A Platform for Reaching into the Environment of a Remote Collaborator

by

Xavier Benavides Palos

Bachelor of Engineering in
Electronic Systems Engineering in Communications

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of

Master of Sciences in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

Signature redacted
Author ....
Program in Media Arts and Sciences,
School of Architecture and Planning
May 6th, 2016

Signature redacted
Certified by..... ....
Pattie Maes
Professor of Media Arts and Sciences
Thesis Supervisor

Signature redacted
Accepted by ....
Pattie Maes
Academic Head, Program in Media Arts and Sciences

A Platform for Reaching into the Environment of a Remote Collaborator

by

Xavier Benavides Palos

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
on May 6th, 2016, in partial fulfillment of the
requirements for the degree of
Master of Sciences in Media Arts and Sciences

Abstract

In this thesis we present ShowMe++, an immersive mobile collaboration system that allows a remote user to communicate with a peer using video, audio and hand gestures. We explore the use of a Head Mounted Display (HMD), depth camera and wearable haptic devices to create a system that (1) enables a remote user to be immersed in another first-person's point of view, (2) offers a new way for the remote expert to provide guidance through three dimensional, real-time hand gestures and voice, (3) allows natural interactions with interfaces of Internet of Things (IoT) devices and (4) provides haptic feedback when interacting with remote or virtual interfaces. Using our system, both users feel present in the same physical environment and can perceive real-time communication from one another in the form of 2-handed gestures and voice. We discuss the design and implementation of the system as well as applications scenarios such as remote maintenance, 3D exploration and remote ghost presence. The user study demonstrates that hand transmission, first person point of view and immersion improve the feeling of co-presence and make remote teaching more effective.

Thesis Supervisor: Pattie Maes
Title: Professor of Media Arts and Sciences

A Platform for Reaching into the Environment of a Remote Collaborator

by

Xavier Benavides Palos

Signature redacted

Advisor...

Pattie Maes
Professor of Media Arts and Sciences
Media Lab, MIT

Signature redacted

Reader...

Ken Perlin
Professor of Computer Science
Media Research Lab, NYU

Signature redacted

Reader..........................

V. Michael Bove, Jr.
Principal Research Scientist
Media Lab, MIT

Acknowledgments

This thesis represents the end of three amazing years, full of experiences and adventures. Every little moment helped me to be who I am today. Pattie, thank you for making it possible. You believed in me since I was a "kid", offering me the opportunity of living my dreams. Thanks to Ken and Mike for taking time out to give me feedback and advice. Thanks to the Fluid Interfaces Group that helped me to work and have fun during my time here. I want to express my gratitude to Sang-won and Harshit for long discussions on research, work and life. My nights at the lab would have been too bored without you. I hope that you will continue practicing Catalan. Judith, thanks for working with me in some of the projects that are part of this thesis. Thanks Roy for being a great source of knowledge and advice.

The journey would have not been possible without the help of Juliana and Cristian. You both offered help at work by proofreading my documents, as well as friendship outside the lab. Thanks for including me in your thoughts and conversations. You both became my little family in the US, sharing fun activities and long night conversations. I will miss you a lot.

I would like to thank people that I had the pleasure to meet over the past 3 years. Javi, thank you for being a close friends since I moved to Boston. You have been the biggest support in my life in the US, offering me everything that I needed. You have given me great advice - I always considered it in my decisions. Ivan and Roger, thanks for offering support and joy even during the hardest times. Thanks for making me feel loved and listened. Martin and Bianca, thank you for taking time to hang out with me. You guys always cheer me up, I wish you all the best.

Thanks to everyone that helped me to discover new places in Boston. Thanks Juliana Cherston, Natasha Jaques, Andrew Schonhoffer, Peter Beshai, Maitane Iruretagoyena, Chang Long, Marc Exposito, Ethel Bermejo, Amna Carreiro, Adria Recasens and many more!

Finalment, gràcies papa, mama, Manel i avis. Sempre m'heu permès somiar i poder perseguir els meus somnis. MOLTES GRÀCIES!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

TODAY more than ever, people are distributed around the world. We live in a globalized world that drives us to have friends and family living in different parts of the planet. Companies are increasingly distributed as well. Most of the projects in today's companies involve people distributed in a large range of offices around the world. Moreover, the exponential growth of Internet helps to find worldwide customers for our products. This is one of the main drivers for improvements and innovations in communications technologies. Using these systems, we can stay in touch and can work together remotely.

Today we enjoy several commercial teleconferencing technologies that are cheap, portable and more immersive than traditional voice-only phone calls. Most of the commercial systems tend to focus on face-to-face communication, offering a window to another part of the world. They allow users to have an easy channel for discussions and meetings. Some commercial tools generate virtual meeting rooms such as Skype or Hangouts, or create shared worksheets (like Google Docs or Dropbox).

However, a lot of collaborations involve physical tasks or shared 3D workspaces. When it comes to performing a physical task collaboratively, existing technologies offer limited ways to convey gestures or interact in the remote user's environment. One problem derives from the users' need to have both hands to interact with their environment, while also using their hands to hold their devices in order to share their workspace with the remote participant. When using teleconference systems on smartphones or tablets, people tend to switch between the front and back camera or they may put the device in a fixed position so as to have freedom of movement [1]. In most cases, the user has to move the camera around in order for the remote person to perceive the entire scene.

In this thesis, we will focus on the subset of remote collaborations problems that involve physical tasks and virtual or physical interfaces. Both of these require the usage of natural interfaces to interact with the content. We propose the use of hands gesture transmission, first person point of view and wearable haptic devices to increase the effectiveness of remote collaborations for physical tasks. We run different user studies in order to prove the contributions of each proposed feature in the remote collaboration experience. We report our findings along with conclusions in the User Study section of this thesis.

This thesis introduces ShowMe++, a solution for the proposed problem in the form of an immersive, mobile system that enables easy communication between remote users about a physical task and remote interfaces. The prototype is designed for use by two users (Figure 1-1). For convenience we will refer to the user who would like to share his point of view and receive guidance as the "novice" user, and the one who is remotely assisting as the "expert" user, even though the roles may well be reversed. Both users wear a HMD and perceive the same view, namely the first-person point of view of the novice. The novice user has two cameras attached to their HMD and the expert has a depth sensor camera attached to theirs HMD. In a future version, we would want the headsets of both parties to have identical sensors, so that roles can easily be reversed. This set up allows the remote expert to perform hand gestures that are superimposed in real-time onto the novice user's immediate environment (Figure 1-2). The novice sees his/her own hands as well as the hands of the expert in real-time in their immediate physical surroundings.

Figure 1-1: In the left picture, the expert uses hand gestures to teach a novice user (right picture). The picture in the middle shows the field of view of both users.

ShowMe++ can be used in a variety of applications where remote collaboration is useful, for example in remote maintenance of complex machinery, training how to operate devices or collaborations around IoT devices. Our proposed system looks to aid in reducing travel expenses by allowing manual problems to be solved collaboratively at a distance.

We have built a proof of concept system to investigate how we can provide a more useful system for remote assistance with manually oriented tasks. The system shares the point of view of the novice user with the expert who in turn can make their hands inhabit that space so as to offer assistance. It also provides haptic feedback to the users through a robotic wearable device attached to their arms. The system is portable and the users also communicate between them using audio.

1.1 Motivation

This thesis is motivated by the idea of not having to split the attention between receiving remote instructions through a communication platform and performing these instructions in the real world. Instead, we render in the novice's visual field the hand gestures of the remote expert who is providing the instructions. One of the advantages is that the users can be more focused on solving the problem as they don't need to split their attention. Moreover, both users feel closer to the problem and to each other as they are immersed in the same space and looking at the problem from the same point of view.

Figure 1-2: Examples of different hand gestures performed during the user study

Our proposed system is especially useful for applications such as set up, repair and maintenance of home appliances, cars or industrial equipment, where we need to perform complex actions that can be only taught by the small nuances performed with our hand gestures.

1.2 Approach

We express emotions and messages unconsciously using arms, hands and legs. This information is known as body language. We perceive these messages while we are talking face to face to someone, and it changes our perception of the verbal message [2]. Moreover in a collaborative setup, the information that we express with our hands goes beyond body language. We use them to put more emphasis on something, to point at things and model gestures or simply to make the information clear to the listener. However, current technologies used for collaborations do not focus on this important aspect of human communication. Rather, commercial tools for remote cooperation offer abstract features such as voice, cursors, spotlights or comments.

With the growth of social media, we increasingly share our daily experiences and moments with others. Sharing these instants, we enable others to feel the same emotions and sensations. We upload our experiences in a first person point of view with systems such as GoPro or Google Glass. However, current technologies for collaboration are mainly focused on supporting face to face interactions. We hypothesize that viewing the scene from the same point of view as the remote user will make remote collaboration more effective in that the remote expert can see what the novice is doing from the same perspective and can give visual as well as auditory instructions augmented onto the novice's field of view.

Touch is often depicted as the sense which cannot be deceived [3], it is how we convince ourselves this is real. As Margaret Atwood [4] writes in 'The Blind Assassin': "Touch comes before sight, before speech. It is the first language and the last, and it always tells the truth". However collaboration technologies don't pay attention to this sense, even though it could help improve accuracy and immersion.

ShowMe++ offers solutions for including hands representation, first person point of view and haptic feedback into a platform for remote collaboration. It explores how the hands of a remote expert can reach into the environment of a novice user to teach actions using gestures, manipulate virtual interfaces (Figure 1-4) or interact with physical interfaces (Figure 1-3) of IoT objects while receiving haptic feedback. The novice user sees the live 3D hands of the remote expert in her field of view. The

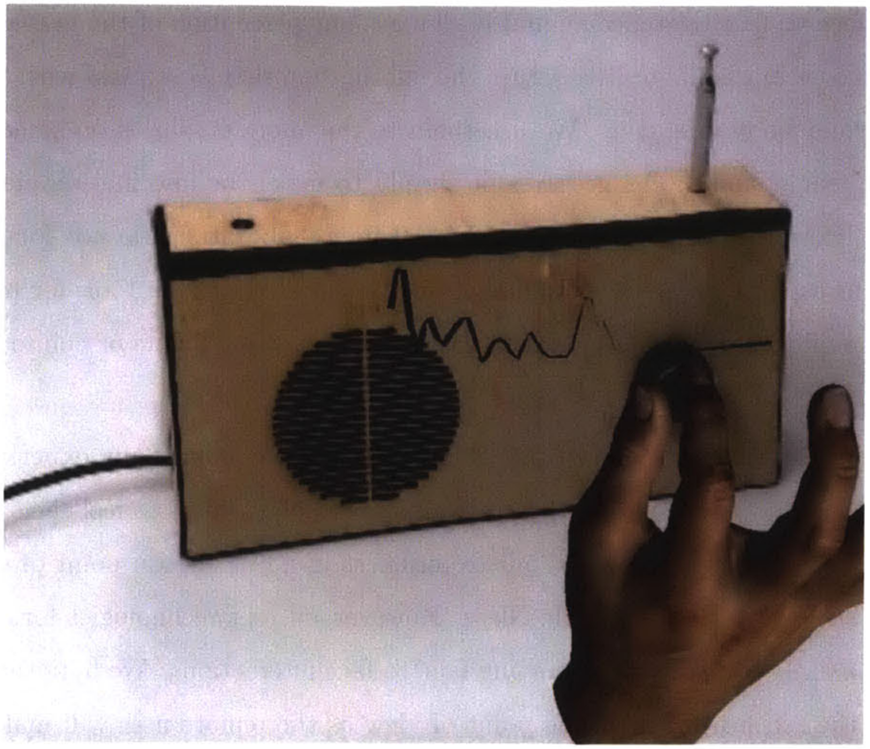

Figure 1-3: Virtual hand of a remote expert manipulating the knob button of an IoT radio

expert user wears a device on his arm that provides haptic feedback for the interactions with remote interfaces such as push buttons or knob buttons. The haptic feedback provided by the wearable device allows the user to increase the accuracy and precision of the interactions.

1.3 Contributions

This work offers several contributions to the field of remote collaboration technologies:

(1) The implementation of a hardware prototype of ShowMe++, achieved by modifying existing devices adding a depth sensor and cameras.
(2) The software implementation for real-time integration of 3D hand gestures into virtual and augmented reality spaces.

Figure 1-4: User manipulating a virtual interface and receiving haptic feedback for the interaction

(3) The implementation of a robotic wearable for haptic feedback.

(4) The evaluation of the system through two user studies/evaluations.

1.4 Road Map to the Thesis

Chapter 2 reviews the related work in the area of collaboration systems that focus on face to face collaboration as well as shared workspaces. Chapter 3 describes the user experience design of our system and discusses the hand representation as well as the first person point of view and the haptic feedback provided by the system. Chapter 4 discusses the implementation of the system detailing the overall system architecture with all its individual parts such as the tracking, network implementation as well as the virtual representations. Chapter 5 presents two user studies to evaluate the effectiveness of hand gestures transmission, first person point of view and haptic feedback in immersive collaborative systems. In Chapter 6 we give an overview of possible usage scenarios for our system. Finally Chapter 7 concludes the thesis with a discussion of conclusions and future work.

Chapter 2

Related Work

This thesis falls at the intersection of Human Computer Interaction (HCI) and Computer Supported Cooperative Work (CSCW). Our system is built on top of several years of research by the academic community. Irene Greif and Paul M. Cashman coined the term CSCW in 1984 at a workshop [5].

One of the most common ways of analyzing a CSCW system is to consider the context of a system's use. Table 2.1 reproduces the CSCW matrix introduced by Johansen in 1988 [6]. It considers work contexts along two dimensions: one, collaborations in the same place or remote collaborations, and two, whether users collaborate at the same time or not.

Our work falls into the category of different place and same time communication systems. In the following sections, we will describe the related work and interactions used in synchronous distributed communications systems.

	Same place	Different place
Same time	Face-to-face	Telephone
Different time	Sticky note	Letter

Table 2.1: Johansen's CSCW matrix for classifying types of remote collaborations

2.1 Shared Spaces

Commercial videoconferencing systems (Figure 2-1) are abundant (Skype, Google Plus Hangouts, Cisco WebEx Conferencing, etc.). Most of these systems enable face-to-face communication from disparate locations but they do not allow remote users to share and reference a common physical workspace. Researchers tried to overcome this limitation using a variety of approaches [7], ranging from projected interfaces [8] [9] to HMD technology. Pioneering work aimed to create an interactive shared drawing surface (Figure 2-2) that both users could work on [10], [11]. Nevertheless, researchers still try to understand and build tools to support collaborative work so as to create a more heightened sense of physical co-presence.

Figure 2-1: Cisco videoconferencing system for remote meetings. It supports face-to-face communication between users.

Over the past several years, researchers have introduced video communication systems (Figure 2-3) that support collaborative work by remote users in a shared virtual space [12], [13],[14], [15]. These systems integrate depth sensors that analyze body interactions and create a shared depth mirror that allows users to work together in the same space. Unlike ShowMe++, these systems are not focused on sharing a first person view of the content and they are not mobile.

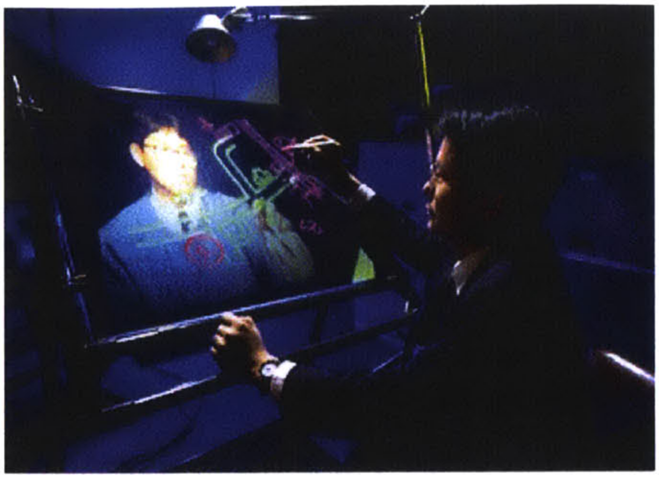

Figure 2-2: ClearBoard is a pioneering system supporting collaborative drawing in a shared space while seeing each other.

Figure 2-3: WaaZam is a networked video environment for people to build their own worlds and have shared experiences at a distance.

Some researchers have focused on proving the efficiency of gestures in shared workspaces. Kirk et al. [16] determined that gestures and visual information improve the speed and accuracy of remote collaboration activities, and Fussell et. al. [17], demonstrated that collaborating users rely more on visual actions than on speech. Tang et. al. [18] confirmed that 35% of the gestures performed in a collaborative task are performed to engage the other user and express ideas. This research motivated our work on ShowMe++ to enable gestural communication in a remote collaboration system.

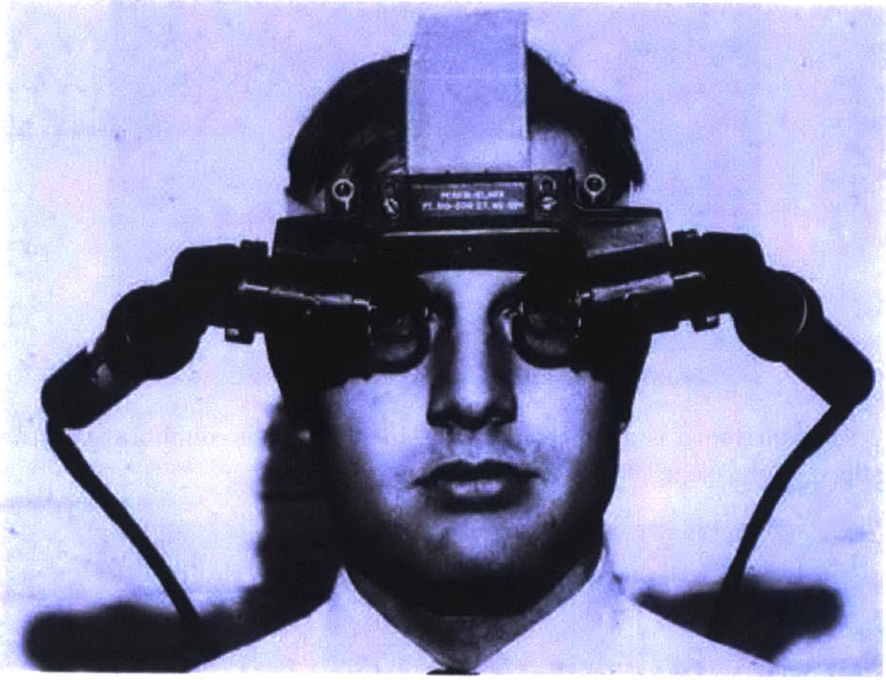

Figure 2-4: The first virtual reality HMD linked to a computer-generated world. Ivan Sutherland designed and presented it in 1961

HMDs have lately attracted considerable attention as a human interface technology (Figure 2-4), even though they have been researched since the 60's [19], [20]. One of the most closely related projects to ShowMe++ is the JackIn project (Figure 2-5), which explores integrating a first person view with out of body vision for human-human communication [21]. One of the users wears a transparent HMD and shares his view to a remote user who sees an out of the body view displayed on a static desktop monitor. The difference between JackIn and ShowMe++ is that we create a mobile system where both users are wearing HMDs and are immersed in the same view, instead of having an out of body view. In contrast with JackIn and other related research [22], in ShowMe++ both hands can be tracked and displayed using 3D hand models (instead of using a flat graphical user interface that supports tele-pointing).

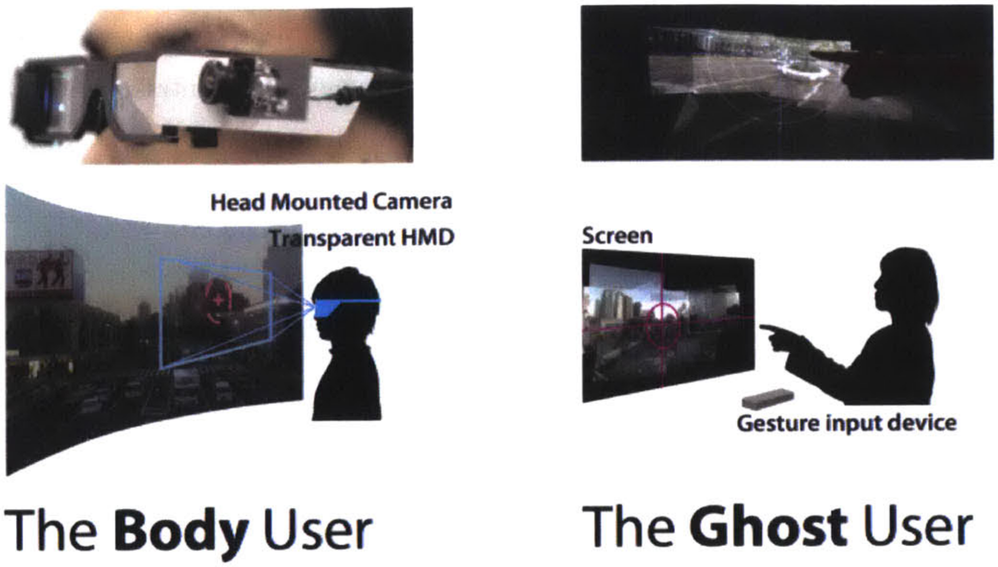

Figure 2-5: JackIn is an immersive experience transmission architecture with a wearable omnidirectional camera for "Human to Human Telepresence"

Another interesting project related to our work is "3D Helping Hands". In this system both users are fused in the same 3D rendering space and the remote expert uses an HMD to perform hand gestures in a shared virtual space [23]. However, in contrast with ShowMe++, the novice user is not using an HMD and has to look through a screen, therefore s/he is not immersed in the same view as the expert.

BeThere [24] is another closely related project (Figure 11). It explores the use of 3D gestures and remote spatial input without any type of HMD. BeThere allows users to leverage a basic pointing gesture and orientation of the finger in order to control a virtual 3D hand. One limitation of BeThere, besides the fact that it only detects one gesture and one hand, is that the whole device is too heavy to hold, forcing the user to use a monopod. The difference between ShowMe++ and these systems is that in ShowMe++ the user is able to perform full hand gestures with both hands and that the novice and expert are completely immersed in the same experience. Moreover, we designed a portable, wearable setup with the novelty of incorporating a wide field of view that tracks full movements of both users' hands and shares this data over the Internet. A similar setup is used in recent work by Oda et. al. [25] which presents a system to guide the user to the perfect viewpoint of an object. To achieve their goal, they employ a 3D model and several pre-recorded angles. In contrast, ShowMe++ proposes a mobile system that can be used in new environments and relies on hand gestures to perform remote collaborative work.

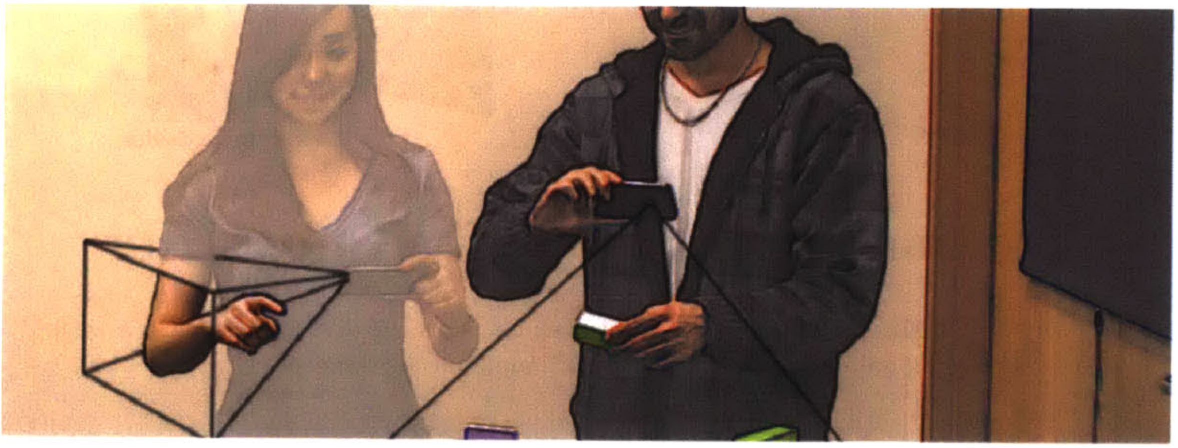

Figure 2-6: BeThere3D is a system designed to explore 3D input for mobile collaborative interactions.

2.2 Haptic Feedback

Our work is informed by theories about the sense of touch and its role in sense of self, memory and education. A good overview of the sense of touch, its physiological and neurological basis can be found in [3]. They cover in detail topics such as tactile perceptual organization, tactile attention, the social aspects of touch and technologies of touch (including virtual reality). Embodied cognition offers us new ways to think about bodies, mind and technology [26]. When a person holds a tool, his sense of self extends to absorb the end-point of the tool [27].

Another important application area for touch and haptic feedback is in education. Tanhua et al. have proved that touch is one of the most important senses in everyday life, however it has been not used extensively to support learning at school. For that, they have conducted a study with 8th grade physics students [28].

Some notable projects that recreate the illusion of touch at a distance are inTouch [29], which is a system that creates the illusion that two people, separated by distance, are interacting within the same physical environment (Figure 2-7). Another project by Brave et al [30] presents a new approach to enhance remote collaboration based on touch and physicality. Physical Telepresence [31] presents a shape display as a shared workspace for remote collaboration.

Figure 2-7: inTouch is a project to explore new forms of interpersonal communication through touch

There are many approaches for implementing haptic feedback in virtual and augmented reality. Early and ongoing efforts have focused on creating haptic gloves [32], which react when the user reaches for a virtual object. Another solution that has been extensively used to provide haptic feedback is the PHANToM, a six-degree of freedom force feedback device that has been used by the VR and AR community. The system developed by Araujo et al. [33], uses a robotic arm attached to a table to provide feedback in VR. They allow interactions with a VR object beyond position and texture. Meyer et al. describe an ultrasonic and electrostatic surface haptic device, which can create tactile perceptions of surface features or textures [34]. A similar approach was implemented on friction-based touch displays in the work of Israr [35]. AIREAL [24] is a haptic technology, which delivers tactile sensations in free air, without requiring the user to wear a physical device (Figure 2-8).

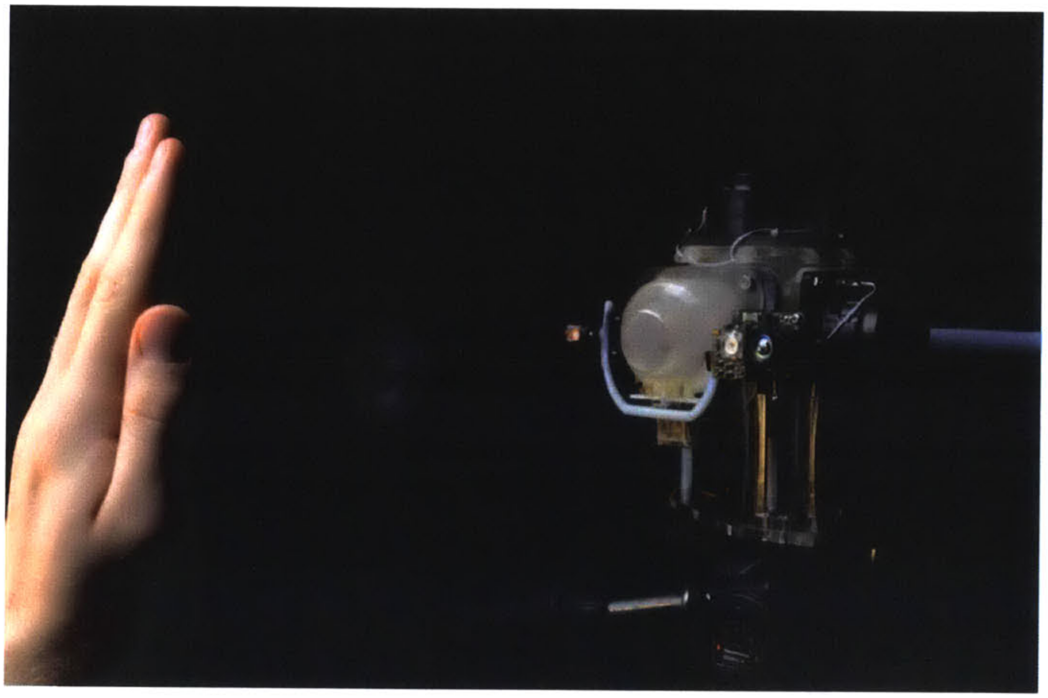

Figure 2-8: AIREAL is a low cost, highly scalable haptic technology that delivers expressive tactile sensations in mid air.

Our system differs from previous solutions in that it provides a wearable that allows a user to interact accurately with different user interfaces. The device is focused on specific physical affordances such as knob buttons and push button. Moreover, it has the form factor of a wearable device, enabling free movement around the scene.

Chapter 3

System Design

SHOWME++ is a remote collaboration platform that has been evolving in order to support different scenarios for collaboration. In this section we explain how the system has been extended, and the different experiences that are currently supported. All of the proposed improvements share two main features: hand gesture transmission and support of remote collaboration.

The initial system was created to improve remote collaboration around physical tasks. Later, the system was updated with new software features that enabled remote control of interfaces of IoT objects. Then, the same hardware and a lot of the same software was used to create shared virtual spaces for collaboration. Finally, we added haptic capabilities to improve the accuracy and precision for remote interactions performed with the hands.

3.1 Hand Gesture Transmission

Hand gestures are part of our daily life (Figure 3-2). We use them to point to things, hold objects, feel textures, express emotions and even as a temperature sensor for liquids or surfaces (Figure 3-1). However, digital platforms for collaboration don't give hands the importance that they deserve.

Figure 3-1: Hands are often used as sensors to measure the temperature or to feel textures of objects around us.

Hands are especially important in collaborative and learning environments. We cannot restrict the freedom of movements that can be performed in such environments, as an infinite number of movements are performed to show actions or instructions. Therefore, the system needs a depth sensor to capture any hand gestures from the user and a means for reconstructing the hands' poses in the virtual environment. The depth sensor needs to be placed in such a way as to meet certain requirements:

- It needs to be placed in a spot that offers mobility to the user and system, and
- It needs to be placed in a comfortable location for the user to perform gestures.

After considering different options, we decided to attach the depth sensor on the front part of the HMD. This location offers a great area and range of interaction for the user (Figure 3-3). It also covers the area where users perform most of the actions

Figure 3-2: An expert uses his hands to teach ceramics to a novice user.

with the hands. Therefore, the user can turn their body around and the system will continue tracking the gestures.

3.2 ShowMe: Basic System for Sharing First-person POV

The first version of ShowMe++ consists of two remote setups with HMDs [36]. One side is used by a novice user and is equipped with 2 cameras in order to reproduce a live feed of the environment into the HMD. The other end is used by an expert user and is equipped with a depth sensor (Figure 3-4). Both users are immersed in the same video feed. We use the depth sensor to capture the hands of the expert, and then superimpose the virtual copy of the hands onto the novice user's field of view.

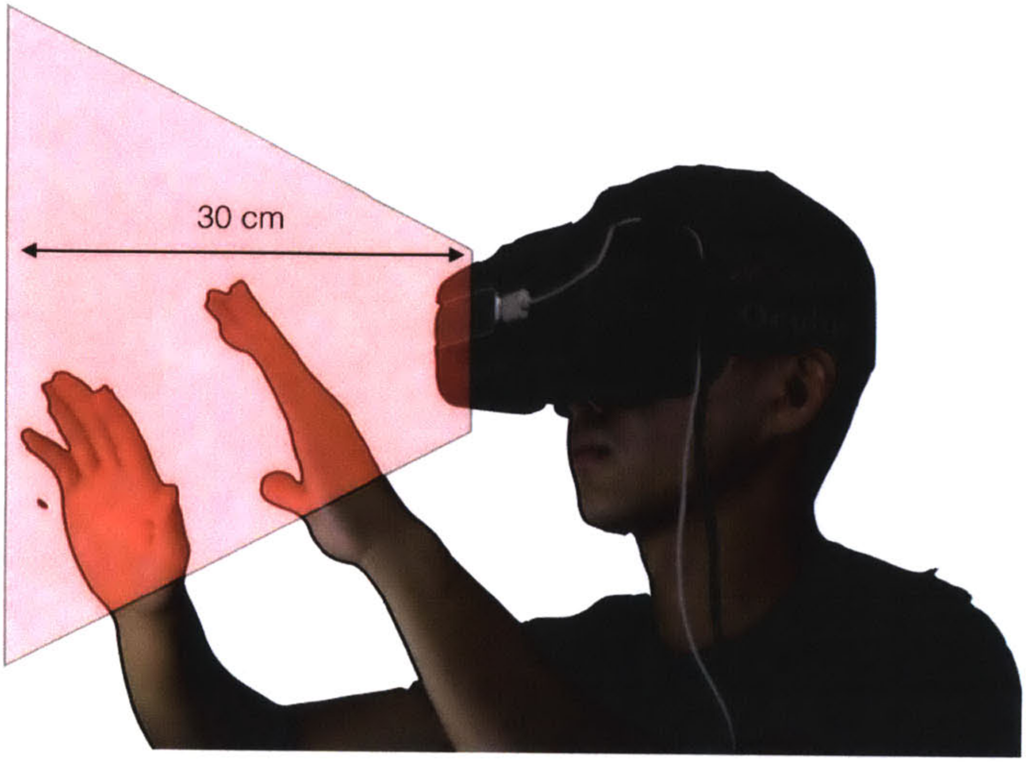

Figure 3-3: Area of interaction for the user wearing the system.

The simplicity of the hardware involved, and the freedom of interactions and gestures that can be performed allow the system to be used in any environment with Internet connection (Figure 3-5). The system is also compact and as such can fit into a back pack or briefcase and used anywhere, anytime.

3.3 Remote-IO: Adding IOT Integration

Lately, we have seen an emergence of Internet connected devices in our houses and workplaces. This opens up the possibility for a remote expert to interact remotely with these devices and fix a problem. We updated our system to make it possible to make actual changes to the physical interfaces of remote IoT objects simply by making hand gestures. This way the system can fill the gap between remote distance

Figure 3-4: Diagram of basic ShowMe++ system.

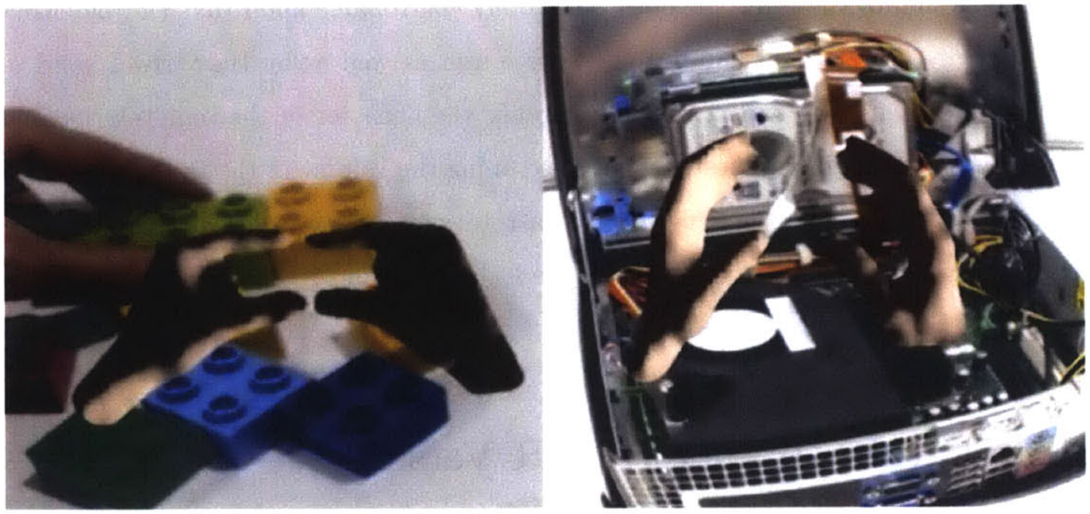

Figure 3-5: The image shows different scenarios with the overlaid hands used for teaching.

and physical presence through remotely controlled Internet devices.

This version of the system is built on top of the previous ShowMe++ system, but adds to it support for remote operation of Internet connected devices [37], [38]. The remote expert is able to operate devices in the novice's environment and bring about physical changes by using the same hand gestures they would use if they were physically present. We built a smart radio where the knobs of the radio can be controlled by local and remote user alike (Figure 3-6).

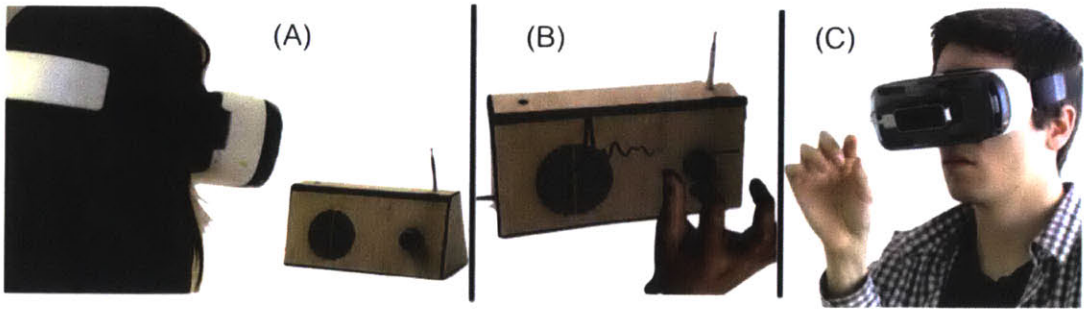

Figure 3-6: The smart radio prototype can be remotely controlled by an expert.

The design of the radio includes a remotely controllable knob that can be either operated by the physical hand of the novice user as well as by the virtual hand of the remote user. The expert user can modify the behavior of the remotely located radio by performing hand gestures such as tuning the volume of the radio by placing the hand over the knob and performing a rotation gesture with the fingers. Once the system detects such a gesture, the knob button will physically move and turn according to the virtual hand gesture.

3.4 DiveIn: Immersive VR Version

Motivated by the huge increase of 3D data provided by sensors, and the difficulty to collaborate remotely around such content, we decided to adapt the ShowMe system to support these types of situations. We developed a compact and portable system to collaborate around entirely virtual 3D models, such as geological models and simulations. The so-called DiveIn (Figure 3-7) system enables 2 or more workers to immerse themselves in a virtual 3D model and interact with that model using natural hand gestures.

Users can use simple hand gestures to perform operations such as zooming in or out and navigating in the virtual model. They can perform other gestures to enter data, change views, etc. Users are aware of the location, the gaze and the hands of the other persons currently working with the model.

To create this system, we used the same hardware as the previous systems. However, we updated the software to create a virtual world that immersed both users. First, we import the data into the environment and then users can interact with the content through natural hand gestures.

3.5 Feel-IO: Adding Wearable Haptic Feedback

After testing the ShowMe system with different people and scenarios, we realized that haptic feedback plays a crucial role in making the experience realistic and effective. Certain actions, such as when we are teaching instructions to a collaborator, require haptic feedback in order to be executed correctly and accurately. Haptic feedback is also relevant, if not necessary, for interacting effectively with a remote physical interface such as in the Remote-IO version of the system. We built an integrated wearable robotic device into our platform to provide portable haptic feedback to the user (Figure 3-8).

We designed a system that can be worn on the user's arm, and is hidden when we don't need to use it. Once the user is reaching for an interface, the wearable device opens up and provides the needed physical affordance for such interaction. The current implementation offers knob buttons, push buttons and regular 3D objects such as cubes or spheres that can be grabbed in the virtual space (Figure 3-9).

Figure 3-7: A user is immersed in a virtual world with geological 3D models. The user is performing a gesture to the other user who is also present in the virtual world.

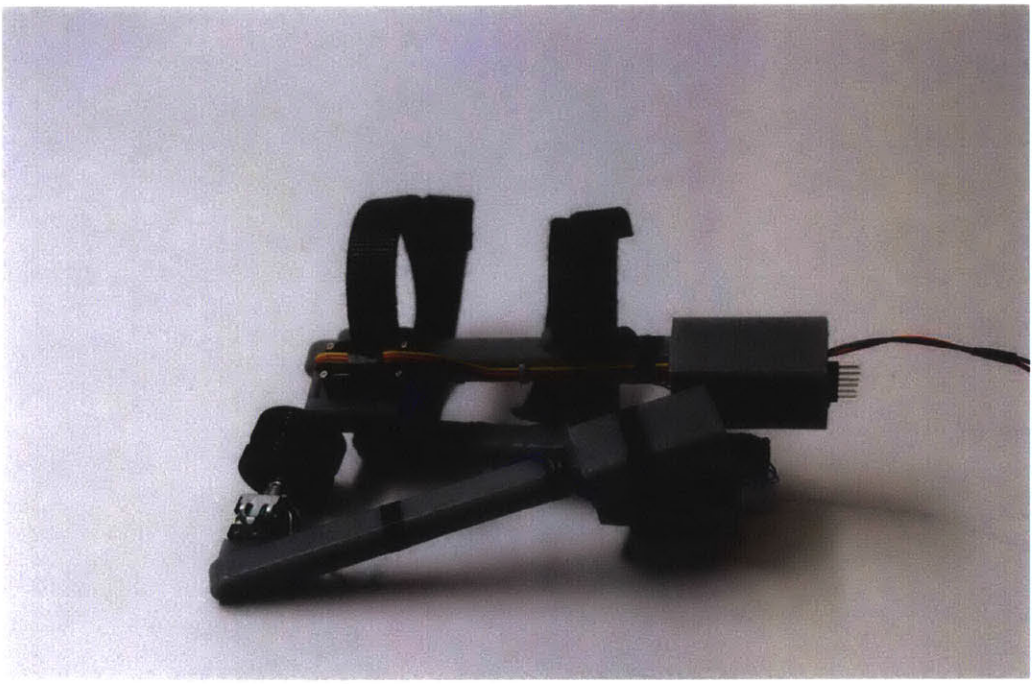

Figure 3-8: The wearable robotics device that is worn on the arm of the user.

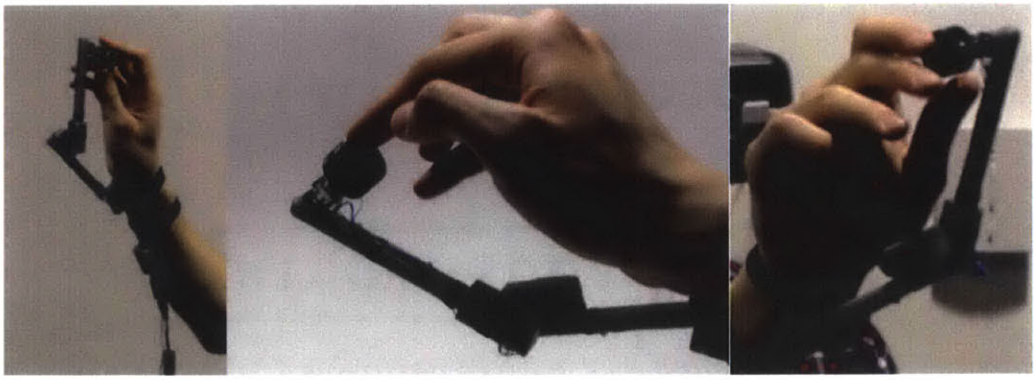

Figure 3-9: The wearable robotics device that is worn on the arm of the user.

Chapter 4

Implementation

THE ShowMe++ system is currently implemented using two Oculus Rifts connected to an Apple MacBook Pro 13", a pair of headphones and the internal camera microphones. For most of the applications, the HMD of the novice user also has two Logitech Pro 9000 cameras attached and the expert user's HMD has a Leap Motion depth sensor attached to capture and track her/his hands (Figure 4-1). For other applications the expert and novice user wear a robotic wearable system that provides haptic feedback for virtual interactions. We designed the system to be mobile, comfortable and small enough to fit in a daily handbag or backpack.

Both computers are sharing information remotely via Wi-Fi, which allows the user to wear the system in mobile situations. ShowMe++ uses the Oculus Rift Development Kit 2 which has a horizontal field of view higher than 90° and a diagonal field of view higher than 110° to create an immersive experience.

4.1 Depth Sensors

The Leap Motion uses optical sensors and infrared light to capture the environment. These sensors are directed upward and have a field of view of 150 degrees wide and 120 degrees deep. It captures images at a rate higher than 60 fps and has an effective range of approximately 25 to 600 millimeters above the device (Figure 4-2).

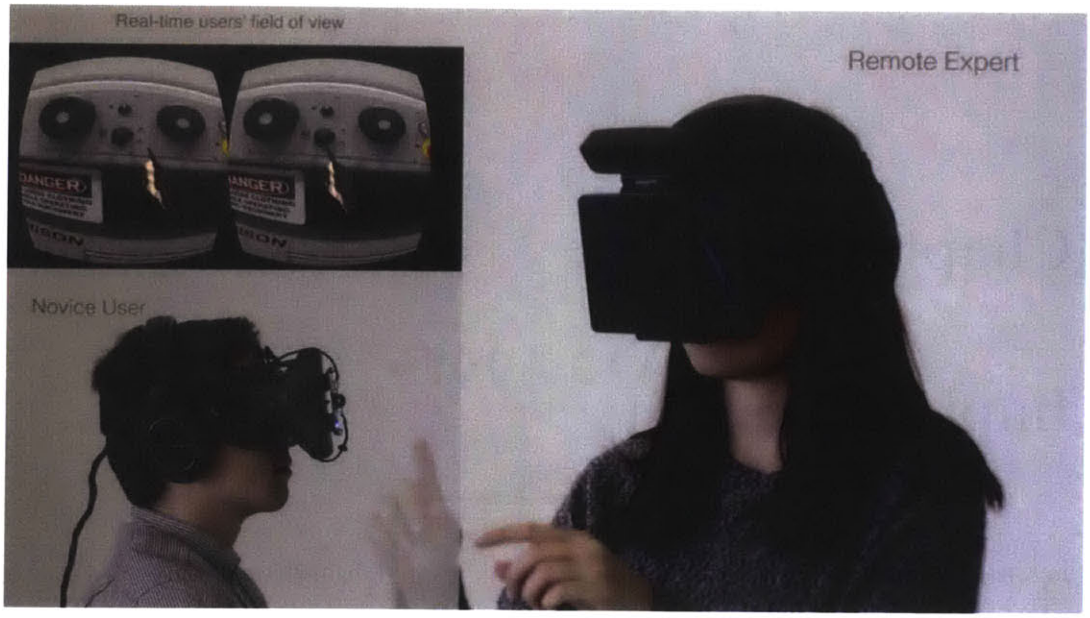

Figure 4-1: This figure shows users wearing the ShowMe++ system. Both the novice's field of view and the virtual hands are displayed in the figure.

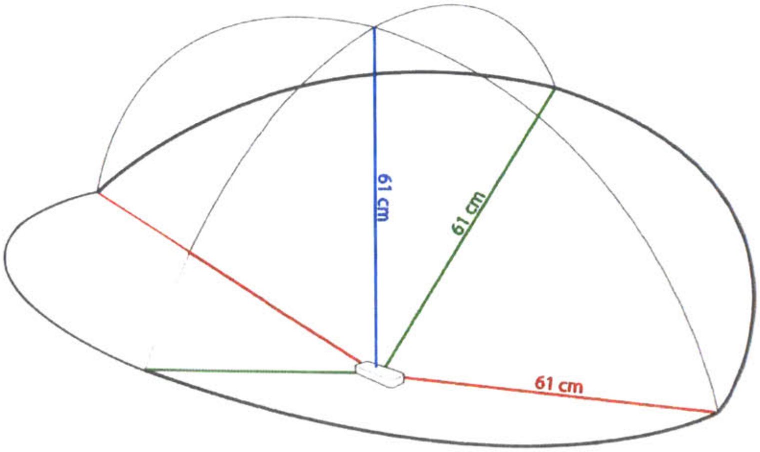

Figure 4-2: Diagram that represents the Leap Motion range of tracking.

To capture the data from the environment, the ray of light enters on the Leap

Motion cameras and the lens bends the light ray so it hits the sensors. The Leap Motion records it as a grey scale brightness value at a specific pixel location. As we can not have perfect lens, the light does not land on the perfect spot of the sensor. Leap Motion has software calibration map (Figure 4-3) that corrects this imperfection, allowing us to calculate the true angle of the original ray of light. Having the distortion-free image, and using the angles from both images, the system can triangulate the 3D location of a feature identified in both images.

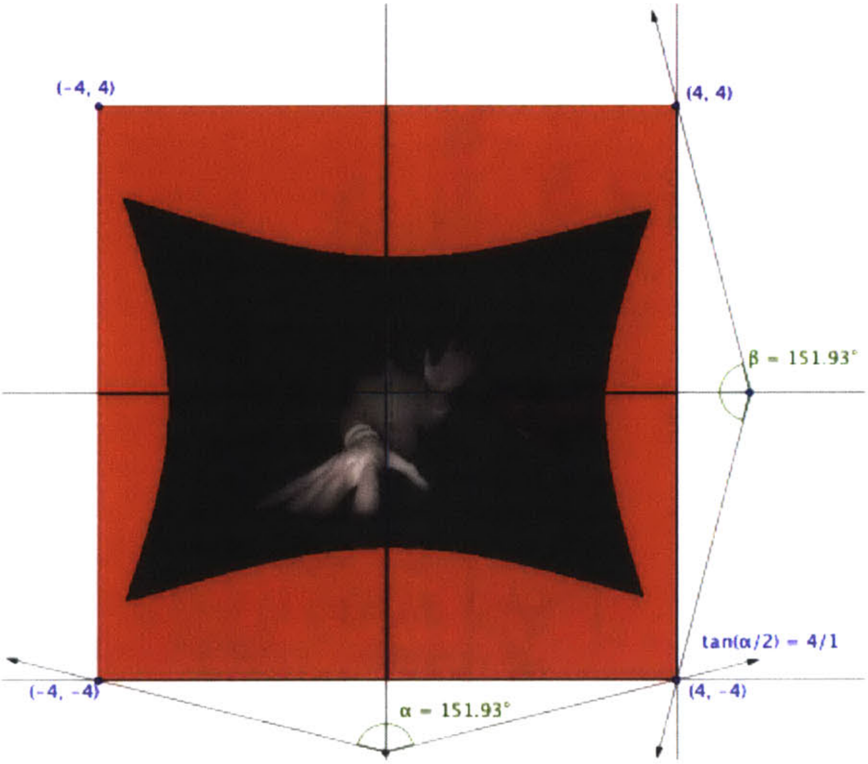

Figure 4-3: Software calibration done by the Leap Motion SDK.

We now understand how the system tracks the position of a feature in the range of view. However, we need the device to track the hands and fingers of the users. To do so, the Leap Motion software uses an internal model of a human hand to predict the posture of the hand and fingers even when they are not visible. The software uses the visible parts of the hand, the internal models, and past observations to calculate the most likely positions of the parts not visible.

The internal model of the hand is represented in Figure 4-4. The system tracks the position of each finger with four points. However, the model for the thumb does not quite match the standard anatomical naming system. A real thumb has one less bone than the other fingers. For ease of programming, the Leap Motion internal thumb model includes a zero-length metacarpal bone so that the thumb has the same number of bones as the other fingers.

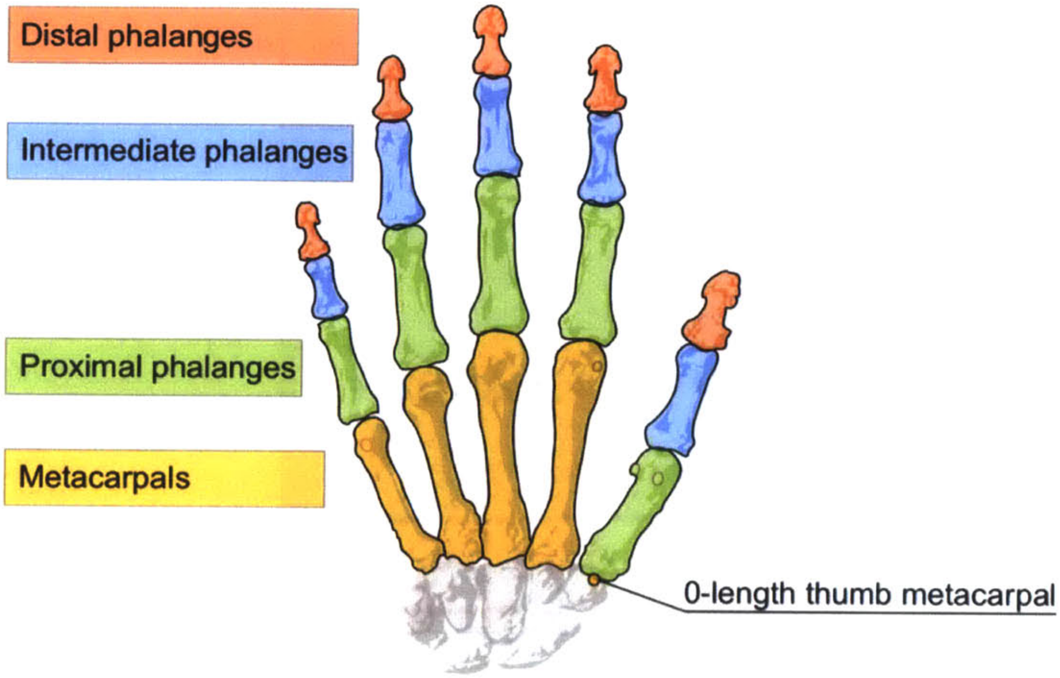

Figure 4-4: Internal model of tracking used by the Leap Motion software.

4.2 Network Protocol

The system is set up with a peer-to-peer connection (Figure 4-5). We use two different methods to share the two types of data: the data from the video streaming and the data exchanged about position, rotation and velocity of virtual objects and hands.

To create a live video feed from the cameras attached to the HMD of the novice user to the HMD of the expert user, we hard-coded a system of UDP Sockets. The 640x480 images are captured frame by frame in the novice's computer, and they are compressed in a JPG format to be sent through the socket. The image is sent and decompressed in the expert's computer, and then it is presented as the next frame of the video for the expert user.

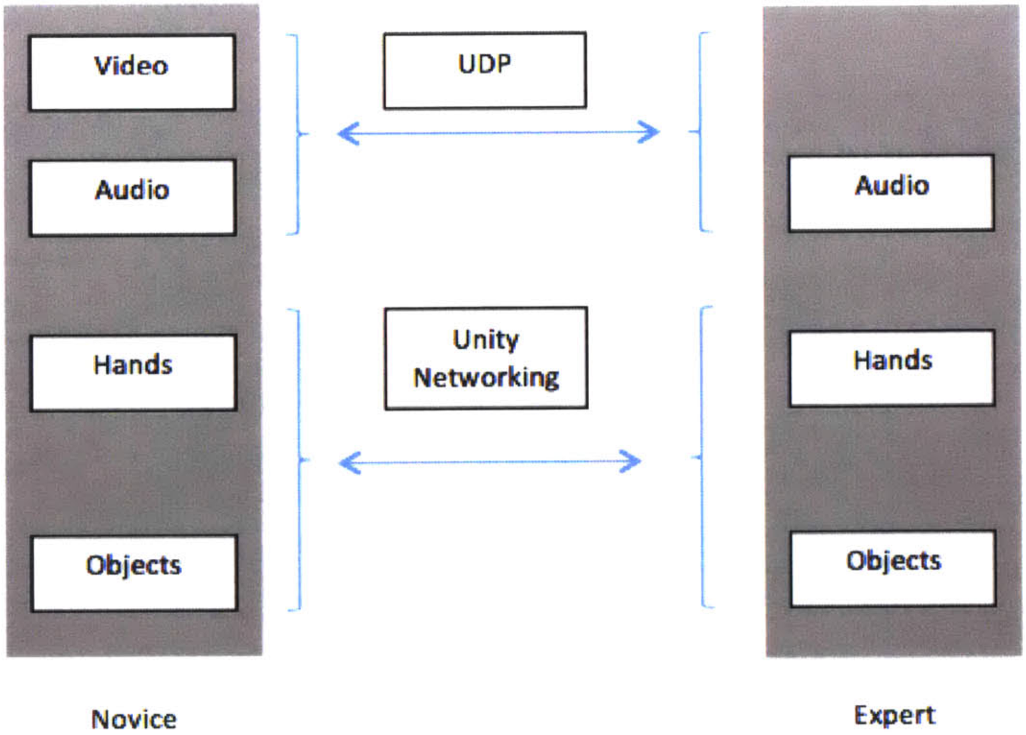

Figure 4-5: Network protocol implemented in the system.

On the other hand, we use the Unity Network Library to synchronize the virtual objects in the scene as well as the hand of the users. It offers us a simple way of connecting the transformation of the objects present on both ends.

4.3 3D Game Engine

We used a 3D game engine to design our experiences and the software for the ShowMe++ platform. We used the Unity Editor (Figure 4-6) as it enables the creation of 2D and 3D games, apps and experiences for multiplatform. There is a 3D virtual space where you design the visual part of the experience. In order to provide intelligence to the scenario, it uses scripts in C# or JavaScript that are assigned to different object in the environment. In our case, we designed the whole experience with C# scripts.

Figure 4-6: Unity3D development screen.

Unity offers a built-in physics engine to handle the physical simulation. However, it is still possible to adjust parameters and change different options of the physics. Moreover, it also includes a rendering engine that offers lighting, cameras, materials, shaders, textures and particle effects.

4.4 HMD

To create an immersive experience we decided to use HMDs that offer the illusion of being in a different world. For our prototype we are using an Oculus Rift DK2 that has a field of view of 110 degrees and can refresh the display up to 90 fps (Figure 4-7).

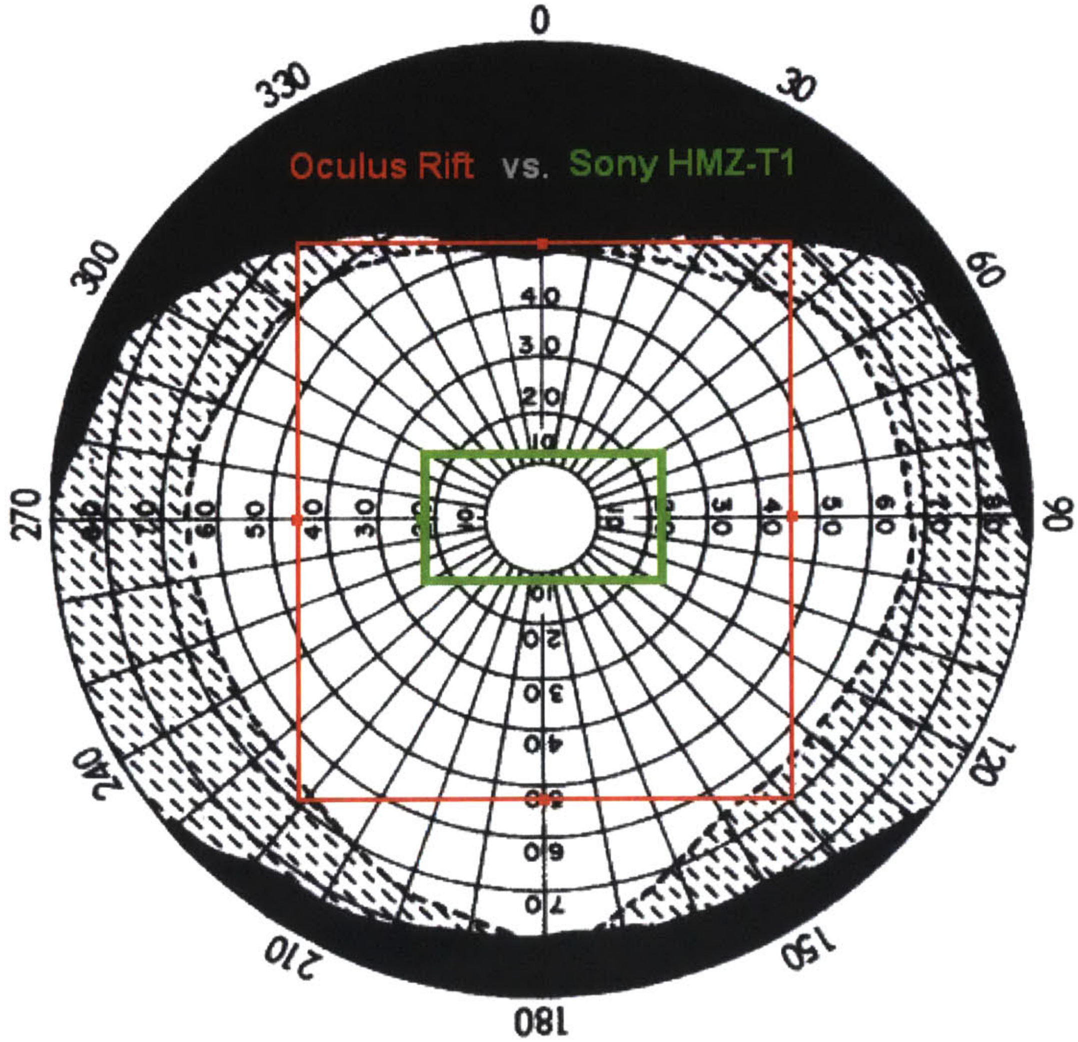

Figure 4-7: Comparison about the field of view between Oculus HMD and Sony HMZ-T1, one of the first personal HMD.

To create the trick of immersion, there are two important things inside the HMD. One of them is the lenses. VR headsets use either two feeds sent to one display or two displays, one per eye. The lenses (Figure 4-8), which are placed between the user's eyes and the pixels, focus and reshape the picture for each eye to create a stereoscopic 3D image by angling the two 2D images.

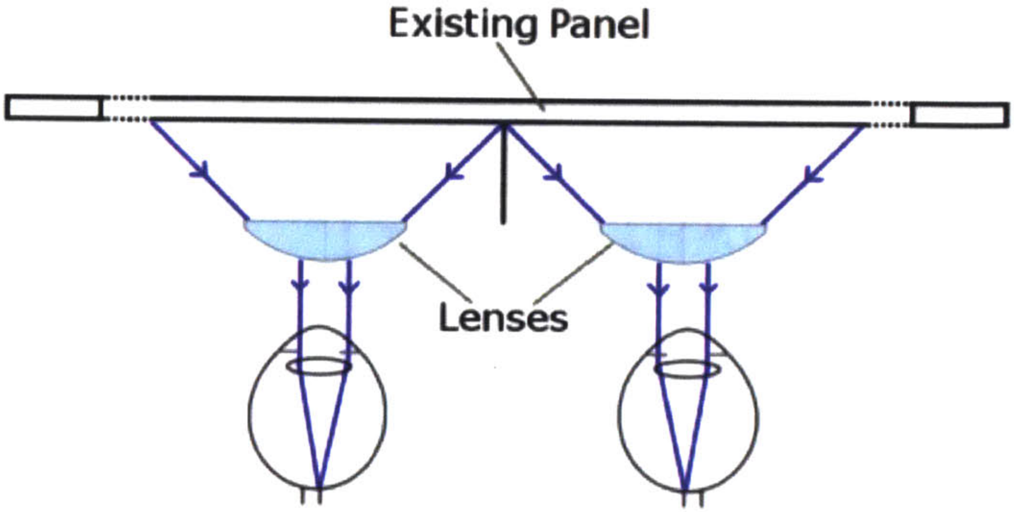

Figure 4-8: The lenses of the HMD are directing the light from the screen to our eyes.

The other important feature is the head tracking. Current HMDs include an IMU that tracks the rotation of the head. It is a critical part of the system, as it provides the smooth virtual rotation when moving our head (Figure 4-9).

4.5 Robotic Wearable

To implement the robotic wearable for haptic feedback, two different aspects had to be considered: the 3D design and the electronics.

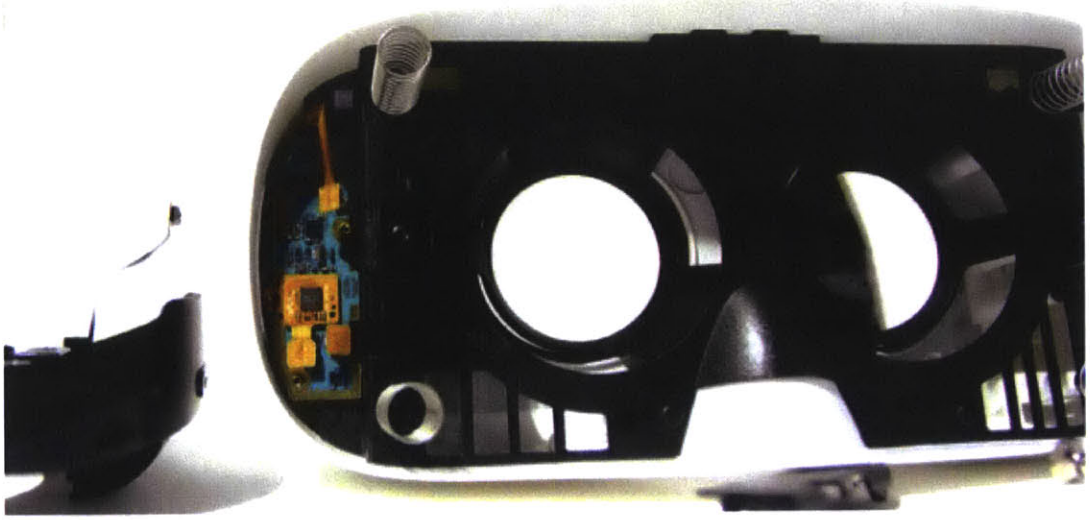

Figure 4-9: IMU embedded in the Gear VR Headset

4.5.1 3D Design

We designed the structure of the wearable using the Rhinoceros CAD software (Figure 4-11). It is designed with three separate arms actuades by two servo-motors. Each arm mesures 22cm long and remains in folded position by default. The arms can unfold to offer capabilities to the user or hide by moving backwards towards the user's arm (Figure 4-10).

4.5.2 Electronics and Communication

The system is implemented with an Arduino Mini Pro that serves as the brain for the wearable device. It sends, receives and processes the information that flows between the computer, the servos and the sensors of the wearable device.

The wearable is made with two servos; three 3D printed resin pieces that work as the skeleton, the physical interfaces and the Arduino (Figure 4-12). Two servos control the movement of separate parts of the arm. We used the Goteck Metal Gear Micro Servos that each can apply a torque of 2.5kg. The servos move the arms to the

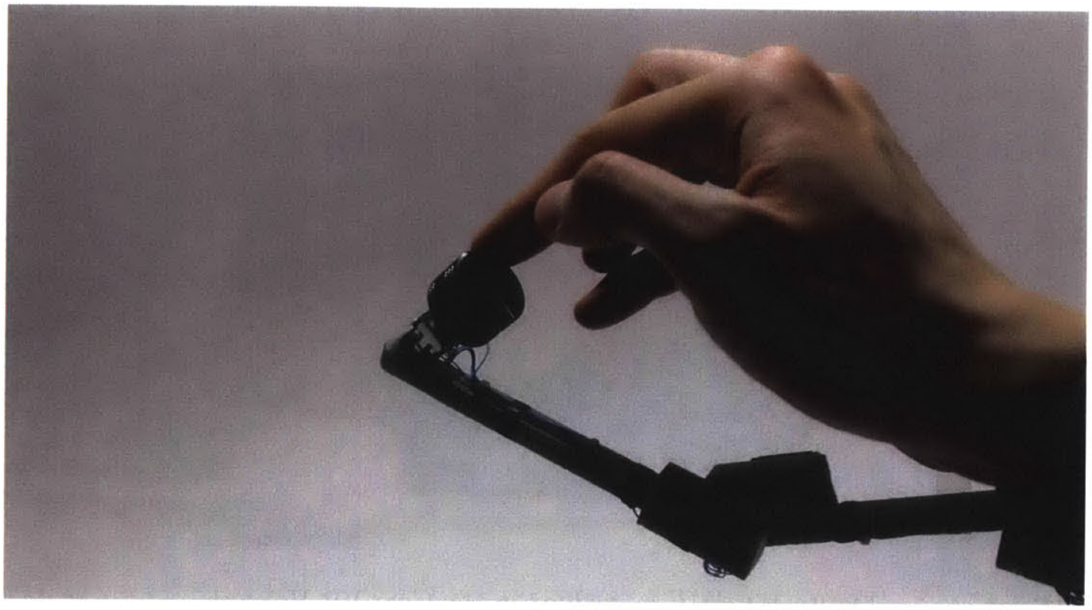

Figure 4-10: Wearable unfolded, allowing the user to interact with the knob button.

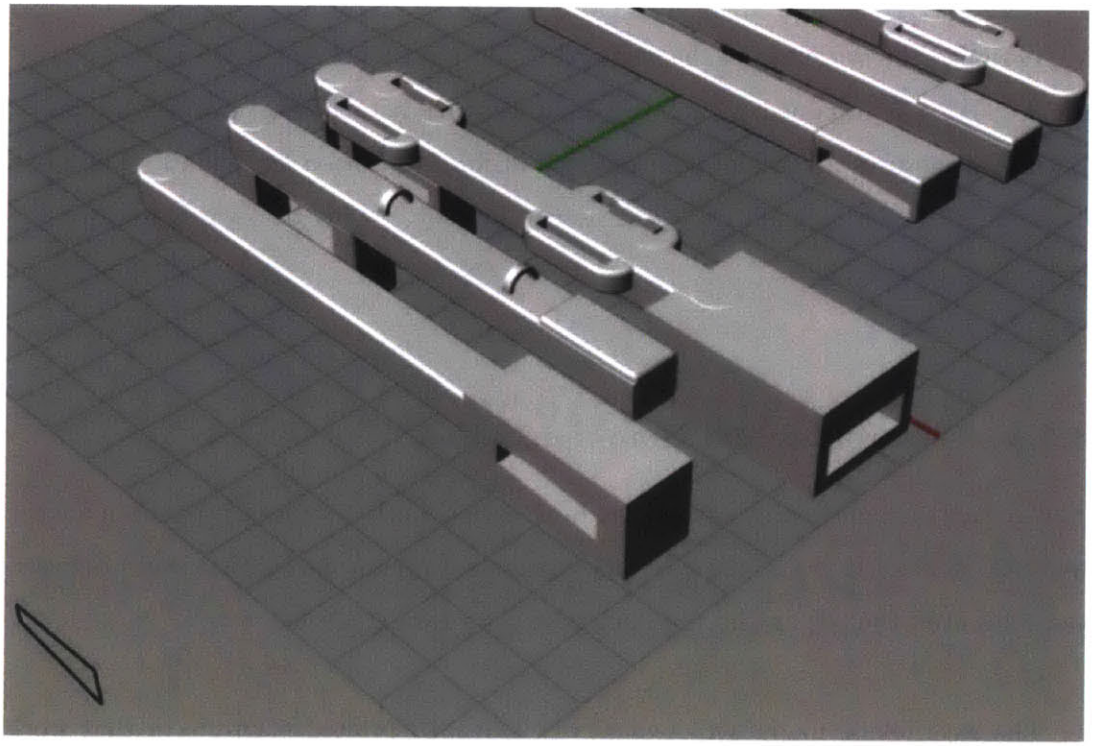

Figure 4-11: 3D design of the wearable haptic device in the Rhinoceros software.

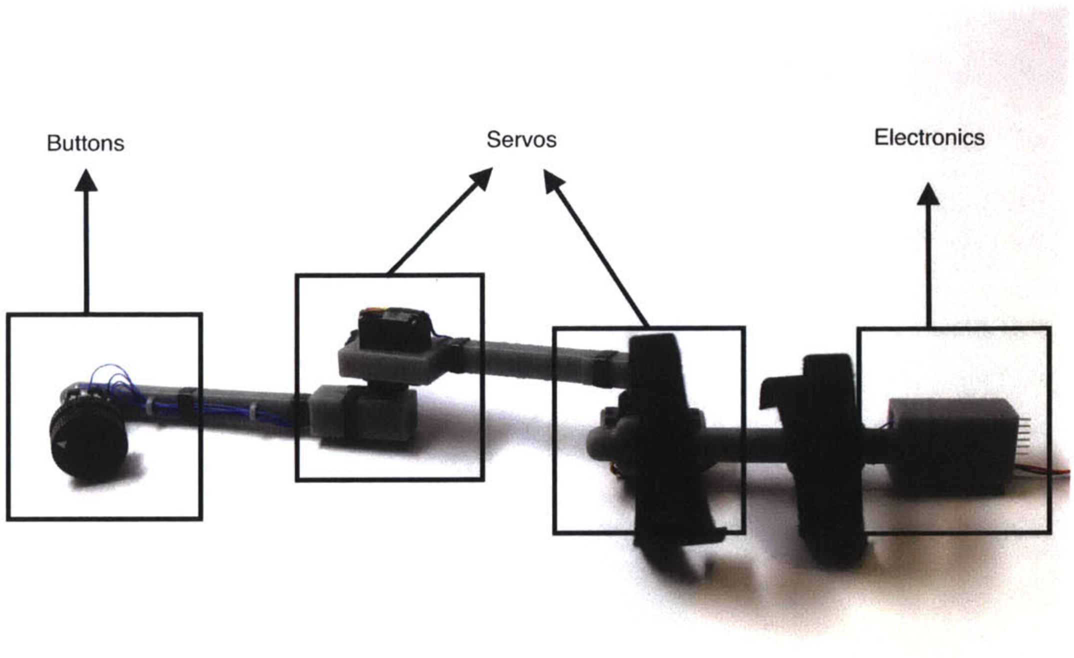

Figure 4-12: Distribution of the different components that are embedded in the wearable device (Button, Knob, Servos and Electronics)

two positions, folded and unfolded on the user's arm. The entire wearable device is powered from a computer's 5V USB port.

The wearable device also offers the haptic feedback while we are grabbing a virtual object. For that, we need to add different object shapes onto the tip of the arm. These physical shapes (cubes, spheres...) are made of conductive material, so we use a capacitive sensor to detect when we are touching it (so the virtual object follow the hand) or when it is not being touched (the virtual object remains still).

Chapter 5

User Study

5.1 First Person Point of View and Hand Representation

We conducted a user study to test if users were able to perform different tasks successfully and to evaluate whether ShowMe++ is an effective system for remote guidance. The emphasis of the evaluation was to test the usability of the system and explore which kind of hand gestures and practices were used in ShowMe++ versus a standard videoconferencing setup. In addition, we were specifically interested in understanding the tradeoffs of viewing through someone else's point of view.

We recorded the participants' behavior during the task performance and compared the experience with the same task when performed on a videoconference system. After performing all tasks, users filled a qualitative questionnaire summarizing their experience and offering feedback for possible improvements.

5.1.1 Participants

Eighteen subjects (11 males, 7 females) from our laboratory were recruited in pairs to participate in the study. All participants were familiar with teleconferencing systems to keep in touch with relatives or for work purposes. Each pair of participants had some type of relationship between them (friends, couple or workmates). The duration of each study was approximately 30 minutes and the ages of the participants ranged from 20 to 28, with a mean of 24.

5.1.2 Set up

The test scenario employed two separated rooms in the same building. Given a pair of users, we randomly selected who would perform what role - the expert user (who was going to teach the tasks) or the novice user (who was going to perform the tasks). In the videoconferencing setup we used two tablets with Skype installed (one for the novice and one for the expert). The two participants were located in different rooms and were able to place their tablets in a mount in order to have freedom of movement. In the case of ShowMe++, the expert user was located in one of the rooms wearing an HMD with the depth sensor attached to it, and in the other room, the novice user was sharing his/her view to the expert through the two cameras attached to their HMD.

5.1.3 Conditions

In the first five minutes of the study, we explained to both users which roles they were going to perform and what the tasks were about. Before starting each task, we privately taught the expert how to solve the tasks, and to memorize and identify what s/he was going to teach the novice user.

Each pair of users needed approximately 30 minutes to complete the user study and questionnaires.

5.1.4 Measurements

We conducted the experiments using a within-subject design. Each task was divided in three different sub-tasks or activities, meaning that each pair of participants performed six sub-tasks using the two systems.

The subtasks were slightly different for each condition, to ensure that the novice user was completely inexperienced every time. Each of the subtasks took approximately 2 to 3 minutes to complete. We describe the three different tasks below.

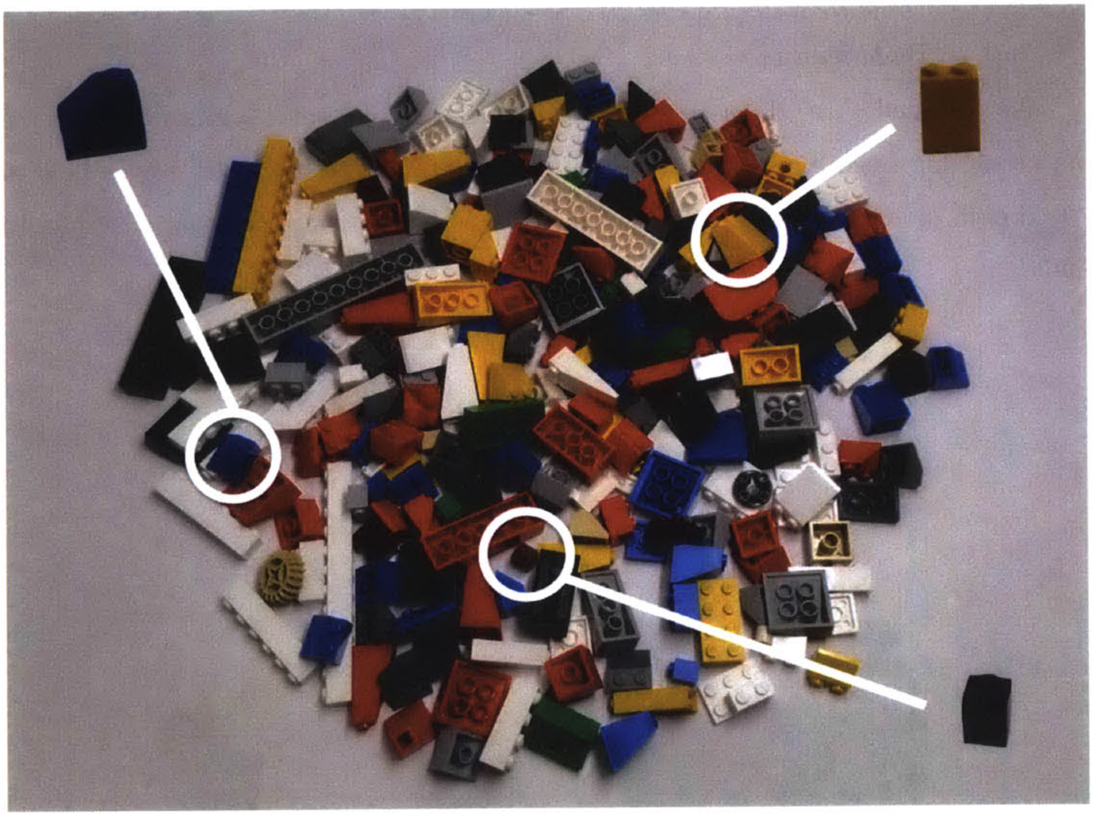

Figure 5-1: User Study - Task 1: 390 Lego Pieces are spread out on a table. The novice user has to locate a specific, unique Lego piece in this pile (the yellow, brown or blue one).

Task 1: Selecting a particular Lego piece

We randomly distributed 390 Lego pieces with an equitable distribution of colors onto a table. Three unique Lego pieces were hidden in three different locations (Figure 5-1). The expert user knew in advance the color, shape and the location of these pieces. The remote expert taught the novice where the pieces were located (one piece for each of the subtasks).

Task 2: Replace components and plug in PC cables

We placed a Dell OptiPlex GX280 desktop computer on a table to be repaired by the novice user. The expert user taught the novice user how to (1) extract two RAM memories, (2) plug-in an 8 pins header and (3) plug-in a 16 pins header into the motherboard.

Task 3: Building a Lego model

We designed three different models with Lego Duplo, some regular Lego pieces and some big Lego pieces (Figure 5-2). The expert had to teach the novice how to build each model. The novice did not have any other guidance on what s/he had to build.

Figure 5-2: User Study-Task 3: Three different types of Lego models, which the expert has to instruct the novice to build.

5.1.5 Evaluation

Once the tasks were concluded, the participants filled out a post-test questionnaire. We analyzed the data collected as well as their comments and remarks. All the users successfully completed the tasks with both ShowMe++ and the Skype system. In the ShowMe++ condition, users were able to guide and receive guidance through hand gestures from the expert and novice user.

Many of the users took time to get used to ShowMe++. During the first minutes they felt disoriented because of the HMD, and did not know whether to use speech, visual cues or hand gestures, which biased the results against it: "it was difficult to coordinate eye-to-hand coordination during the first task" (novice). These results were somewhat expected since ShowMe++ is not as familiar as Skype and we did not give the users much time at all to get used to this new style of interaction. Several of the novices complained about the fact that their view of the workspace was indirect (because of the HMD) and therefore somewhat restricted: "It was a little distracting viewing my field of vision through a digital device and not my own eyes" (novice). In contrast, some of the users acting as experts remarked that a first person point of view offers the best angle to teach another person, as the expert can see where the novice is looking and at the same time offers the same perspective for their own hands.

Despite some of these drawbacks, in these tasks and under these conditions, enabling the viewing of first-person hand gestures helped users to successfully perform the tasks, which suggests that ShowMe++ can be effective for remote collaboration around manual tasks. Overall, users found ShowMe++ to be more helpful for task 1 and 3, and rated ShowMe++ as almost equal to videoconferencing for task 2. Figure 5-3 shows the results of the questionnaire regarding the helpfulness of ShowMe++ and videoconferencing for each task.

These results coincide with the time spent for each pair of users completing the tasks in that users spent on average less time the complete the tasks with ShowMe++ than with the Skype set up. However, given the number of users we tested, these timing results are not statistically significant. Users on average needed 8 seconds less to successfully finish task 1 using ShowMe++ than with Skype (31s versus 39s), 6 seconds more using ShowMe++ in task 2 (48s versus 54s), and 5 seconds less in task 3 (43s versus 48s). Some of the comments were "Task 1 is the best for ShowMe++ scenarios, there were so many different color pieces and sizes, that it was almost impossible to know to which he was referring during Skype" (novice), "Finding the Lego piece was the most difficult task to describe. I couldn't tell which piece was being looked at with just voice and no pointing" (expert). Most users agreed with the usefulness of being able to use gestures to communicate between novice and expert in order to point to objects, indicate sizes or communicate orientation of elements. Users also agreed that ShowMe++ worked less well in Task2. They felt that the resolution of the Oculus Rift was not accurate enough to interact with small components. We will continue updating the hardware with new HMDs, to meet the requirements needed for such interactions.

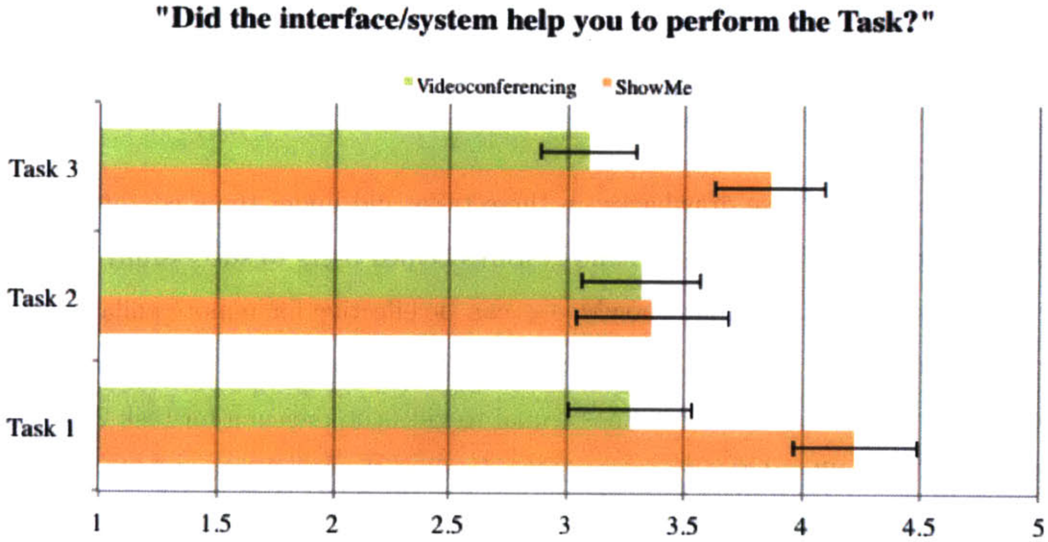

Figure 5-3: Helpfulness comparison between ShowMe and videoconferencing for each task.

One of the curiosities we noticed was that some experts were performing hand gestures even when they were not using ShowMe++ and the remote novice user could not see them. Research has shown that hand gestures play a crucial role in the learning process and enrich communication among collaborators [39]. In the case of videoconferencing, the body language was lesser and slightly more abrupt than with ShowMe++. In contrast, the gestures that experts performed while using ShowMe++

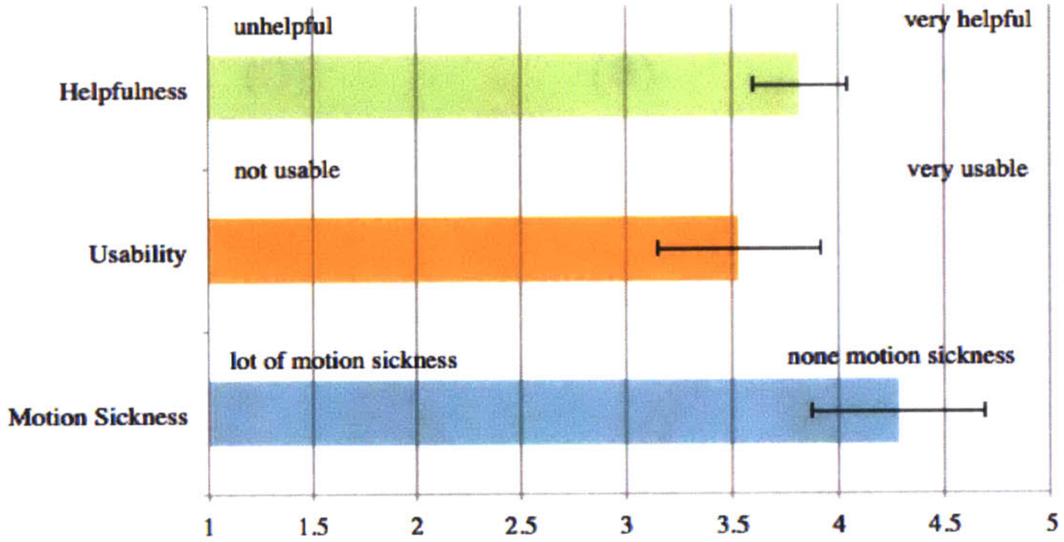

Figure 5-4: Result of the post-test questionnaire. Users rated ShowMe as a system that does not cause motion sickness and that is usable and helpful. Motion Sickness (mean=4.28, std=0.4) Usability (mean=3.53, std=0.38) Helpfulness (mean=4.38, std=0.22).

were much more smooth, specific and clear. The most common gesture was to point at an object (Figure 5-5 - A), followed by the gesture to express that an action was done correctly (Figure 5-5 - B - C). Participants used both hands to show the orientation of a piece, and how to connect multiple pieces (Figure 5-5 - E), and one or two hands to show the size of a specific component or to simulate the shape of it (Figure 5-5 - D). The majority of gestures performed included pointing, size description and orientation, and emphasizing numbers (Figure 5-5 - F).

Finally, users were asked for the helpfulness, usability and motion sickness they experienced while using ShowMe++ (Figure 5-4). Users were also asked which system (Skype or ShowMe++) they would like to use for real collaborative and learning tasks in the future, taking into account that ShowMe++ was a proof of concept and still a research project. 94.5% selected ShowMe++ as their preferred system over videoconferencing. These results suggest that ShowMe++ is a usable and helpful system for remote collaboration tasks that potentially could be used effectively in

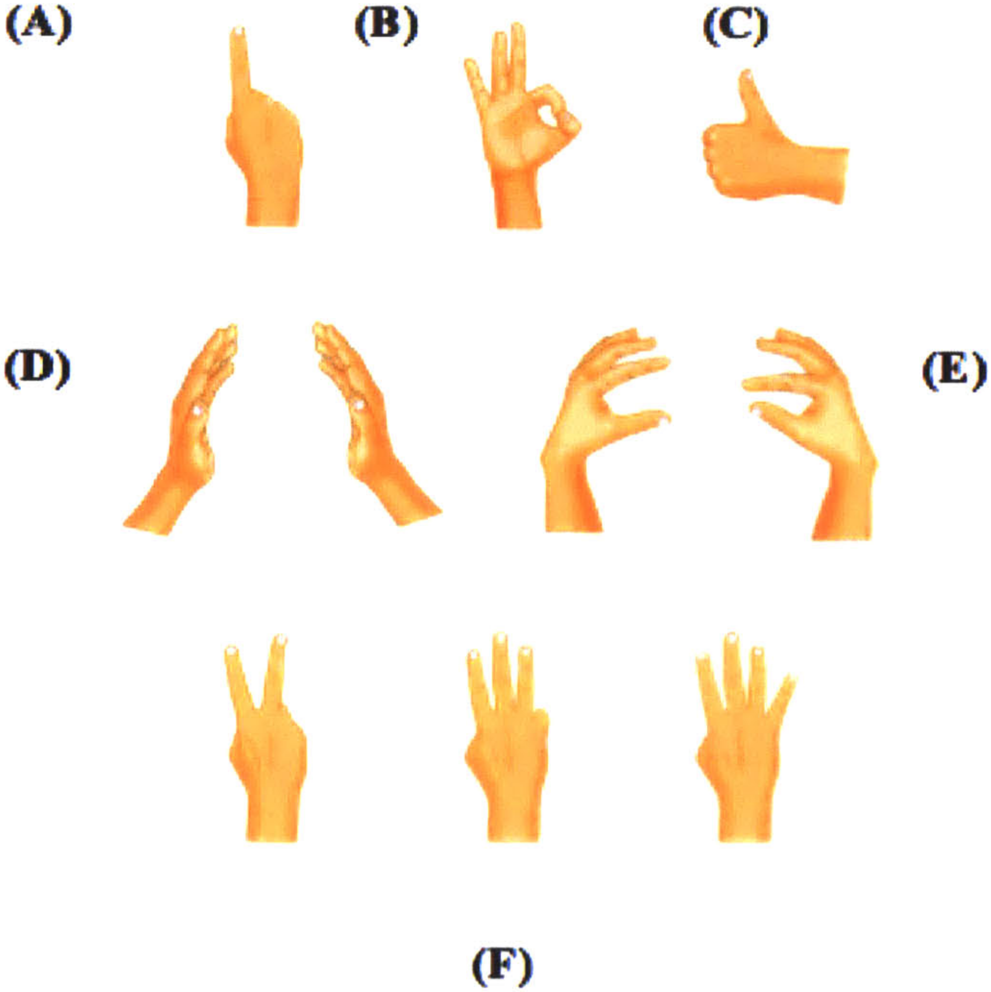

Figure 5-5: Example of hand gestures performed to transmit information to the other user during the user study.

real life applications.

5.2 Haptic Feedback in a Virtual Collaborative

We ran a preliminary evaluation with several users to gather qualitative data about the haptic feedback system. The goal of the study was to let users try the wearable to

(1) analyze if they felt the haptic feedback was realistic and convincing, (2) understand the ease of use and (3) analyze if they felt tired while or after using the system.

5.2.1 Participants

Ten subjects (7 males, 3 females) from our laboratory were recruited to participate in the evaluation. All participants received instructions about the system prior to the trial, and they performed the test in pairs. The duration of each evaluation was approximately 10 minutes and the ages of the participants ranged from 23 to 30, with a mean of 26.

5.2.2 Set up

The test scenario was done in two separated rooms where users where sitting in a stool that can rotate 360 degrees. The users had an Oculus Rift DK2 with a Leap Motion attached to the headset and the wearable haptic feedback device on the arm of their dominant hand. They were immersed in a collaboratively virtual world that had different scenarios, interfaces and objects to grab and interact with.

5.2.3 Conditions

In the first 3 minutes of the study, we let the users try the system to get familiarized with it. Prior to the trial, we explained to them how the system works and what they would feel and experience while in the virtual world. It is important to notice that all the participants had previous experiences with VR.

5.2.4 Measurements

The users were immersed in a shared virtual space with an industrial scenario. The users had three different activities that they had to perform with haptic feedback. The tasks included a knob button turning to a specific value, pushing a specific button, and transporting spheres from one table to a basket (Figure 5-6).

We also asked each participant to perform an individual timed task. Each participant had to tune a virtual radio with and without the haptic feedback. We tested the time improvement offered by the wearable device.

The main goal for the study was to get qualitative feedback about users' feelings and opinions about the system.

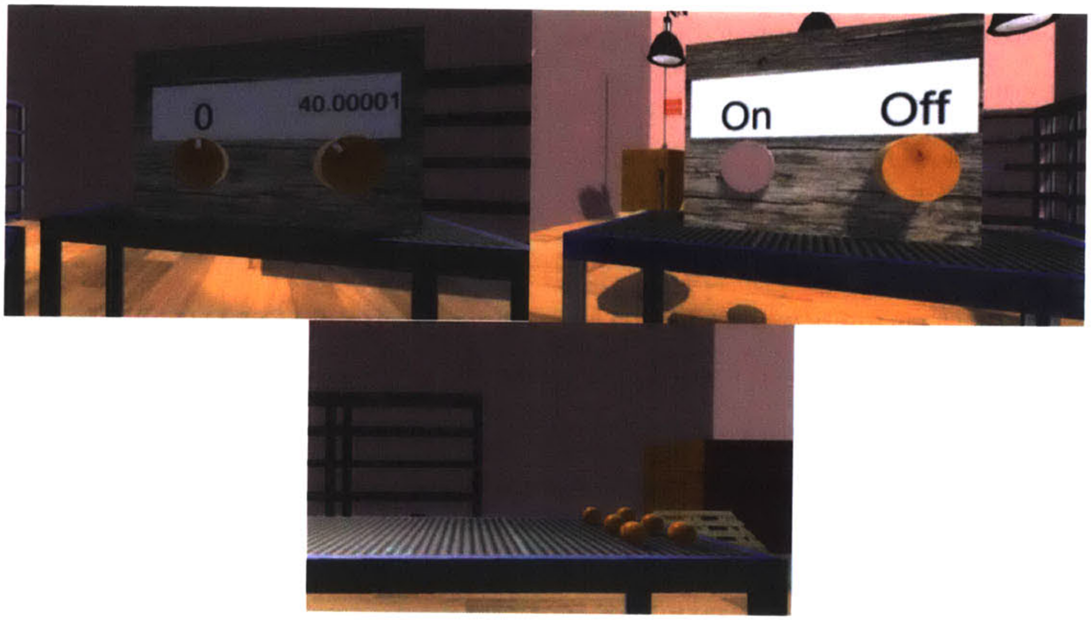

Figure 5-6: Top-left: the knob button interface used in the study. Top-right: the push button interface used in the study. Bottom: the spheres that needed to be moved in the test.

5.2.5 Evaluation

After the test, users were asked a series of questions about their experience during the activities that they had just performed. All of them found the system really easy to use, as they didn't need to learn anything new to interact with VR interfaces. Users were also asked about realism of the interaction and most of them found the interactions very realistic (Figure 5-7). None of the users felt any fatigue using our device. They all agreed that after using the wearable, they got used to it and remarked the improvement offered for actions that require high accuracy in VR.

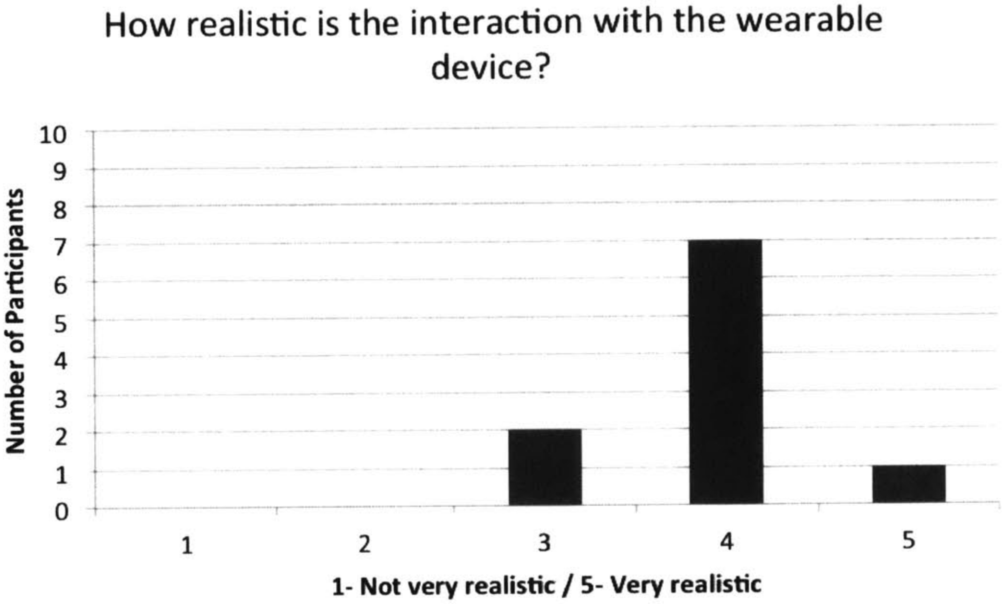

Figure 5-7: The graph represents the answers of the users about realism.

All of the users pointed out that the system provides higher precision with the knob button interface. However, a couple of users found a small misalignment between the position of the real knob button and the virtual one. Nevertheless, it was not a problem for them to interact with it, as they both rated the interaction as very useful. Regarding the use of the knob button interface, one user said "It is really difficult to tune a virtual radio without the wearable", another said "it would be great to use the wearable as an always-available knob button to interact with my TV or speakers".

The users were asked to move virtual spheres from one table to a basket in the scene, with and without the help the wearable device. Some of them noticed that the device does not reproduce the weight while grabbing objects. Although these users wanted to feel the object's weight, they all remarked that the haptic feedback helped them feel more immersed in the system. User immersion and engagement with the system will lead to better collaboration between peers in a virtual workspace.

As part of the experiment, we measured task completion time with and without haptic feedback. We immersed the user in a virtual room that has a virtual radio with an interactive knob button. The angle of the knob button represents the radio station's frequency, which is displayed on top of it. We asked the participants to tune the radio to 102Mhz. It was the first time that these users were interacting with a virtual interface with and without haptic feedback. Therefore, none of them had any previous knowledge relevant for the interaction.

The positive qualitative feedback about the haptic device is in line with the lesser time users spent in completing the task using the device, as opposed to without it. The timing results are not statistically significant due the low number of users we tested. Users on average needed 21 seconds less to successfully finish the task with the haptic feedback than without. (33s versus 17s) The results are shown Figure 5-8.

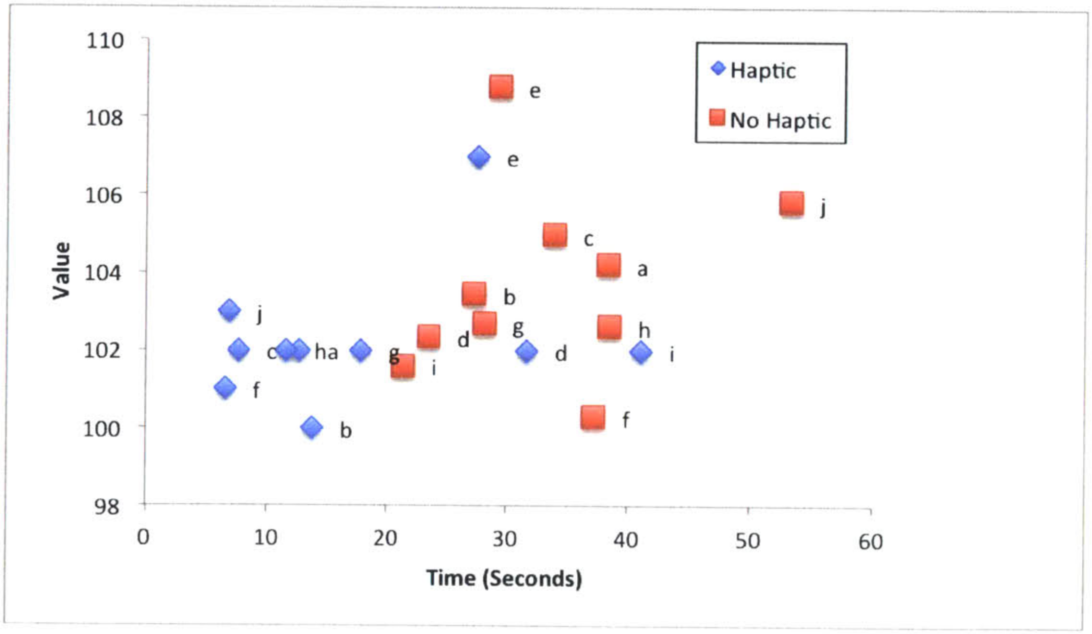

Figure 5-8: In the figure the tunned value is plotted vs. time, where square's mark non-haptic feedback and diamond's are with haptic feedback

We also measured the accuracy users achieved when rotating the knob button. Accuracy was measured as the average of the final value selected by the users. We got an average value of 103.69 (std of 2.41) without feedback and 102.3 (std of 1.82) with feedback. These results mean that the values collected from users with haptic feedback were closer to the desired value (102MHz), which once more is in line with the qualitative feedback received from the users.

Chapter 6

Applications

6.1 Remote Assistance

The ShowMe++ system can be used in a variety of applications for remote collaboration, for example in remote maintenance of complex machinery or in the training process for how to operate devices (Figure 6-1). Our proposed system looks to aid in reducing travel expenses (cost saving impact) by allowing manual problems to be solved at a distance (time saving impact).

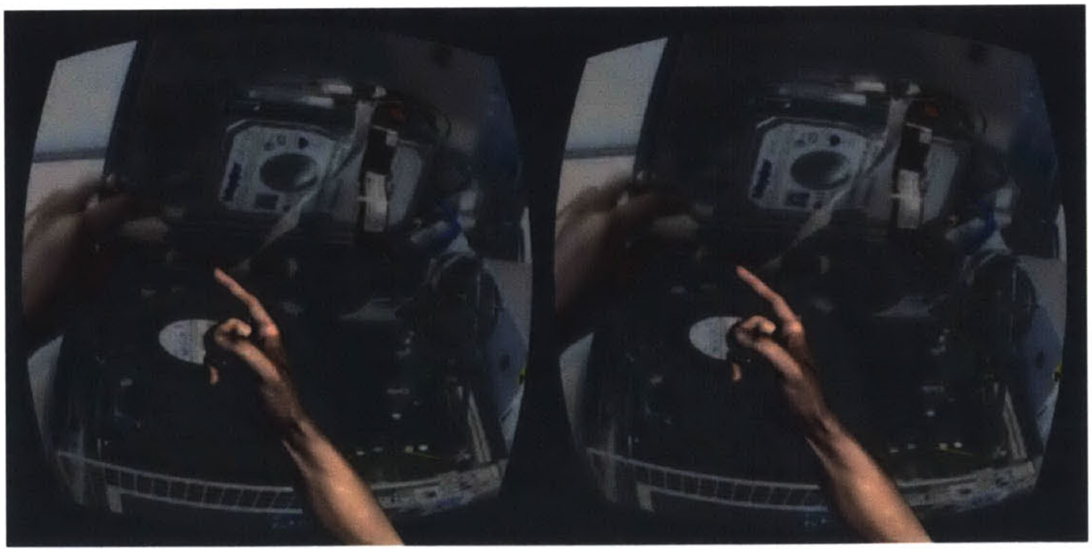

Figure 6-1: An expert remote user is teaching how to replace the RAM memory of a computer.

6.2 Remote Manipulation of IoT devices

Another possible application involves mounting the novice's HMD on a tele-operated robot. It allows the remote user to move around a space and interact with light switches, radios, TV or other home appliances. The local user will see changes in the environment and in the physical position of the interface.

Another application relates to remote help. A remote expert trains the novice operator how to program or use an industrial machine. Both of them wear the system so the remote user can show the operator how to manipulate the smart machine, by performing the actions (Figure 6-2).

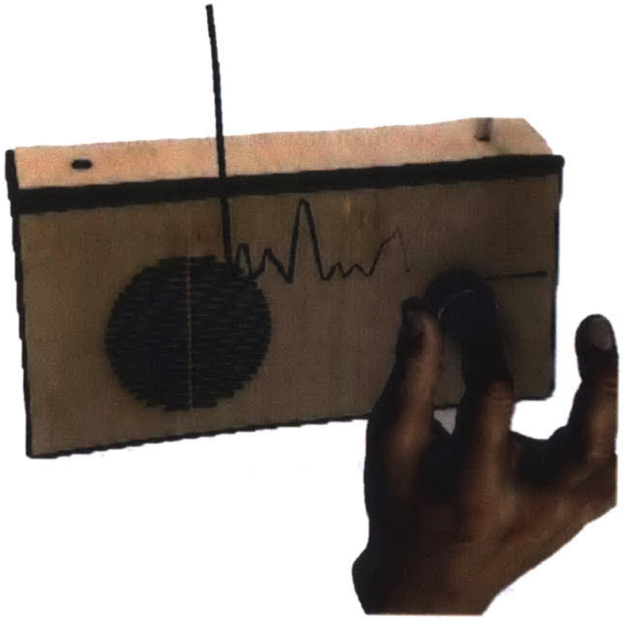

Figure 6-2: A remote expert manipulates the knob button of an Internet connected radio.

6.3 Collaboration around 3D Models

The system can be used to enable workers to immerse themselves into 3D models from any location, using simple, portable equipment (Figure 6-3). They can interact with 3D models in this virtual world along with collaborators using audio and hand gestures. Applications include collaboration around design and architectural models, geological models, biological models and more.

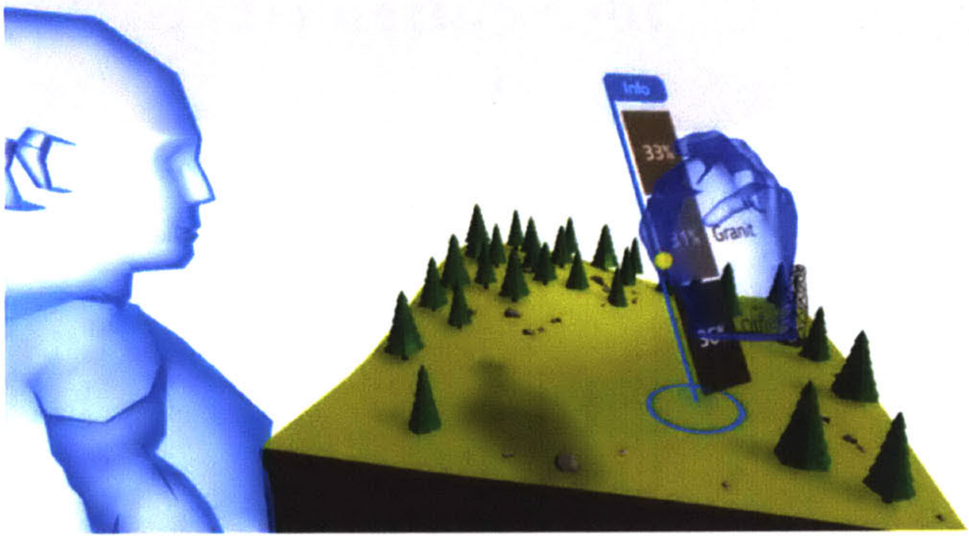

Figure 6-3: Users are interacting around a 3D geological model in a VR environment.

Another application is inspired by the game FoldIt [40], an online puzzle video game about protein folding. The objective of FoldIt is to fold the structures of selected proteins, using tools provided in the game. Since the structures presented are in 3D, we implemented an application for our system, where a user can use their hands to manipulate proteins (Figure 6-4). Moreover, the user is immersed in the environment with another user, and both of them can work closely to solve the game and can see each other's actions.

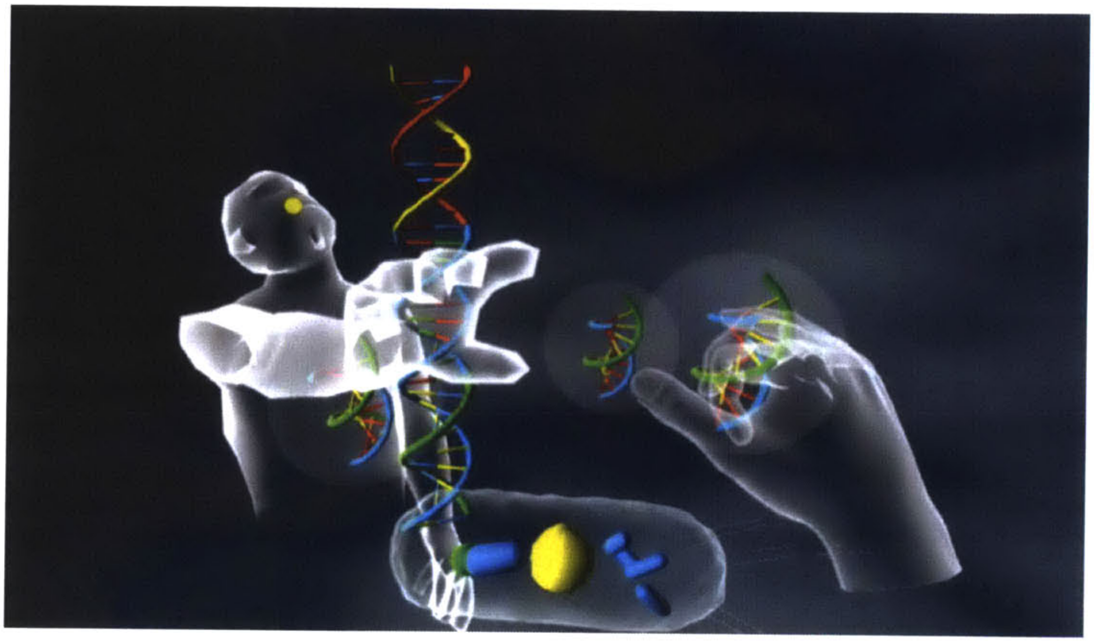

Figure 6-4: Two users manipulating and collaborating around proteins and cells in a virtual shared environment.

Chapter 7

Conclusions and Future Work

We described ShowMe++, an immersive mobile collaboration system that merges the advantages of videoconferencing with 3D hand gestures and haptic feedback to achieve a mixed reality experience. We offer a solution for remote collaboration around manual tasks where both users are immersed in the novice user's point of view, which is augmented with real-time visualization of the expert's hand gestures. We showed through preliminary studies that participants were able to successfully collaborate around remote tasks, and found ShowMe++ to be a helpful and usable system. In particular, the system enabled them to establish a common understanding of the space, objects, sizes and orientations.

We are encouraged by the feedback from the users to continue adding more features to the system. The communication between novice and expert was very effective, and mainly focused on hand gestures and almost no voice. We observed that the richer the variety of hand gestures used was, the more fluent the resulting collaboration. Novice users were mimicking what the experts were pointing at, whereas when using Skype novices relied on the experts' verbal description of objects. One of the most used techniques for novices while videoconferencing was to randomly point to multiple elements waiting for the approval of the expert followed by "is it this one?" or "this?".

Since there is not a captured 3D model of the novice's local environment, the expert's 3D hands can not be properly depth merged into it, and as the results have showed, it is still worth the complexity of creating 3D reconstructed hands rather than a simple plain segmentation of the expert's hands if we do not have a captured 3D model of the novice's local environment.

Some users mentioned that it was difficult to understand other user's emotions during remote collaboration. The reason is because in some scenarios, peers are represented as full inexpressive avatars and in others as virtual hands. In future work, we could capture the states of the users and represent them in the virtual model with the recent advancements in affective computing and physiological sensors. For example, we could change the color of the avatars or hands to represent these emotions. This would offer users deeper insight into the other users emotional state during the collaboration process.

We could also consider analyzing brain activity to improve shared virtual workspaces. Technology can now capture and categorize brain signals. We could collect this information to make virtual workplaces more pleasant for the users.

Study participants mentioned that they would like to use the haptic wearable device for scenarios unrelated to VR. We could use the wearable as an extension of our physical body. It could offer us just-in-time access to interfaces in the physical world. For example, user point his hand at a TV, the robotic arm could unfold and provide an interface to flip between TV channels. If the user moves his hand towards the speaker, the robotic arm could open to provide an interface to change the volume.

As part of the future work, we would like to integrate our system with home appliances and teleconferencing systems. Additionally we plan to study the use of our system in fields such as online education, remote learning, and remote assistance.

Bibliography

[1] Amelia Schladow Nicholas Wallen Francis Zane Federico Casalegno. Leonard Giusti, Kotval Xerxes. Workspace configurations: Setting the stage for remote collaboration on physical tasks. *NoriCHI, ACM Press*, 2012.

[2] Julius Fast. Body language. *Simon and Schuster*, 1988.

[3] C. Spencer A. Gallace. In touch with the future: The sense of touch from cognitive neuroscience to virtual reality. *OUP Oxford*, 2014.

[4] Margaret Atwood. The blind assassin. *McClelland and Stewart*, 2002.

[5] Jonathan Grudin. Computer-supported cooperative work: History and focus. *Computer*, 1994.

[6] Johansen Soren. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 1988.

[7] Keiji Uemura Keisuke Tajimi, Nobuchika Sakata and Shogo Nishida. Remote collaboration using real-world projection interface. *SMC IEEE*, 2010.

[8] Hiroshi Ishii Xiao Xiao. Mirrorfugue: Communicating hand gesture in remote piano collaboration. *TEI, ACM Press*, 2011.

[9] Benjamin Cohen Ran Stone. Pavel Gurevich, Joel Lanir. Teleadvisor: A versatile augmented reality tool for remote assitance. *CHI, ACM Press*, 2012.

[10] Hiroshi Ishii Jobayashi and Minoru. Clearboard: A seamless medium for shared drawing and conversation with eye contact. *CHI, ACM Press*, 1992.

[11] Scott L. Minneman John C. Tang. Videowhiteboard: video shadows to support remote collaboration. *CHI, ACM Press*, 1991.

[12] Anthony Tang Kori Inkpen Seth Hunter, Pattie Maes and Susan M Hessey. Waazam! supporting creative play at a distance in customized video environments. *CHI, ACM Press*, 2014.

[13] Stefan Agamanolis and V Michael Bove Jr. Multilevel scripting for responsive multimedia. *IEEE multimedia*, 1997.

[14] Ermal Dreshaj. Holosuite: an exploration into interactive holographic telepresence. *Massachusetts Institute of Technology*, 2015.

[15] Seth Hunter Nick Gillian Pattie Maes Anette von Kapri, Kevin Wong. Inreach: Manipulating 3d objects remotely using your body. *CHI, ACM Press*, 2013.

[16] Danae Stanton Fraser David S. Kirk. The effects of remote gesturing on distance instruction. *CSCL, ACM Press*, 2005.

[17] Susan R. Fussell Darren Gergle, Robert E. Kraut. Action as language in a shared visual space. *CSCW, ACM Press*, 2004.

[18] John C. Tang. Findings from observational studies of collaborative work. *Man-Machine Studies*, 1991.

[19] Ivan. E.Sutherland. A head-mounted three dimensional display. *AFIPS, ACM Press*, 1968.

[20] C.Comeau J.Bryan. Headsight television system provides remote surveillance. *Electronics*, 1961.

[21] Jun Rekimoto Sunichi Kasahara. Jackin: Integrating first-person view with out-of-body vision generation for human-human augmentation. *AH, ACM Press*, 2014.

[22] Matthew Turk Tobias Hollerer Steffen Gauglitz, Cha Lee. Integrating the physical environment into mobile remote collaboration. *MobileHCI, ACM Press*, 2012.

[23] Weidong Huang Franco Tecchia, Leila Alem. 3d helping hands: a gesture based mr system for remote collaboration. *VRCAI, ACM Press*, 2012.

[24] David Forsyth Brian P.Bailey Rajinder S.Sodhi, Brett R.Jones and Giuliano Maciocci. Bethere: 3d mobile collaboration with spatial input. *CHI, ACM Press*, 2013.

[25] Steven Feiner Barbara Tversky. Ohan Oda, Mengu Sukan. Poster: 3d referencing for remote task assistance in augmented reality. *3D User Interfaces*, 2013.

[26] D. Kirsh. Embodied cognition and the magical future of interaction design. *TOCHI, ACM Press*, 2013.

[27] Bolognini N. Ladavas E. Frassinetti, F. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, 2002.

[28] Pystynen J. Raisamo R. Tanhua-Piiroinen, E. Haptic applications as physics teaching tools. *In Haptic Audio-Visual Environments and Games (HAVE), IEEE*, 2010.

[29] Dahley A. Brave, S. intouch:a medium for haptic interpersonal communication. *CHI, ACM Press*, 1997.

[30] Ishii H. Dahley A. Brave, S. Tangible interfaces for remote collaboration and communication. *CSCW, ACM Press*, 1998.

[31] Follmer S. Olwal A. Ishii H. Leithinger, D. Physical telepresence: shape capture and display for embodied, computer-mediated remote collaboration. *UIST, ACM Press*, 2014.

[32] Burdea G. Popescu G. Boian R. Bouzit, M. The rutgers master ii-new design force-feedback glove. *ASME, IEEE*, 2002.

[33] Jota R. Perumal V. Yao J. X. Singh K. Wigdor D. Araujo, B. Snake charmer: Physically enabling virtual objects. *TEI, ACM Press*, 2016.

[34] Wiertlewski M. Peshkin M. A. Colgate J. E. Meyer, D. J. Dynamics of ultrasonic and electrostatic friction modulation for rendering texture on haptic surfaces. *HAPTICS, ACM Press*, 2014.

[35] Israr A. Poupyrev I. Apa Kim, S. C. Tactile rendering of 3d features on touch surfaces. *UIST, ACM Press*, 2013.

[36] Judith Amores, Xavier Benavides, and Pattie Maes. Showme: A remote collaboration system that supports immersive gestural communication. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1343–1348. ACM, 2015.

[37] Xavier Benavides, Judith Amores, and Pattie Maes. Remot-io: a system for reaching into the environment of a remote collaborator. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 99–100. ACM, 2015.

[38] Xavier Benavides, Judith Amores, and Pattie Maes. Invisibilia: revealing invisible data using augmented reality and internet connected devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 341–344. ACM, 2015.

[39] Roberta Klatzky Laura Valenzeno, Martha Alibali. Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, 2003.

[40] Foldit: Solve puzzles for science. *https://fold.it/*.