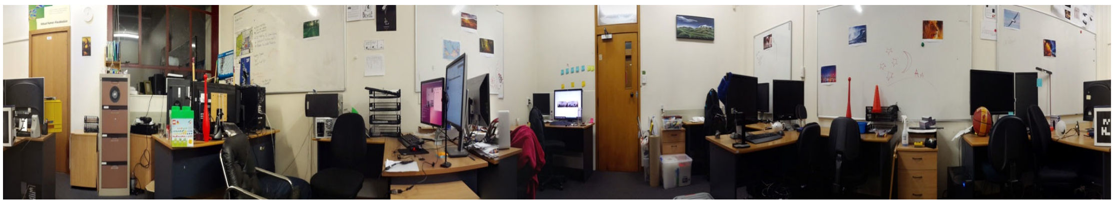


FIGURE 30: Panorama used for the experiment

To facilitate audio communication, both users were provided a headset with two earpieces and a microphone. The audio connection was established using voice over IP. All sessions were recorded on video via webcams in both rooms.

6.2 Subjects

Both experiments had 24 participants aged between 18 and 45, and subjects from the first experiment were not included in the second, resulting in 48 different participants in total for both studies. Gender was equally distributed in both cases. Gender equality played an important role for the second experiment as presence is perceived differently by gender [68]. Students and University personnel were recruited, as well as members of the public. Subjects did not know each other prior to the experiment and collaborated in pairs of their own gender.

Using only gender specific groups was chosen due to the fact that two different mediums were being used and we wanted to avoid any biases based on gender.

6.3 Experiment One: Awareness

This section describes the experiment aiming to explore what effect different visual cues would have on awareness, based on the interfaces developed as a result of an earlier pilot study. The pilot study suggested the use of both exocentric and egocentric cues creates a higher level of awareness, and as such all interfaces included both types of cues, with the exocentric cue changing between conditions.

6.3.1 Experimental Design

Subjects were asked to perform a collaborative task during the experiment. They were placed in two different rooms and did not know where their collaborative partner was located. One user would be using a head mounted display, while the other user would be using a tablet. Both users could see an identical panoramic image on their device. The panorama image was taken beforehand with the Photo Sphere stitching application. The panoramic image was a panorama of the room where the Glass user was located – to mimic the scenario that the glass user had just captured a panorama and wants to collaborate with the tablet user.

There were four experimental conditions in the study that depended on the type of awareness interface the participant was using. A within-subject design was used, and all participants had to go through four conditions. The order of the conditions was counter balanced using a Latin square technique. For collaboration, participants were connected with a microphone and a headset.

The first experiment consisted of following conditions:

- *Audio* – Only the panorama image was available. No further visual cues were implemented.
- *Centred Radar (CR)* – A radar implemented in the middle of the screen that indicates the users own view point and the remote users viewpoint in a different colour.
- *Context Compass (CC)* – A horizontal line on the upper screen with a red rectangle representing a user's own viewpoint, and a blue rectangle representing the other user's viewpoint.
- *Classic Interface* – A radar in the upper right corner using the same principle as the Centred Radar. A frame indicates the viewpoint of the second user.

The interfaces are explained in more detail in Section 5.1.2, *Interfaces*.

6.3.2 Procedure

The participants were explained the experiment together in a meeting room and informed about the procedure. All conditions were explained at this point, and were explained a second time once the participants were assigned to a device and in their respective environment. Questions could be asked at any point during the introduction. Once the participants confirmed they were ready, they were taken to their room and introduced to their device. For both participants there was a researcher to provide assistance.

Before the start of the task, each participant was asked to fill out a questionnaire asking basic demographic questions, such as age, gender, perceived level of English and experience with tablets and head mounted displays. After the questionnaire they were introduced to the interface of their first condition and given some time to get familiar with it. On both sites the sessions were recorded on video for further analysis. After each condition the participant filled out a questionnaire asking questions about collaboration, presence and perceived task performance. After all the conditions, they were asked to give any comments they might have and rank the conditions from favourite to least favourite based on four terms (4 = favourite, 1 = least favourite): Awareness, Collaboration, Ease of Use and Presence.

6.3.3 Scenario and Task

The first experiment's task involved the users taking turns choosing an object in the panorama, and then giving clues about the object in question, such as its colour, shape or surrounding objects to their partner for them to guess the object they had chosen. Once an object was correctly identified, the users would switch roles of who choose an object and who guessed what it was. Users were instructed that after each turn they needed to pick an object in another part of the room to make use of the awareness interface and not choose objects in the same location. Time was limited to two minutes per condition.

It was specifically explained by the researchers that it was not a competitive task and that the participants were to work together, as a previously conducted pilot study revealed that participants would start picking tiny objects making it increasingly hard for the user to find the object in question. This resulted in an uneven competition, where the Glass user had a significant advantage being seated in the actual room and thus could see small details not visible on the tablet user's display.

6.3.4 Dependent Variables

To measure how much the user was aware of their partner's actions, we asked users to fill out the following survey. The questions are presented in detail in Section 7.1.1, *Quantitative Results*.

Collaboration

Three questions were asked to determine the collaboration between the participants. They were measured on a 7-point Likert scale. The questionnaire was based on one used in previous studies on remote mobile collaboration [69].

Presence

Questions were based on a previous questionnaire for physical presence ("sense of being there") in virtual environments [70]. As the original questionnaire targeted virtual environments, we only used the 3 questions out of the 13 that we found suitable for our case. Presence was measured on a 5-point Likert scale.

Perceived Task Performance

Three questions on a 7-point Likert scale were asked to rate the subjects perceived task performance, based on a survey previously used in [46]. In this questionnaire the user could evaluate their own task performance while using the system.

In addition to the survey questions, we measured the number of turns completed and the difference in head orientation between users:

Number of Turns

During the experiment both researchers were taking notes on the number of turns that were made during the tasks, which was later verified by video analysis. A turn was defined as a user finding an object, the other user confirming it was right and then switching roles, starting a new round.

Head Orientation Difference

The difference in head orientation between the Glass user and the Tablet user represents whether both users are looking at the same part of the panorama and how far they are from each other. The head orientation results were measured using log files that recorded the orientation of the device every half second.

6.4 Experiment Two: Social Presence

Building on the first experiment, the goal of the second experiment was to explore if interaction in the panorama would increase the users' sense of Social Presence. Pointing and drawing tools were enabled in the software for both devices.

6.4.1 Pilot Study

Both users were given an identical list of items of furniture. There was a total of four lists, differing in artistic styles and objects that would be switched through each task. The list always contained pictures of six objects (Figure 31) and users were asked to place the objects collaboratively inside the room to simulate a realistic location. The furniture photographs provided were a mixture of more unusual furniture (such as baroque style furniture and a 2 metre tall mirror) and more common objects. Most of the photographs provided approximate sizes and dimensions, while some of the more common objects were shown without measurements.

The idea of the second experiment was for the users to collaborate and discuss the locations in a realistic manner, using the interaction techniques provided for pointing and sketching. In the pilot study it turned out that one participant would often take the lead and choose the location of all the objects, rarely using the interaction tools provided. The more passive collaborator would simply agree and add one or two comments. To compensate for this flaw during the pilot study some things were changed for the final experiment: In the final experiment the list of objects was split in half, giving one half to each participant. The task of placing the objects remained the same, but with only one participant knowing what the object looked like, each participant was motivated to elaborate on what the object looked like and use the interaction methods in order to explain the dimensions. To further enforce collaboration, the user without the image of the object was the one who was tasked to find a place for the object. This resulted in more conversation, as the "blind" user had to know size and shape in order to find a realistic spot for the furniture.

FIGURE 31: Furniture examples:
"Contemporary" (left), "Baroque" (middle) and "Minimalism" (right)

6.4.2 Experimental Design

As with the first study, the experiment involved a collaborative task between two participants in different rooms using Glass and Tablet devices. Both subjects could see a panorama, which was the environment in which the Glass user was placed.

Although the pilot study did not show any awareness interface being significantly better, the Context Compass was chosen as the interface based on comments by previous participants that it provided a clear view of the panorama with little distraction by navigation tools.

Designed as a within-subject study all participants had to go through all four conditions that were counter balanced using a Latin square technique. In this experiment, the conditions varied in terms of interaction:

Audio – Only the panorama image and the awareness interface were visible.
Drawing – Both users could draw shapes on the panorama using the touch surface of the tablet or the touchpad on the Glass device. Drawing was done with one finger. Lifting the finger and touching again would result in a new shape being drawn.

Pointing – Similar to the drawing condition, but instead of drawing a continuous line only a pointer in the shape of an arrow would be visible and be used in a sense of pointing to objects or locations. The arrow would always be visible on screen even if not in use.

Pointing and Drawing (P+D) – The user is able to switch between the interaction modes. An icon on the lower side of the screen would indicate the mode, showing an arrow for pointing mode or a pen for drawing mode. The current mode could be switched by tapping two fingers on the device.

The default mode was set to pointing to facilitate orientation on the touchpad for the Glass user. Ideally, the pointer would be used to choose where the user would like to begin drawing, then the user would switch to drawing to create the desired shape or object, and then switch back to pointing.

When drawing, if the user wished to clear the drawings on his screen, a 3-fingered tap erased the existing shapes. Erasing the drawings on one device would also erase the drawings on the remote device.

6.4.3 Procedure

The procedure was the same as described for Experiment One, introducing the experiment to both participants and guiding them to their room. However, this time there was a more extensive training session once they were seated and equipped. Before each task there was a short training on how to use the interaction modes correctly. After they confirmed being comfortable the main task would start. After the experiment they were asked to fill out a questionnaire on Presence, Collaboration and Usability. Additionally they were asked to comment on whether they enjoyed using pointing or drawing more, and for which tasks they thought it more appropriate. Lastly they were asked to rank the conditions based on Presence and Interaction. The researchers took notes during the experiment. During the experiment time stamped data about user interaction and head orientation was logged by the system. The logged data allowed for the interaction to be reconstructed for later analysis.

6.4.4 Scenario and Task

In the second experiment, the task to be completed was a collaborative furniture placing task. Both users were given a list of furniture objects where they could see the names of all the objects (e.g., “Mirror”), but each user could only see the pictures of the objects for one half of the list each. The subject with the picture was asked to describe the object to the user without the picture, while the user without the picture was asked to find a suitable place inside the room. The user with the picture could then confirm or deny if the location was realistic. Once both participants had agreed on a location they would switch roles and move on to the next object on the list.

There were a total of four lists, one for each condition, which differed in artistic styles and objects. Each list contained six objects that the users had to place inside the room in a realistic location. Participants were given a maximum of three minutes to complete the task, although it was emphasized that their performance was not being measured, and that it was more important that they worked together to find the optimal solution for each object.

FIGURE 32: Participants during the experiment

6.4.5 Dependent Variable

To measure the user's perceived presence, performance and interface usability, subjects were asked to fill out surveys on Social Presence, Task Performance and System Usability. The questions are presented in detail in Section 7.2.1, *Quantitative Results*.

Social Presence

The social presence questionnaire consisted of 8 questions on a 7-point Likert scale. The questions designed to measure the "Sense of Being Together" [71].

Perceived Task Performance

For Perceived Task Performance, the same questionnaire as for the Awareness experiment was used, however while the awareness experiment only used 3 out of 4 questions, the fourth question was included in this experiment as it was considered an explicit interaction task.

Usability

Usability was tested with The System Usability Scale (SUS). SUS was developed by Brooke in 1996 as a tool to quickly measure the subjective rating of usability of an interface or product, and was shown to be highly robust and flexible enough to assess different technologies [72]. The method produces one single score, which allows it to be easily interpreted to give an estimate of the usability of a system.

The SUS questionnaire contains 10 statements where the user can choose his level of agreement on a five-point Likert scale ranging from "Strongly agree" to "Strongly disagree". The statements switch between positive and negative, and there is a clear guideline on how to measure the score of a SUS properly [73]. The score of the SUS ranges from 0 to 100 and provides a point estimate of percentage. Acceptable scores start at 70, with scores over 90 indicating a highly usable product. Scores under 70 should be seen with concern and be subject for further improvement [72].

6.5 Summary

Two user studies were conducted during the course of the project. The first experiment aimed to find the best interface for awareness as awareness, which is deemed important for Social Presence. User's were seated in different rooms and asked to look for different objects that should be explained to the other user, in order for him or her to guess the right object.
The second study introduced interaction to the panorama, and the experiment focused on the impact of interaction on Social Presence. Participants were asked to rearrange a room together with furniture designs that were given to them.

7 Results

In this chapter, the results from Experiment One and Experiment Two are presented individually. Both sections follow the same structure: First, the quantitative results are presented, including questionnaire results and logged data such as head orientation and pointing and drawing data. Following this, observations and qualitative feedback is provided. Data is shown as tabular form and bar graph form, where whiskers represent Standard Error. The chapter concludes with a summary of the results.

7.1 Experiment One

All survey questions were analysed with a non-parametric test, namely the Friedman-test. Post hoc analysis was further conducted with a Wilcoxon signed-rank test with Bonferroni correction applied, resulting in a significance level set at $p<0.0083$.

7.1.1 Quantitative Results

Collaboration

Subjects were asked the following questions:

1. How easy was it to work with your partner?
2. How easily did your partner work with you?
3. How easy was it to be aware of what your partner was doing?

The results are shown in Figure 33.

Glass

Generally, Audio was rated the worst interface type for all three questions. There was a statistically significant difference in all three collaboration questions Q1 to Q3 (Q1: ($\chi2(3)=13.500$, $p = 0.004$; Q2: $\chi2(3)=12.808$, $p = 0.005$; Q3 $\chi2(3)=15.894$, $p = 0.001$)). For Q1 there was a significant difference between the Classic – Audio condition (Z= -2,648, $p = 0.008$). Q2 and Q3 did not show significant results with the Bonferroni correction applied. However, Q2 showed a trend towards Classic – Audio (Z= -2,414, $p = 0.16$), while Q3 tended towards the three visual interaction types compared to Audio: Classic – Audio (Z= -2,612, $p = 0.009$), Context Compass – Audio (Z= -2,553, $p = 0.1$) and Centred Radar – Audio. (Z = -2,491, $p = 0.013$). There was no difference between the visual cue conditions.

Tablet

All collaboration questions showed a statistically significant difference. Similar to Glass the Audio condition scored the lowest on the tablet. For Q3 the Audio case was rated significantly worse than all of the visual conditions. (Q1: $\chi2(3)=7.888$, $p = 0.048$; Q2: $\chi2(3)=10.642$, $p=0.014$; Q3 $\chi2(3)=16.590$, $p = 0.001$). Post hoc analysis showed a significant difference for Q3: Classic – Audio ($Z= -2{,}855$, $p = 0.004$), Centred Radar – Audio ($Z= -2.825$, $p = 0.005$). Context Compass – Audio came close with ($Z= -2.609$, $p = 0.009$). There was no difference between the visual cue conditions.

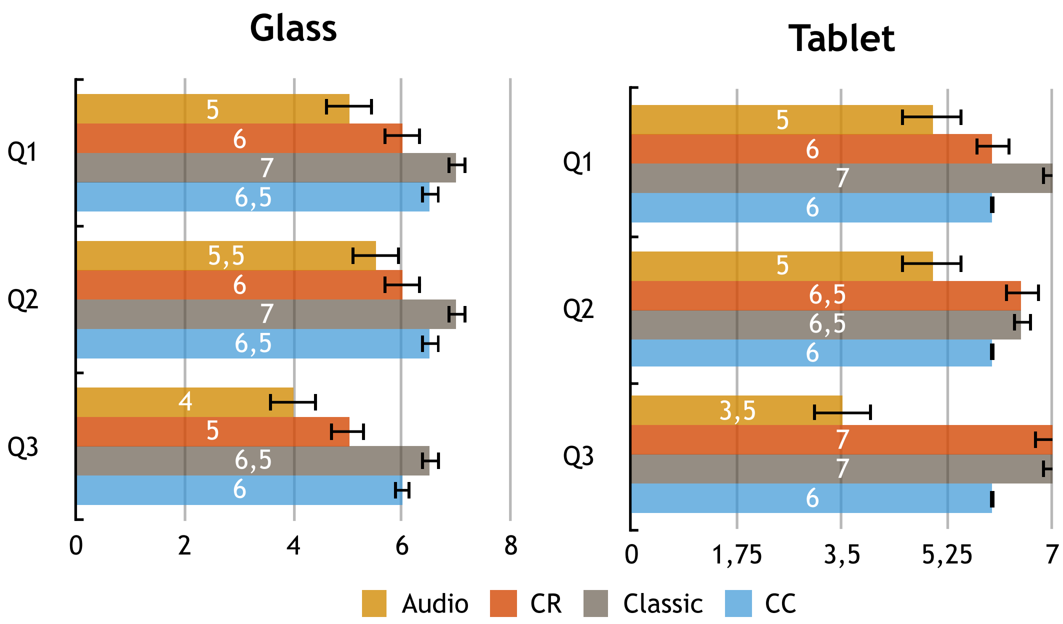


FIGURE 33: Median scores of the questions 1–3 on Collaboration. Glass (left) and Tablet (right)

Presence

Subjects were asked the following questions:

4. How strong was your sense of presence in the panorama view?
5. How strong was your sense of "being there" in the panorama view?
6. How easily were you able to look around the panorama view?

There was no statistical difference found on either medium, or across any of the questions. The results are shown in Figure 34.

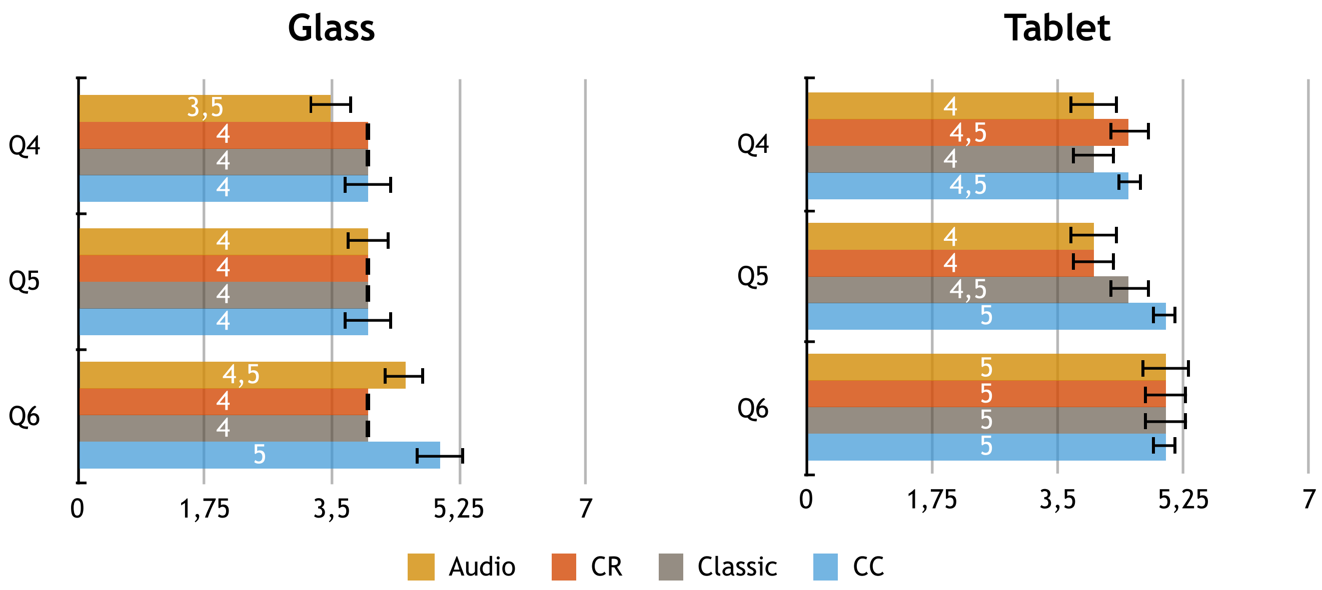


FIGURE 34: Median scores for the questions 4–6 for Presence. Glass (left) and Tablet (right)

Perceived Task Performance

Subjects were asked the following questions:

7. How easy was it to be aware of what your partner was doing?
8. How do you think that you managed to do the task in the system?
9. How efficiently did you feel that you managed to do the task?

Figure 35 shows the median for each of these questions. In Q7 and Q9 there was a significant difference on Glass (Q7: $\chi2(3)$=8,935, p = 0.03 ; Q9: $\chi2(3)$=11,118, p = 0.011), however post hoc analysis did not show any significant difference between the conditions. A trend was visible for Q9 between all visual cues and Audio: Context Compass and Audio (*Z*= -2,555, *p*= 0.01), Centred Radar and Audio (*Z*= -2,360, *p*= 0.18) and Centred Radar and Audio. (*Z*= -2,223, *p*= 0.026). On the tablet, Q8 showed a statistically significant difference ($\chi2(3)$=10,121, p = 0.018), however the post hoc test did not show a significant difference. However, it came close between Audio and Centred Radar (*Z*= -2,342, *p* = 0.02) and Audio and Classic (*Z*= -2,572, *p* = 0.01), and the first two visual cues competing Classic and Centred Radar (*Z*= -2,333, *p* = 0.02).

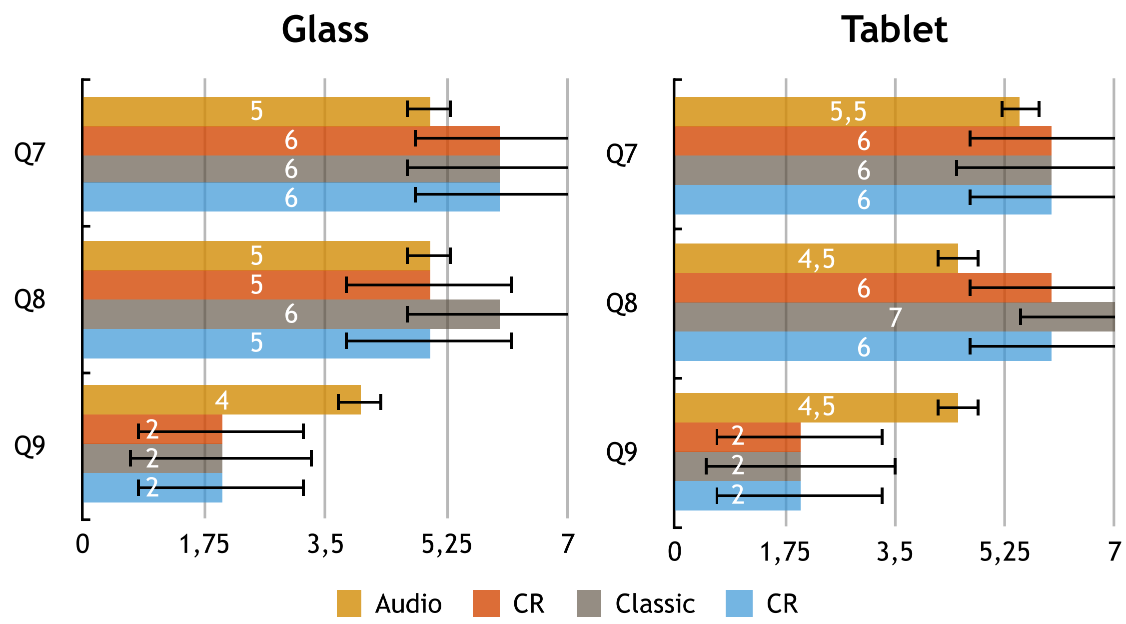


FIGURE 35: Median scores on questions 7–9
for Perceived Task Performance. Glass (left) and Tablet (right)

Ranking

Users were asked to rank each of the conditions in order in terms of Ease of Use, Awareness, Presence and Collaboration. Figure 36 and 37 below show the median (1 = worst, 4 = best). In all cases the Audio only condition was ranked lowest and there was no difference between the visual cues.

Glass

The ranking was as follows for all for terms. Interfaces that received the same median result are presented with their average score in brackets: Classic (Mean on Presence: *2,92*), Context Compass (Mean on Presence: *2,59* and Ease of Use: *2,83*), Centred Radar (Mean on Ease of Use: *2,67*) and Audio. There was a statistically significant difference in the Awareness ranking ($\chi2(3)$=16.3, $p = 0.001$), Collaboration ranking ($\chi2(3)$=17.417, $p = 0.001$) and Ease of Use ranking ($\chi2(3)$=16.3, $p = 0.001$). For the awareness ranking there was a significant difference between Context Compass – Audio (Z=-2.87, $p = 0.004$) and Classic – Audio (Z = -2.756, $p = 0.006$). Similar results showed the ranking for Collaboration and Ease of Use: Context Compass – Audio (Z=-2.87, $p = 0.004$) and Classic – Audio (Z=-2.858, $p = 0.004$) for Collaboration and Context Compass – Audio (Z=-2.832, $p = 0.005$) and Classic – Audio (Z=-2.756, $p = 0.006$) for Ease of Use.

Tablet

The results for the tablet device had the same ranking as those on the Glass, except for Ease of Use, where Centred Radar was ranked second and Context Compass was ranked third. Similar to the results for the Glass device, the Audio condition ranked significantly lower than some of the visual conditions in terms of Awareness, Collaboration and Ease of Use. There was a statistically significant difference in the Awareness ($\chi 2(3)$=15.4, p=0.002), Collaboration ($\chi 2(3)$= 11.5, p = 0.009) and Ease of Use ($\chi 2(3)$= 17.1, p = 0.001) ranking. For the Awareness ranking there was a significant difference between Centred Radar – Audio (Z=-3.14 and p = 0.002). Significant results were also found for Ease of Use between Centred Radar – Audio (Z=-2.818, p = 0.005) and Classic – Audio (Z=-2.806, p = 0.005) in favour for the visual cues.

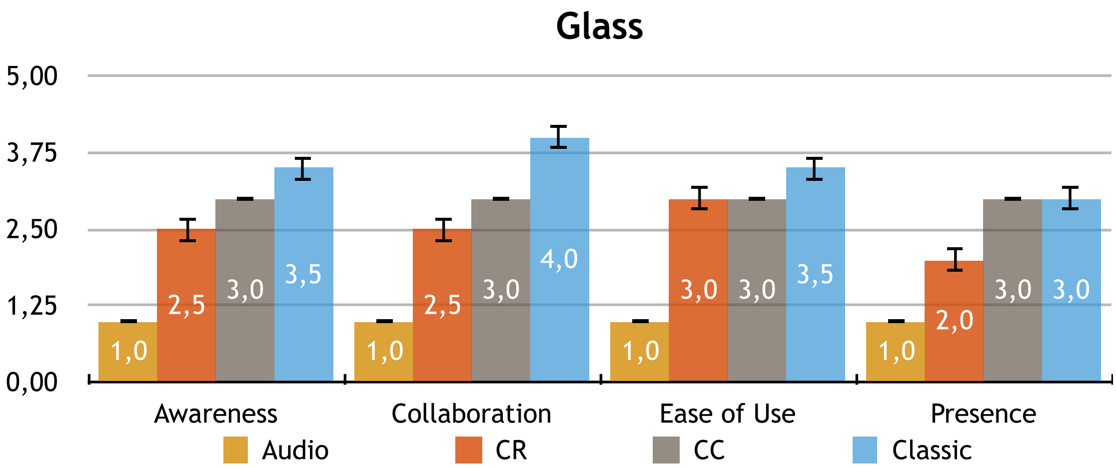


FIGURE 36: Median ranking scores for Glass
on Awareness, Collaboration, Ease of Use and Presence

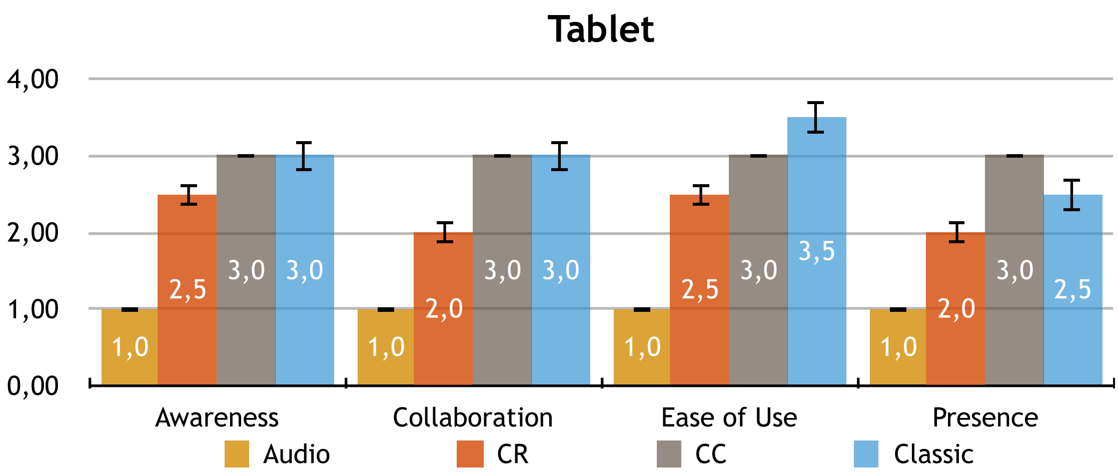


FIGURE 37: Median ranking scores for Tablet
on Awareness, Collaboration, Ease of Use and Presence

Turn-Taking

Figure 38 shows the average number of turns completed for each task. Users completed fewer turns in the Audio only condition compared to the visual conditions, but there was no difference between the turns-taken in the visual conditions. There was a statistically significant difference in the amounts of turns taken per condition. ($\chi 2(3) = 12.467$, $p = 0.006$). There was a statistically significant result in the number of turn-taken for Context Compass – Audio ($Z = -2.698$, $p = 0.007$) possibly due to faster alignment of orientation between the users.

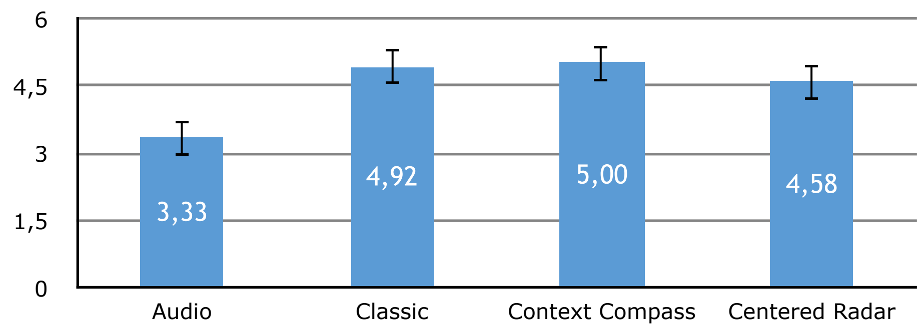


FIGURE 38: Average scores for turns completed

Head Orientation

Figure 39 shows the average difference in head orientation between the Glass and Tablet users during the experiment. A repeated measures ANOVA with a Greenhouse-Geisser correction determined that the difference in mean head orientation was statistically significant between the conditions, ($F(1.501, 21.017) = 21.032$, $P < 0.042$) however the post-hoc test did not indicate where the significance was. Without the Bonferroni test there would be a significant difference between all Audio – Visual Cue conditions. (Audio – CR ($p = 0.22$); Audio – Classic ($p = 0.047$), Audio – Compass ($p = 0.045$).

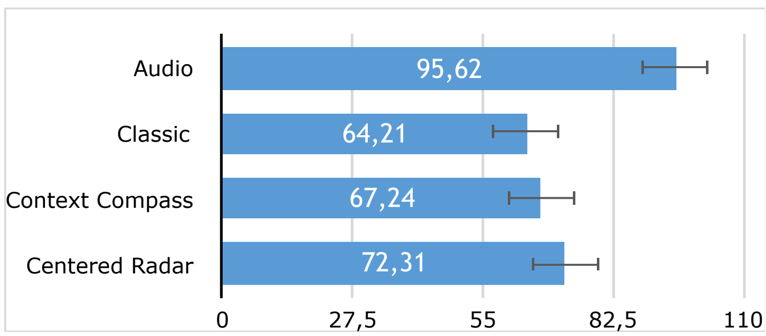


FIGURE 39: Average differences in head orientation (in degrees)

7.1.2 Qualitative Results

Observation

Participants exhibited a wide variety of different behaviours during the experiment. The descriptions of the scene that were used to achieve common ground were highly visually oriented. Some of the Glass users were looking and describing objects that they could see in the real world rather than the image displayed on Glass. Since the Glass display is located at the upper right corner of the eye there was some minor misalignment between what they were seeing and their head orientation as was shown to the remote user. Other users, however, only used the panoramic display without looking at the real environment, and thus the misalignment did not have any effect.

A few users used hand gestures when describing the shapes or to point at a location while describing the object verbally to the remote user, even though these hand gestures were not visible to the remote user. Some users asked questions about the shape or the name of the object. When using the audio only condition, some users developed strategies such as describing their orientation using the hands of a clock as a metaphor. Table 2 below shows some of user's speech from the experiment. Confirming the correct object was important for turn-taking, and patterns of using "reference points" were utilised for orientation inside the image.

The panoramic image used maps the room onto a 360 degree cylinder. Due to this the image doesn't rotate up or down when the user looks up and down. Different users held the tablet device differently; some held it in front of their face at a 90 degree angle, while others held the device at a 45 degree angle when they discovered moving the device up and down didn't rotate the panorama.

User's device	Time(s)	User's speech	Comments
Glass, Tablet	6	That's right I can see it. Yes, you found it! I can see… can you see…? Are you looking at it right now?	Confirming visibility of the object
Tablet	1	That's on top of the drawing cabinet, right	Confirming object location

Tablet	2	If you look where I'm looking you will see… Is it where my centre is…? Just next to where you look now, there is…	Used orientation cues to explain the location of the object
Glass, Tablet	3	Then if you go from there to your left, you will find … If you go 180 degrees from there… 90 degrees from that, you will find… If you keep going around, you will find…	Referencing to surroundings
Glass, Tablet	2	What letter does it start with? What colour is it?	Used language cues to guess the object
Glass	1	Perfect and completely with you	Agreement
Glass, Tablet		If you go to 3 o'clock… At my 6 o'clock…	Alternative mental representation during the Audio condition

TABLE 2: Examples of user expressions in Experiment One

Feedback

There was no consensus from the quantitative data about which of the visual cues was best in the first experiment. While there was a significant difference observed compared to the Audio only condition, it can only be said that visual cues were superior to no visual cues at all. This matched the results of the qualitative data, where there was some preference indicated from the qualitative comments, but no clearly superior visual interface. While the users considered the visual cues useful in all cases, there was criticism of the centre point as some users felt that it blocked their view of the image, even though all interface elements were semi-transparent.

Centred Radar

Feedback about the Centred Radar matched the goal of the interface: To merge all the information on one single point. *"Overall I think that the centered radar was the best because all the important info was centered at the point of view"*. However, it was also mentioned that there was too much information, which distracted from the panoramic image. One user mentioned a higher cognitive load when the Centred Radar, along with the Context Compass.

Context Compass

The Context Compass was described as the easiest interface to use, along with the Classic radar condition. It was positively mentioned that the Compass did not take up too much space and it was easier to concentrate on the panoramic image. It was also mentioned that the movement of the viewpoint on a line would be easier to match up with the collaborator. One user stated the circular motion of the other two visual cues made the participant feel like she and the other user were chasing each other trying to align their views. However, as the line is a more abstract interpretation of a circular movement of the head, it was been mentioned by one user that it would require a higher cognitive load and felt less intuitive.

Audio

Positive feedback for the Audio only condition was that it provided a clear image with no distracting elements; however, overall almost all users preferred having visual cues. There was one participant pair that had better quantitative results using the Audio condition than any other interface, and on further inspection it was found these participants devised a clock metaphor that allowed them to determine each other's position quickly and without any other visual cues.

Centre Point

The centre point was almost unanimously described positively, even being described as "the most important cue", and was often used as a orientation point for describing objects *"it is left of the centre point"*. While all the user interface elements were semi-transparent, the filled circle was perceived as blocking the view by some users. *"also if the objects is very small, the big blue dot can cover it which makes seeing the object difficult in the panorama view."*

The Frame

Feedback regarding the frame showed that it was not considered important or was not explicitly mentioned by the users. Only one user was satisfied with the frame. *"I liked the frame, I think that the frame is enough for most of the users using these types of interfaces, because of the analogy with the interface used in panoramic photos."*

7.2 Experiment Two

The analysis of results for Experiment Two was conducted using in the same manner as those of Experiment One.

7.2.1 Quantitative Results

Social Presence

Figure 40 shows the final composite Social Presence score for each interface condition and device used. For Glass users the Drawing condition had a significantly lower Social Presence than the Audio only condition, and Pointing and Drawing. There was no difference found for tablet.

Based on the work of [71] Social Presence was measured as a single dimension. The results of all eight questions of the Social Presence questionnaire were analysed with a Friedman test, which revealed a significant difference between the conditions for Glass users. ($\chi 2(3)=18.130$, $p < 0.0005$). The significance level was set to $p = 0.0083$ with an applied Bonferroni correction. The following conditions were found to be significant: Drawing – Audio (Z=-3.794, p<0.0005) and Drawing – Pointing+Drawing (Z=-3.103, $p = 0.002$), suggesting that the drawing interface produced the lowest presence on the Glass device.

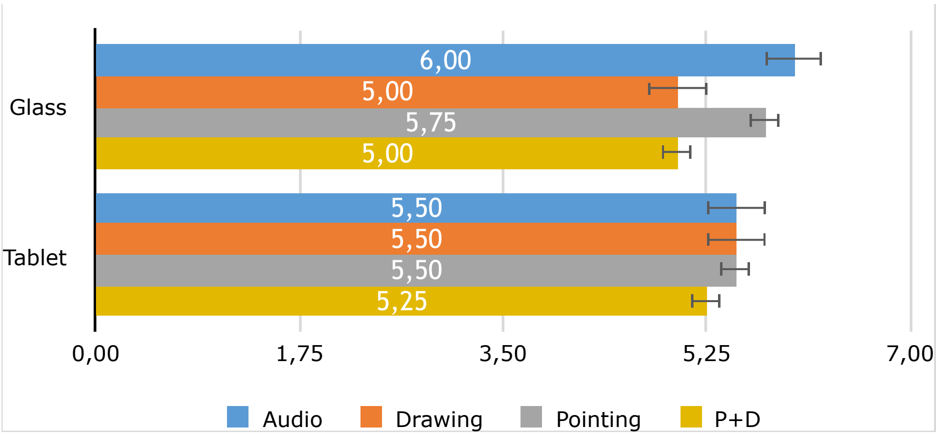


FIGURE 40: Median total scores for questions 1–8 for Social Presence for both devices

Perceived Task Performance

Users answered the following questions about perceived task performance:

1. How easy was it to be aware of what your partner was doing?
2. How well do you think that you managed to do the task in the system?
3. How efficiently did you feel that you managed to do the task?
4. How problematic was it to show the other person things in the environment?

An additional question about interaction was added to the Perceived Task Performance questionnaire for the second experiment. Figure 41 presents the median score for each of the Perceived Task Performance questions. The results of the Perceived Task performance were only significant for the Glass device. Q3 and Q4 were statistically significant. ($\chi2(3)=11.490$, $p = 0.009$) and Q4 ($\chi2(3)=8.429$, $p = 0.38$). The Wilcoxon signed-rank test showed a significant difference between Pointing – Drawing ($Z=-2.842$, $p = 0.004$) for Q3, meaning that Pointing was significantly more efficient in completing the task. While Q4 did not show any significant results between the groups with applied Bonferroni correction, it showed a trend towards Pointing – PointingDrawing ($Z=-2.558$, $p = 0.01$) and Pointing – Drawing ($Z=-2.288$, $p = 0.02$), suggesting Pointing was a better way of communicating ideas with a remote partner.

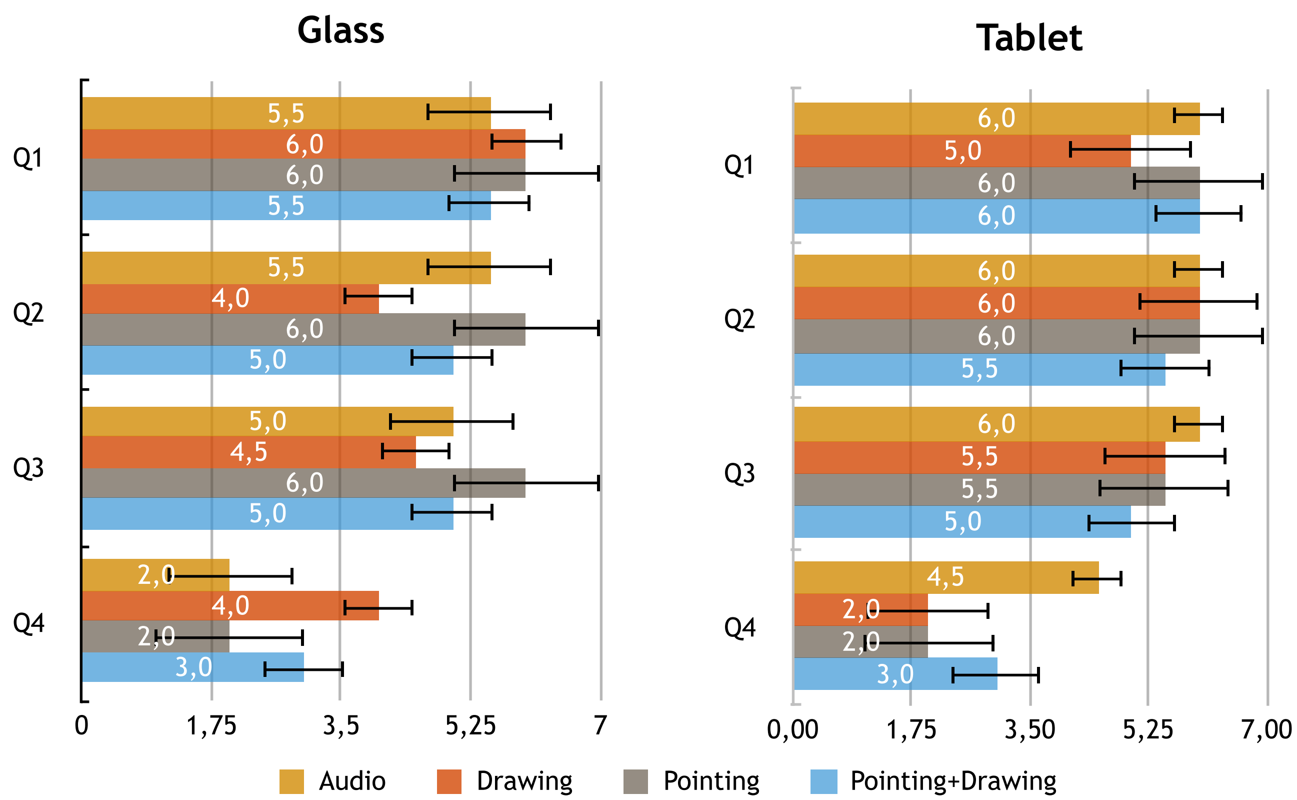


FIGURE 41: Median scores on Perceived Task Performance for the Presence study on Glass and Tablet

System Usability Scale

Overall, the average SUS scores for tablet users was *77.1 ±18.9* for the Audio condition, which indicates a “good” usability [72]. Furthermore, the Drawing and Pointing conditions scored *70.4±21.4* and *74.2±21.5* respectively. Both can be perceived as “good”. However, the dual condition Pointing and Drawing only scored *62.9±24*, indicating more work has to be done to improve the usability of this interface. Some tablet users mentioned they would have preferred an icon to change modes, rather than needing to use a tapping gesture.

For Glass users, the overall scores were the highest for Audio *(75.6±10.8)* followed by Pointing *(72.1±17.1)*, both of which indicate "good" usability. The dual mode Pointing and Drawing scored *58.1±18.6*, and the Drawing only scored *53.8±19.2*, both indicating "unacceptable" usability. A repeated measures ANOVA determined that the mean SUS scores differed statistically significantly between the conditions ($F(3, 33) = 5{,}625025$, $P = 0.003$). There was a significant difference between the highest ranked Audio and the Drawing condition ($p = 0.03$). The average results of the SUS scores are shown in Figure 42.

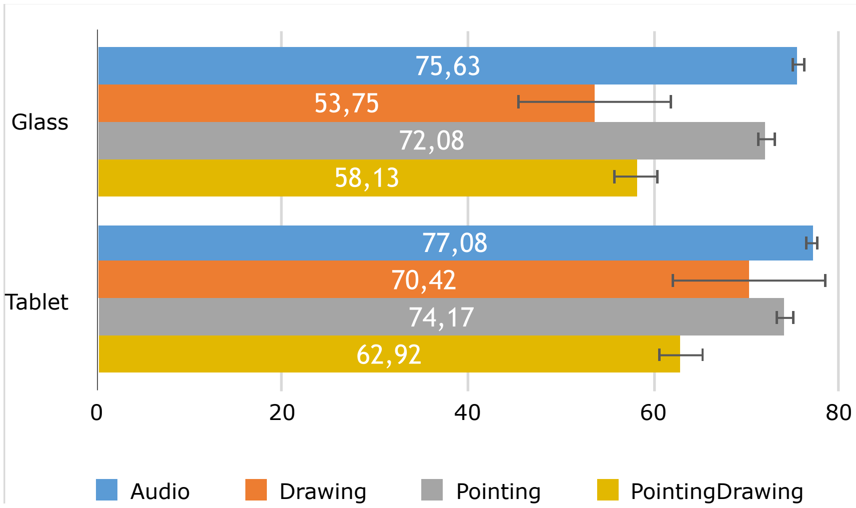


FIGURE 42: Average SUS scores for Glass and Tablet

Ranking

Users were asked to rank each of the conditions in preference for Interaction and Presence. Figure 43 shows the median results (1 = worst, 4 = best). Figure 43 shows the median ranking scores for both devices across all conditions. Conditions with the same median value are marked with their means as well for better understanding.

For Interaction preference, Glass users ranked the conditions as follows (from highest to lowest): Pointing, Audio, Pointing+Drawing and Drawing. The Presence preference was ranked as: Audio, Pointing, Pointing+Drawing and Drawing. There was a statistically significant difference for the Interaction ($\chi2(3)$=15.300, $p = 0.002$) and Presence ($\chi2(3)$=19.300, $p < 0.00005$) ranking. There was a significant difference in Interaction for

Drawing – Audio (Z=-2.803, p = 0.005) and Drawing – Pointing (Z=-2.766, p = 0.006). There were also significant differences in Presence for Drawing – Audio (Z=-3.100, p = 0.002), Drawing – Pointing (Z=-3.013, p = 0.003) and Pointing+Drawing – Drawing (Z=-2.804, p = 0.005).

The Interaction and Presence preference rankings were different for tablet users. For the Interaction preference, Drawing was set the highest, followed by Pointing, Pointing+Drawing, and finally Audio. For Presence, the Drawing condition is again first (Mean: *3*), followed Pointing (Mean: *2,92*), Audio and Pointing+Drawing. There was a statistically significant difference in the Interaction ranking (($\chi 2(3)$ = 9.100, p = 0.028) and Presence ranking ($\chi 2(3)$ = 7.891, p = 0.048), the latter with a significant difference between Drawing and Pointing +Drawing (Z=-2.645, p = 0.008). The post hoc test did not show the difference between the Interaction conditions.

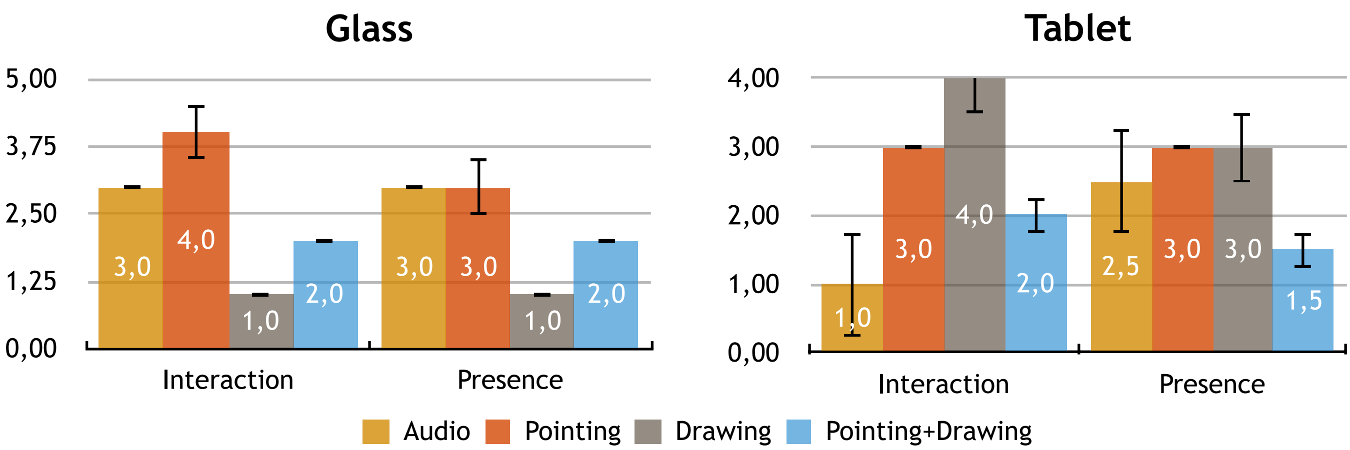


Figure 43: Median ranking scores on Interaction and Presence for Glass (left) and Tablet (right)

7.2.2 Qualitative Results

The logging of vector points allowed us to later reconstruct the interaction. The logs were time stamped, and for analysis the first seconds were taken out as the researcher would activate the application and then put the Glass device on the participant, thus generating data that is not part of the experiment. The logs were rerun on a tablet device and screenshots taken.

Observation

General Observations

Some users were found to take a leading role, by suggesting (using drawing and/or pointing) a location for the object, with their partners simply agreeing on the suggested location. These roles did not depend on which device users were using, even though Glass users had difficulty with the Drawing interface and often stopped using the tool. For tablet users, being able to hold the tablet on a 45 degree angle made drawing and pointing gestures easier to issue, while Glass users found it difficult to draw, and sometimes point, using the side touch panel.

It was observed during the second experiment that there was an increased usage of words such as "here" and "there", as the user was able to use the annotation to help describe the location of the object, with "here" being more commonly used than "there" for tablet and around equal for Glass. During the Audio condition none of the participants used the words "here" or "there", but rather relied on lengthy descriptions on the locations that could lead to ambiguities. Interestingly, tablet users used the word "here" more commonly compared to Glass, although they were placed in a different room. However, the most common usage of those words was for confirmation of a location. Table 3 shows some example verbal expressions used during the experiment. All participants took the descriptions of the furniture seriously and described them mostly in great detail.

User's device	Time(s)	User's speech	Comments
Glass, Tablet	1	Can we have a fishtank there? Can you see the whiteboard just here? Maybe we could put the lamp here?	Use of "here" or "there" during the interaction conditions.
Tablet, Glass	3	Do you mean the desk on the .. ahm .. (*pause*) As you come through the door the desk on your left is empty.	Ambiguities during the Audio condition
Glass, Tablet	2	The first picture is a lamp. It is quite a tall lamp. It is probably as tall as you. It is a big and low table. It is very elegant. We need to put it on the floor.	A user describing the object on the list
Glass, Tablet	2	On the next desk with the file racks. We can put the book shelf next to the file racks. In front of where the orange cone is.	Descriptions of the location
Glass, Tablet	1	Here? Yep, there it is.	Confirming a location

TABLE 3: Examples of user expressions in Experiment Two

Use of Pointing or Drawing

Although all participants had their own way of indicating location or shape, there were common practices used depending on the device. Tablet users generally appeared to be comfortable with drawing and sometimes tried to draw the shape of the object. Figure 44 demonstrates an example of a user's attempt to draw an orange juicer on their tablet. Users with the Glass device were often observed trying to draw a specific object, but many quickly gave up and ended up using more abstract representations, such as rectangles or circles.

Geometric shapes were also used by tablet users, for example to indicate a spot on the wall for a picture, and, as many tablet users felt comfortable using the drawing condition, they would sometimes just draw an arrow to point to a specific location, instead of using the pointer itself when they had the option to. In the reverse case, Glass users often favoured staying in the pointing mode, making repeated circular motions with the pointer to indicate object size and shape. Figure 45 demonstrates some examples of the drawing behaviour, showing a tablet user's drawing of an arrow to indicate the chosen spot for discussion (a), attempted drawing by a Glass user (b), and examples of geometric shapes: a circle marking the spot for discussion (c), and a rectangle to depict the location of a portrait (d).

Due to the limited size of the touch interaction space and the separation of touch surface from the display surface in the Glass device, users had difficulty with drawing on Glass and often chose to only use only arbitrary shapes. One of the biggest difficulties during the drawing only condition on Glass was that the user had no real indication on where their drawing would start on the panorama. One user came up with a unique solution to this problem by starting drawing at the top of the image and drawing a line downwards to the point where he wanted the line to be. This user's images were less messy, although it did require some communication with his partner to explain his drawing technique.

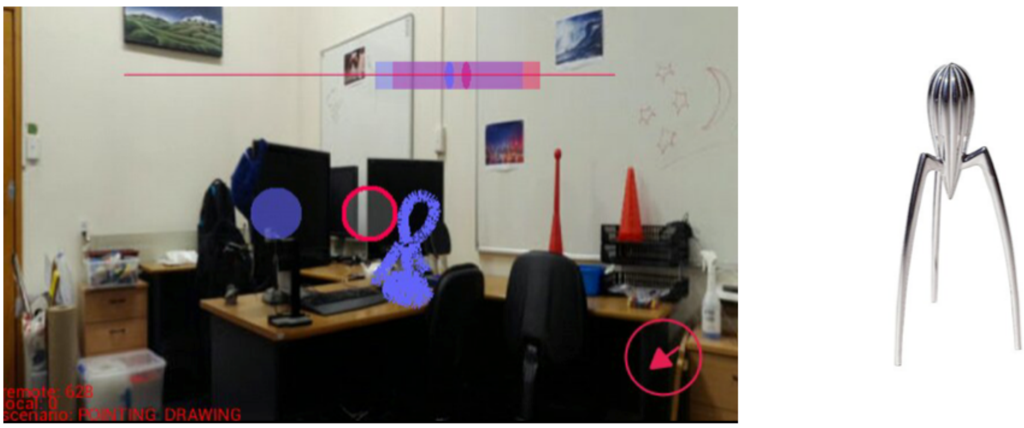
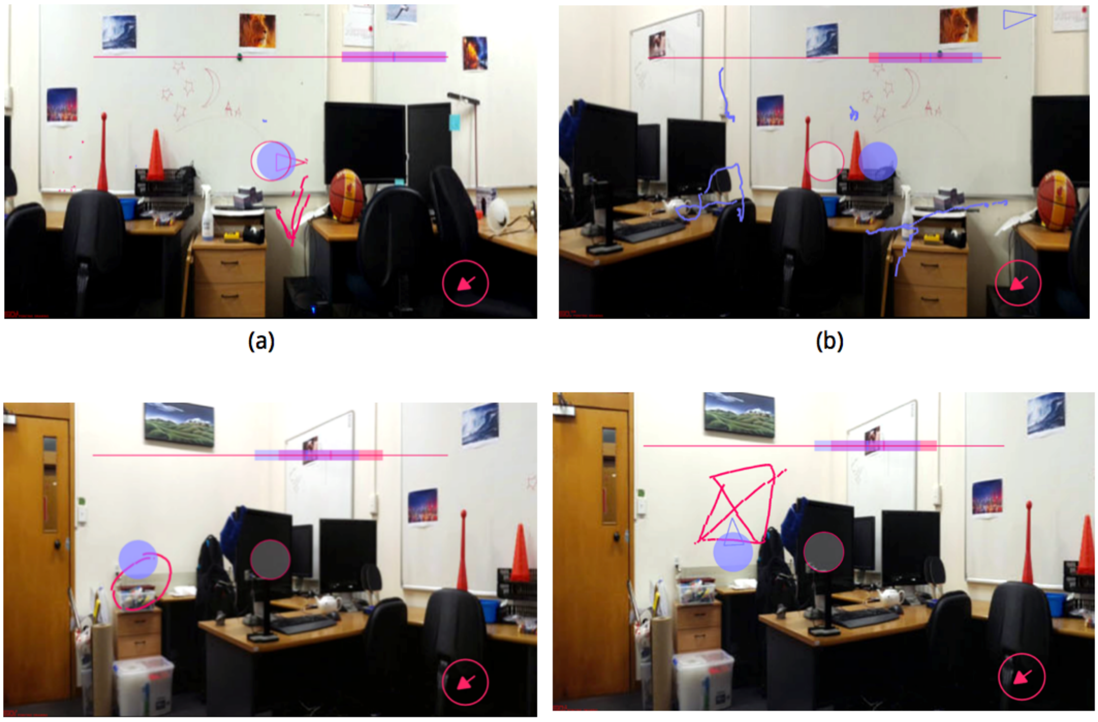


FIGURE 44: Example of a user trying to draw a specific shape during the study

When agreeing on a location, both participants would typically use the pointer tool to point at the same location, confirming to the other user that they agreed with the location.

Engagement

Users engaged with the remote user, using orientation cues to orient themselves to their partners view, and annotating on the panorama to improve their communication. One user on the Glass pointed out how this established an intimate connection with the other participant even though they did not know each other beforehand, due to the fact that they were able to see the same panorama and share the experience of placing the furniture.



(a) (b) (c) (d)

FIGURE 45: Drawing observation examples.
Tablet user drawing an arrow (a), Glass user attempting drawing (b),
Use of geometric shapes in (c) and (d)

Feedback

After the experiment, users were asked for their preference of pointing or drawing, and asked for feedback on where they deemed each tool as more appropriate.

Glass

Out of 12 participants on Glass, there were 10 that preferred using the pointing only condition, one user that preferred them equally, and only one user who preferred the drawing mode, although he still described drawing as “difficult”. Users mentioned drawing on the Glass to be “too cumbersome” and “too difficult”, with one user stating they did not feel “in control” as the size of the touchpad was too small to be conveniently used, some may have had difficulty with mapping of the 2D motion on the touch pad to the 2D drawing output on the screen. One Glass user had difficulties with the two-finger mode switching and described it as “fuzzy”. Two subjects felt that pointing would be sufficient for this type of task, but that in an entirely unknown and complex environment where more detailed explanation would be required, drawing could potentially be superior. Some others appreciated having the drawing option, but felt that the touchpad was too small to be used effectively.

Overall, most users found the pointing to be useful for “showing a spot”, and that it is “easier and quicker for simple objects”. The perceived benefits of drawing were to show dimensions and orientation and to offer more specific details. However it was often felt that simple shapes would be sufficient for this, and that “circles and rectangles are enough to get an idea across”. Only one user found no use in drawings whatsoever, describing the function as “not useful at all” and that “for outlining pointing is enough”. Finally, two users felt that it was as easy to talk as to point, that one interaction method is enough, and due to the difficulties with the interaction interface, audio was sufficient to reach a quick conclusion with the awareness cues provided being enough for collaboration.

Some users felt that the mapping from a forward/back motion of the glass touchpad to a left/right action on the screen felt “reversed”. During the pilot study the mapping was in the opposite direction to the final experiment, and there were also complaints that the mapping felt “wrong”, indicating that, when using this sort of mapping, it may be necessary to allow users to choose the direction of the mapping that feels natural to them.

Tablet

Contrary to Glass, on the tablet only two users preferred pointing, one user preferred them equally, and the remaining nine mainly preferred the drawing tool. Two users described drawing as “fun”, while another never used the pointing tool regardless of task, explaining “you can just draw an arrow for pointing”. One user was confused by the method of switching tools. While general consensus was that pointing was perceived as not being necessary, it was still used during the experiment to show a specific location or point to small objects, with drawing mainly being used to highlight large objects and shapes. One user suggested that the icon that indicates the current mode should be “clickable” in order to switch modalities.

More generally, some users mentioned that with the Audio only interface they could concentrate more on the image and the task, as opposed to having navigation and interaction tools that distracted them.

7.3 Summary

Experiment One: Awareness

This experiment has shown that visual cues can significantly enhance social awareness when sharing a Social Panorama space, compared to using audio only. Users were able to use the visual cues to quickly follow each other around the social space and establish common ground, which resulted in the feeling that they could collaborate better. However, it was found that there was no difference between the visual cues used. This may have been because all the visual cue conditions included the centre point in the interface, which many subject felt was the most important visual cue. Some users developed strategies to improve understanding when using the Audio condition, for example using a clock metaphor or stating the number of degrees to turn. This indicates that any sort of cue - as long as it is understood by all participants involved – can provide good enough guidance for awareness of the remote user’s viewpoint.

Experiment Two: Presence

In this experiment we found that adding interaction tools did not increase the Social Presence for tablet users. On the Glass device, the drawing condition actually decreased presence compared to other conditions. Glass users found the drawing condition to be very difficult to use, suggesting that if a Social Panorama interface is not easy to use it will have a negative impact on Social Presence. The Glass drawing interface also scored lower on Perceived Task Performance when compared to the Pointing only condition. The Audio condition scored surprisingly high on the SUS and on some of the presence questions. Users felt that the Audio did not distract them from their task resulting in a higher immersion in the panorama. The task in the experiment was of low complexity, and thus easy to solve even without interaction tools. Drawing usability problems did not occur on tablet, however there weren't any significant differences between the conditions, and users performed equally well despite tending towards using the Drawing interface. Drawing usability problems did not occur on tablet, however there weren't any significant differences between the conditions, and users performed equally well despite tending towards using the Drawing interface. On tablet, although all modalities except “Pointing+Drawing” received a “good” score on the SUS the overall values on Social Presence were low with 5,5. This indicates that good usability does not necessarily increase Social Presence.

Users were observed using different strategies to describe an object: Some would try to draw very specific forms and shapes, while most others would rely on basic geometric shapes to get their idea across. Due to the difficulties of drawing on Glass, users relied on the pointing tool to outline shapes and used more verbal communication. The results confirm the observations from previous studies [12] that drawings were more commonly made to explain shape and dimensions and pointing gestures were usually used to reference to a location or an exact object.

8 Discussion and Limitations

This chapter interprets the results of the previous chapter, and outlines some of the limitations of the study.

8.1 Discussion

During the first experiment the Audio interface scored significantly lower than the visual interfaces on most of the questions and on both devices for perceived task performance and collaboration. However there was no visual interface that was superior to the others in terms of awareness or establishing common ground. The main concept of how the interface worked was understood well by the users, resulting in no further differences. Many of the participants were University students, where it can be assumed that they have experienced a similar interface in an application or game. More important was the fact that both types of cues, egocentric and exocentric, were implemented as these provide different benefits of an overview and a more detailed view. The former is used for awareness of the current viewpoint of the other user, and the latter on the overall current location in relation to the local user.

The second study provided a few surprises. While Audio scored low in collaboration and perceived task performance in the first study, the condition performed surprisingly well in some cases. In terms of Social Presence (Glass), Ranking of Presence (Glass) and Usability on both devices it scored the highest. This may be explained in part due to the interaction difficulties users were having with the Glass device, which suggests the ease of use of an interface can have an impact on Social Presence. It was also mentioned by participants that the permanent use of the touch panel felt unnatural, and all information they needed to perform the task quickly was already provided by the awareness cues and the possibility to talk. Although it did require more words and lengthy descriptions it was not perceived as bothersome.

While previous research has indicated Social Presence is vital in terms of satisfaction inside a virtual learning environment [74], there have also been studies that were not able to demonstrate a significant relationship between co-presence and satisfaction [75], which may mean that the presence felt was more of a physical presence as opposed to being socially connected. Some users mentioned during both experiments that it was much easier to concentrate on the panorama environment without the “distractions” of visual cues or interaction, suggesting a greater immersion in the environment. One possible solution to this would be to have the exocentric cues fade out once the user stops moving, and to put more importance on egocentric elements to lessen the cognitive load and distraction by too many interface elements.

The environment used in the study was fairly simple, and Audio only could have been sufficient to resolve the tasks. In more complex and detailed environments, there may be too many ambiguities for speech alone, and other modes of interaction and communication might be able to resolve the tasks much more conveniently. Nonetheless, these results underline the importance of Audio for remote collaboration and presence.

Using the Glass device, some users relied on the real setting as opposed to the image displayed on the screen, due to the higher level of detail in the real world compared to that of the small display. Other users were completely immersed in the panoramic image, with results similar to a study of panoramas on mobile devices, where users did not realise there were not looking into the real world [18]. Even those users who had difficulties with the interaction were still very actively involved in the task, and developed other means of communication, although this may be due to the collaborative nature of the tasks.

Glass users who favoured using the real world over the panorama only used the touchpad to "make a point" and confirm things with their partner. For these users, only minimal cues and interaction were favoured, suggesting that tablet and Glass users have very different interaction patterns.

Based on the results, reducing the complexity of interaction is favourable, as most users only used one interaction method the whole time. Although tablet users had more "freedom" in terms of possible interaction methods, most favoured the drawing interaction, only using the pointer for short interactions to "make a point" or confirm something. On the contrary, glass users had difficulty with the drawing, and tended to favour the pointing interface as it reduced the complexity of interaction. Overall, modalities have to be targeted to the specific task and device to support the user, and this research has given directions in where to go to "tailor" the needs of each user more appropriately.

It seems that awareness and interaction both did not have an effect on Social Presence in this study. In the first study awareness scored equally for cues compared to no cues, and the second study did not show interaction improving presence for this task.

8.2 Limitations

One of the main limitations of both experiments was the small number of participant groups, with only 12 per experiment, and it is possible that some of the results that came close to a significant level would have been more conclusive with more participating groups.

In the application, the panorama was mapped onto a cylinder, while a more “realistic” panorama would have been mapped on a sphere. This was a deliberate decision as a spherical panorama would have added another dimension to work with, resulting in a more complex scenario (e.g. trying to search for your partners viewpoint in two dimensions rather than one). We also felt this was not a significant issue as many current user applications which capture panoramas also only offer cylindrical panoramic.

There were some limitations regarding the interface. Drawing issues on Glass lead to insecurity as the small touch pad was difficult for more complex drawings. In addition, the “clear” function, issued as a three finger tap, proved difficult for some users, especially on Glass where the small touchpad would sometimes not recognise all three fingers. This could have led to further frustration using the Drawing interface, even though it did not affect the actual drawing.

All experiments were conducted indoors in a controlled environment, even though the original use case also described outdoor collaboration. This choice was made for several reasons. First, the system was always intended for indoor use as well, and in such an early stage it made sense to keep the environment simple and reduce any confounding factors that could bias communication. Furthermore, the present configuration of the application does not have any verbal communication integrated, thus the participants had to use additional equipment to provide verbal communication.

In terms of methodology, all questionnaires were based on those used and validated by other researchers beforehand. However, the concept of Presence is very broad with lots of different definitions, and it could be seen as a limitation that the study only assessed one definition of presence. In further studies, different concepts of presence could be further examined, for example examining the concept of co-presence, which relates to the feeling of being physically together. Unfortunately, as there is no firm definition on what presence is or how to measure it, it is difficult to compare studies even when they aim to measure the same definition of presence as the methods vary significantly.

The type of task during the experiment may result in different outcomes. Both experiments involved simple tasks that could be solved easily with audio, explaining the high scores in many presence and usability related questions. More complex tasks could yield a different result. Tasks involving complex and more detailed outdoor imagery that require more precise tools could lead to a better understanding of possible gesture representations, that would be difficult by speech alone.

One final limitation was the fact that the Frame as an interface element was only introduced into one of the interfaces, the Classic interface, for the awareness study. This was been done deliberately as the Centred Radar was designed to be as simple as possible, and the Context Compass already integrated a similar design using rectangular shapes. However, at least one user mentioned that the Classic interface was his favourite because of the frame. Although ultimately there were no measured differences between the visual cues, this may have biased the results.

9 Conclusion and Future Work

9.1 Conclusion

Panoramic imagery can provide an immersive and holistic impression of an environment. With the ubiquity of smartphones equipped with high-quality cameras and the desire of people to share their experiences and feel connected, a panorama can support this by offering the impression of a remotely located user to be close to another with the same freedom of looking around independently as in a real environment. However, the use of panoramic imagery in a mobile collaborative environment has not been researched thoroughly, leaving many gaps in interface guidelines and on how the sensation of connection and shared experiences can be utilised in such an omni-directional view setting.

This research aims to contribute in the field of wearable computing and sharing personal experiences by using panoramic imagery and exploring remote and collaborative interaction modalities and their impact on presence. To achieve this goal a prototype was developed that simulates an already captured panorama that is presented on a tablet and on a HMD, following an extensive needs analysis and brainstorming process. The implemented user interface supported the awareness necessary for successful grounding, and the addition of pointing and drawing provided tools to reference objects or locations in the panorama image quickly.

The UIs for awareness were based on previous research as well as an original design. A pilot study quickly showed the importance of combining exocentric and egocentric cues, which resulted in interfaces that provided an overall view of the space and where the remote user is currently looking, as well as a more detailed view once they overlapped. In addition, established interaction methods for remote collaboration were implemented for pointing and drawing and it was explored whether these modalities can add to the sense of being with another in the same space.

Two formal user studies were conducted that tested the prototype on the best interface to support awareness, the effect of an interaction modality on Social Presence, and its impact on usability. The study on awareness consisted of three different interface types and one audio condition, whereas the study on presence consisted of three interaction conditions and one audio condition. The prototype was tested by twelve groups on each occasion.

From Experiment One we found that additional visual cues like Radar displays or a Context Compass could indeed be used to significantly increase awareness compared to an audio only interface. Users were able to use the visual cues to quickly follow each other around the social space and establish common ground. However there was no difference between the visual cues, and awareness cues did not seem to have an impact on co-presence. The assumption that visual cues support awareness was proven.

From Experiment Two we found a difference on the interaction tools being used for Glass, with Drawing and Pointing+Drawing receiving a significantly lower score than Audio. It was also Drawing that scored the significantly lower than Audio in terms of usability, and in the Presence ranking, showing that a bad usability can have a negative effect on Social Presence. However, "good" usability does not imply a high sense of presence, as the overall values on Social Presence did not exceed 6 on any modality. Bad usability seems to make the user aware of the mediated virtual environment and reduce the sense of presence. There was also a clear preference from users for pointing tools on Glass, as Drawing was perceived as difficult and the extensive use of the touch panel as unnatural. Because of this, the assumption that annotation methods positively influence Social Presence had to be rejected.

Both studies contributed to research in different ways. Visual cues significantly improve collaboration and awareness, and the results of the presence study lays foundations for future studies, specifically in terms of collaboration for wearable computing and panoramas. Future research can take into account following suggestions:

- The use of exocentric cues when the users are active looking around to provide awareness of the other user's position, but fade out once they viewpoints meet.
- Egocentric cues should be considered when both users share the field of view to put focus on details and provide less distractions. Egocentric cues should fade out once there is movement.
- Interaction modalities should be thoroughly tested and adjusted to the device in use. For the users of a wearable device, this implies quick and easy access to its interaction tools that do not need prolonged use of a small touch panel. The pointing method should be considered for Glass, while Drawing is deemed as the better tool for more precise interaction on tablets.
- One interaction modality is sufficient
- Audio support for speech communication

In summary, this thesis described the concept of Social Panoramas, using wearable computers, cameras and displays to have shared social experiences. A working prototype was developed on Google Glass and then used to explore the key issues of awareness and Social

Presence and if the technology can be used to create awareness cues and increase Social Presence.

9.2 Future Work

This research is the first steps into a new field of collaboration using Social Panoramas between HMDs and tablet devices. New wearable devices offer more immersive experiences with users that are in remote locations.

Short-term improvements to the technology would involve applying the lessons learned about interaction and interfaces. The issues on each device, such as the difficultly with drawing on Glass, have to be addressed and further "tailored" for each user. Due to the success of the pointing interaction on Glass, it may be worth exploring if changing shapes of the pointer would be useful for specific tasks, as long as it does not increase the interface complexity too much.

Voice commands present an interesting research area, but care must be taken to reduce disruption of communication between participants. Natural gesture Interaction is another interesting field of research, for example streaming hand gestures between remote users, and allowing users to point at objects using their real hands.

The results of experiments shows there is room for improvement in the user interface. Some users felt various features were too distracting, and an interface that fades unused parts away could be further explored. In particular, when the users are looking in the same direction, the exocentric view could fade out, reducing possible distractions on the screen. Only once the user's views do not overlap would the display fade in again.

The alignment of the centre points could be facilitated by adding a "magnet". When centre points are close together they could automatically overlap to provide a fixed position for further interaction.

To improve awareness of remote annotation, off-screen widgets could be implemented that indicate when the partner is currently drawing/pointing somewhere off-screen. This could especially be useful for larger collaboration groups.

Audio has been shown to be an integral part of communication and Social Presence. Future research could explore the use of spatial audio, where the voice of the other user could be heard depending on his or her viewpoint. Using more of a user's senses could provide a better understanding of where the user is looking, and potentially increase the sense of presence. Visual cues would confirm what the user heard.

Haptic feedback has been shown to have favourable results in terms of presence [71, 46]. With haptic feedback, it would make sense to use it on occasions where user's viewpoints "meet", to enforce the metaphor of touching. This could be either on every view alignment, or every time the users' pointers "meet" to confirm pointing at the same location, as was observed during the study. The device could slightly vibrate each the viewpoints meet.

With the recent popularity of depth sensors, there has been significant interest in more intuitive forms of gesture input – such as natural hand gestures, although these were not explored. The goal of this research was to explore possibilities of current hardware, without adding any additional sensors or cameras. In its present state, any sort of image processing quickly drains the battery power of Google Glass, and data exchange with a server comes with a cost of latency. In the future sensors such as MIME [76] might be introduced to these devices, and provide depth sensors on wearable devices.

Another possibility for future work would be to offer a live-stream of the current field of view of the local user, leveraging a Mixed Reality view to show a combination of captured image and live video. These new interface elements will have to be evaluated in further experiments which focus on usability and Social Presence.

As the current prototype has only been tested in a controlled indoor environment, a higher fidelity prototype should be tested outdoors under real world conditions. An outdoor environment introduces several additional challenges, for example ambient noise, moving backgrounds and more complex surroundings.

In future research, more attention could be given to task performance. Although efficiency was not the focus of the project, testing efficiency could provide a quantitative measure on how collaboration processes improve. Previous research has shown that the amount of remote gestures correlate with faster task performance. [27]

Finally, future work could include the possibility of using spherical panoramas. This would require further research regarding how the interface would operate. A complete implementation of the proposed system, including real time panorama capture, stitching and sharing, would make the idea of Social Panoramas real. While it is not yet known how consumers will use wearable devices such as Google Glass, they do offer new possibilities for collaboration, with much left to explore.

References

[1] A. Weilenmann, T. Hillman, and B. Jungselius, “Instagram at the Museum: Communicating the Museum Experience through Social Photo Sharing Alexandra,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13, 2013, p. 1843.

[2] W. IJsselsteijn, J.Van Baren, and F. Van Lanen, “Staying in touch: Social presence and connectedness through synchronous and asynchronous communication media,” in C. Stephanidis and J. Jacko (eds.), Human-Computer Interaction: Theory and Practice (Part II), volume 2 of the Proceedings of HCI International 2003, 2003, pp. 924-928.

[3] D. Kirk, T. Rodden, and D. S. Fraser, “Turn it THIS way: grounding collaborative action with remote gestures,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’07. ACM, New York, NY, USA, 2007, 1039-1048.

[4] W. Huang and L. Alem, “HandsinAir: a wearable system for remote collaboration on physical tasks,” in Proceedings of the 2013 conference on Computer supported cooperative work companion - CSCW ’13, 2013, p. 153.

[5] F. Biocca, C. Harms, and J. K. Burgoon, “Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria,” Presence Teleoperators Virtual Environ., vol. 12, no. 5, pp. 456–480, Oct. 2003.

[6] D. Trendafilov, Y. Vazquez-Alvarez, S. Lemmelä, and R. Murray-Smith, "Can we work this out?: an evaluation of remote collaborative interaction in a mobile shared environment,” in Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11). ACM, New York, NY, USA, 2011, pp. 499-502.

[7] Google Street View: http://www.instantstreetview.com/

[8] Microsoft StreetSide: http://www.microsoft.com/maps/streetside.aspx

[9] F. Pece, W. Steptoe, F. Wanner, S. Julier, T. Weyrich, J. Kautz, and A. Steed, “Panoinserts,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13, 2013, p. 1319.

[10] A. Au and J. Liang, "Ztitch: A mobile phone application for immersive panorama creation, navigation, and social sharing," in IEEE 14th International Workshop on Multimedia Signal Processing (MMSP), 2012, no. April, pp. 13–18.

[11] W. Xu and J. Mulligan, "Panoramic video stitching from commodity HDTV cameras" in Multimed. Syst. 19, 2013, pp. 407–426.

[12] Photo Sphere: http://www.google.co.nz/maps/about/contribute/photosphere/

[13] TourWrist: http://www.tourwrist.com/

[14] Brown, M.: Autostitch: http://www.cs.bath.ac.uk/brown/autostitch/ autostitch.html

[15] Photosynth: http://photosynth.net/

[16] J. Ardouin, A. Lécuyer, and M. Marchal, "FlyVIZ: a novel display device to provide humans with 360° vision by coupling catadioptric camera with HMD," in 18th ACM symposium on Virtual reality software and technology., 2012, pp. 41–44.

[17] T. Langlotz, D. Wagner, A. Mulloni, and D. Schmalstieg, "Online Creation of Panoramic Augmented Reality Annotations on Mobile Phones," IEEE Pervasive Comput., vol. 11, no. 2, pp. 56–63, Feb. 2012.

[18] J. Wither, Y.-T. Tsai, and R. Azuma, "Indirect augmented reality," Comput. Graph., vol. 35, no. 4, pp. 810–822, Aug. 2011.

[19] R. Kaiser, M. Thaler, A. Kriechbaum, H. Fassold, W. Bailer, and J. Rosner, "Real-time Person Tracking in High-resolution Panoramic Video for Automated Broadcast Production," 2011 Conf. Vis. Media Prod., pp. 21–29, Nov. 2011.

[20] D. Seo, S. Kim, H. Park, and H. Ko, "Real-time panoramic video streaming system with overlaid interface concept for social media," Multimed. Syst., Aug. 2013.

[21] J. Norris, H. Schnädelbach, and G. Qiu, "CamBlend: an object focused collaboration tool," in Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12, 2012, p. 627.

[22] T. Höllerer, S. Feiner, J. Pavlik, "Situated Documentaries: Embedding Multimedia Presentations in the Real World," in Proc. of the Third International Symposium on Wearable Computers - ISWC '99, 1999, pp. 79-86.

[23] Wearcomp papers: http://wearcomp.org/

[24] L.-T. Cheng and J. Robinson, "Dealing with speed and robustness issues for video-based registration on a wearable computing platform," Wearable Computers, 1998. Digest of Papers. Second International Symposium on , vol., no., pp.84,91, 19-20 Oct. 1998

[25] R. E. Kraut, S. R. Fussell, and J. Siegel, “Visual information as a conversational resource in collaborative physical tasks,” Hum.-Comput. Interact. 18, pp. 13-49, June 2003.

[26] S. Bottecchia, J.M Cieutat, and J.P. Jessel, “TAC: augmented reality system for collaborative tele- assistance in the field of maintenance through internet,” in Proceedings of the 1st Augmented Human International Conference, 2010, p. 14.

[27] S. Fussell, L. Setlock, J. Yang, J. Ou, E. Mauer, and A. Kramer, “Gestures Over Video Streams to Support Remote Collaboration on Physical Tasks,” Human-Computer Interact., vol. 19, no. 3, pp. 273– 309, Sep. 2004.

[28] N. Sakata, T. Kurata, T. Kato, M. Kourogi, and H. Kuzuoka, “WACL: supporting telecommunications using - wearable active camera with laser pointer,” in Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings., pp. 53–56.

[29] H. Kuzuoka, J. Kosaka, K. Yamazaki, Y. Suga, A. Yamazaki, P. Luff, and C. Heath, “Mediating dual ecologies,” in Proceedings of the conference on computer supported cooperative work, 2004, pp. 477-486.

[30] J. Ou, S.R. Fussell, S. R. Fussell. X. Chen, L.D. Setlock, and J. Yang, “Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks,” in Proceedings of the 5th conference on Multimodal interfaces, 2003, pp. 242-249.

[31] D. S. Kirk, and D. S. Fraser, “The Impact of Remote Gesturing on Distance Instruction,” in Proceedings of the Conference on Computer Supported Collaborative Learning, 2005, pp. 301-310.

[32] D. S. Kirk, and D. S. Fraser, “Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks,” in Proceedings of the Conference on Human Factors in Computing Systems, 2006, pp. 1191-1200.

[33] R. Poelman, O. Akman, S. Lukosch, and P. Jonker, “As if being there: mediated reality for crime scene investigation,” in Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW ’12, 2012, pp. 1267-1276.

[34] L. Alem, F. Tecchia, and W. Huang, “HandsOnVideo: Towards a gesture based mobile AR system for remote collaboration,” in Alem, L. and Huang, W. (eds), Recent trends of mobile collaborative augmented reality systems. 2011, pp. 135-148.

[35] J.C. Tang, and S.L. Minneman, “VideoWhiteboard: video shadows to support remote collaboration,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’91, 1991, pp. 315-322.

[36] E. Veinott, J.S. Olson, G.M. Olson, and X. Fu, “Video Helps Remote Work: Speakers who need to negotiate common ground benefit from seeing each other” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’99, 1999, pp. 302-309.

[37] M. Hassell, S. Goyal, M. Limayem, I. Boughzala, “Being there: An empirical look at learning outcomes in 3D virtual worlds”, 2009.

[38] M. Siitonen and M. Olbertz-siitonen, “I Am Right Here With You – Constructing Presence in Distributed Teams,” in Proceedings of International Conference on Making Sense of Converging Media (AcademicMindTrek ’13), pp. 11–16, 2013.

[39] H. H. Clark and S. E. Brennan, “Grounding in communication,” in Perspectives on Socially Shared Cognition, L. Resnick, J. Levine, and S. Teasley Eds., American Psychological Association, 1991, pp. 127–149.

[40] J. Rowan, and E.D. Mynatt, “Digital Family Portrait field trial: Support for aging in place,” in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’05, 2005, pp. 521-530.

[41] C. Wisneski, H. Ishii, A. Dahley, M.G. Gorbet, S. Brave, B. Ullmer, and P. Yarin, “Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information,” in Proceedings of the First International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture -. CoBuild '98, 1998, pp. 22-32.

[42] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer, “Integrating the physical environment into mobile remote collaboration,” in Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI ’12, 2012, pp. 241-250.

[43] A. Xu, J. Biehl, E. Rieffel, T. Turner, and W. van Melle, “Learning How to Feel Again: Towards Affective Workplace Presence and Communication Technologies,” in Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI ’12, 2012, p. 839.

[44] R. Schroeder, “Social interaction in virtual environments: Key issues, common themes, and a framework for research,” in R. Schroeder (Ed.), The social life of avatars: Presence and interaction in shared virtual environments, 2002, pp. 1-18.

[45] J. Short, E.Williams, and B. Christie, “The social psychology of telecommunications,” John Wiley & Sons, 1976.

[46] E.-L. Sallnäs, “Effects of Communication Mode on Social Presence, Virtual Presence, and Performance in Collaborative Virtual Environments,” Presence Teleoperators Virtual Environ., vol. 14, no. 4, pp. 434–449, Aug. 2005.

[47] R.L. Daft, and R.H. Lengel, “Organizational information requirements, media richness, and structural design,” Management Science, 32(5), 1986, 554—571.

[48] R.L. Daft, “Essentials of organization theory and design,“ South-Western College Publishing, 1998.

[49] J. Van Baren and W. IJsselsteijn, “Measuring presence: A guide to current measurement approaches,” Deliv. OmniPres Proj. IST-2001-39237, 2004.

[50] C. Hendrix, and W. Barfield, “Presence within virtual environments as a function of visual display parameters,” in Presence: Teleoperators and Virtual Environments, 5, 1996, pp. 274 - 289.

[51] M. Lombard, T.B. Ditton, D. Crane, B. Davis, G. Gil-Egui, and K. Horvath, “Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument,” in Proceedings of the Third International Workshop on Presence, 2000.

[52] Glass website: http://www.google.com/glass/

[53] Oculus VR. Oculus rift website: http://www.oculusvr.com, June

[54] M.R. Endsley, B. Bolte, and D.G. Jones, “Designing for Situation Awareness: An Approach to Human-Centered Design,” CRC Press, 2003.

[55] J. D. Gould and C. Lewis, “Designing for usability: key principles and what designers think” Commun., 3, pp. 300-311, March 1985.

[56] J. L. Gabbard and J. E. Swan II, “Usability Engineering for Augmented Reality: Employing User-Based Studies to Inform Design”, in IEEE Transactions on Visualization and Computer Graphics 14, 3, 2008, pp. 513-525.

[57] J. Hwang, J. Jung, and G. J. Kim, “Hand-held virtual reality: a feasibility study,” in Proc. of the ACM symposium on Virtual reality software and technology - VRST ’06, 2006, pp. 356-363

[58] S. Benford, J. Bowers, L.E. Fahlén, Greenhalgh, C. and S. Snowdon, “User Embodiment in Collaborative Virtual Environments,” in Proceedings of the ACM Conference on Computer-Human Interaction - CHI’95, 1995, pp. 242-249.

[59] A. Mulloni, A. Dünser, and D. Schmalstieg, “Zooming interfaces for augmented reality browsers,” in Proceedings of the 12th international conference on Human computer interaction with mobile devices and services - MobileHCI ’10, 2010, pp. 161-170.

[60] M. Rantanen, A. Oulasvirta, J. Blom, S. Tiitta, and M. Mäntylä, “InfoRadar: group and public messaging in the mobile context,” in Proceedings of the third Nordic conference on Human-computer interaction - NordiCHI ’04, 2004, pp. 131-140.

[61] Wikitude website: http://www.wikitude.com/

[62] G. Lee, A. Duenser, A. Nassani and M. Billinghurst, “AntarcticAR: An Outdoor AR Experience of Virtual Tour to Antarctica,” in ISMAR2013 - International Symposium on mixed and augmented reality, 2013, pp. 29-38.

[63] R. Suomela and J. Lehikoinen, “Context compass,” in Digest of Papers. Fourth International Symposium on Wearable Computers, 2000, pp. 147–154.

[64] P. Baudisch and R. Rosenholtz, “Halo: a technique for visualizing off-screen objects,” In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’03, 2003, pp. 481-488.

[65] Processing http://www.processing.org/

[66] Ketai Library https://code.google.com/p/ketai/

[67] OSC protocol for processing http://www.sojamo.de/libraries/oscP5/

[68] W. Barfield and S. Weghorst, “The sense of presence within virtual environment: a conceptual framework” in Proceedings of the 5th international conference on human–computer interaction, 1999, pp. 699–704.

[69] A. Henrysson, M. Billinghurst and M. Ollila, “Face to face collaborative AR on mobile phones” in Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality - ISMAR’05, 2005, pp. 80-89.

[70] H. Q. Dinh, N. Walker, C. Song, A. Kobayashi,and L. F. Hodges, “Evaluating the Importance of Multi-sensory Input on Memory and the Sense of Presence in Virtual Environments,” in Proceedings of the IEEE Virtual Reality - VR '99. 1999, pp. 222-228.

[71] C. Basdogan, C. Ho, M. A. Srinivasan, and M. Slater, “An experimental study on the role of touch in shared virtual environments,” ACM Trans. Comput. Interact., vol. 7, no. 4, pp. 443–460, Dec. 2000.

[72] A. Bangor, P. T. Kortum, and J. T. Miller, “An Empirical Evaluation of the System Usability Scale,” Int. J. Hum. Comput. Interact., vol. 24, no. 6, pp. 574–594, Jul. 2008.

[73] J. Brooke, “SUS-A quick and dirty usability scale,” Usability Eval. Ind., vol. 189, p. 194, 1996.

[74] A. Colaço, A. Kirmani, H.S. Yang, N.-W. Gong, C. Schmandt, and V. K. Goyal, “Mime: compact, low power 3D gesture sensing for interaction with head mounted displays,” jn Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST ’13, 2013, pp. 227-236.

[75] Google Street View on Google Play: https://play.google.com/store/apps/details?id=com.google.android.street