icon *: An exploration into learning across distances with a wearable device*

Dhairya Dand

B. Tech in Computer Science, VJTI, 2010

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Masters in Science in Media Arts and Sciences at the
Massachusetts Institute of Technology.
September 2014

Signature redacted

Author ....
Dhairya Dand
Program in Media Arts and Sciences
August 15th 2014

Signature redacted

Certified by ....
Pattie Maes
Alexander W. Dreyfoos (1954) Professor of Media Technology
Program in Media Arts and Sciences

Signature redacted

Accepted by ....
Pattie Maes
Interim Academic Head
Program in Media Arts and Sciences

***icon** : An exploration into learning across distances with a wearable device*

Dhairya Dand

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
on August 15th 2014,
in partial fulfillment of the requirements for the degree of
Masters in Science in Media Arts and Sciences at the
Massachusetts Institute of Technology.

Abstract

There are 7 billion people on this planet. Each one of us is a learner and a teacher. Collectively we hold the entire of humanity's knowledge and boundless compassion. We enrich our lives every single day by unknowingly learning and teaching small things to others.

We have a plethora of tools to communicate at our disposal, tools that make it possible to learn and teach others. However these tools don't operate in our natural hands-on learning environments. They demand their own computational enclosures either on our desks, laps, palms or in front of our eyes. Moreover, they don't support a seamless transfer of intentions that can be achieved by an effective communication tool.

This thesis proposes, documents and reflects on the creation a wearable shoulder-worn wearable communication device 'icon' and a set of tools 'iconHub' and 'iconView' that try to facilitate learning across distances between people, while offering a eyes-free and hands-free experience.

Thesis Supervisor:
Pattie Maes
Alexander W. Dreyfoos (1954) Professor of Media Technology
Program in Media Arts and Sciences

icon : An exploration into learning across distances with a wearable device

Dhairya Dand

The following people served as readers for this thesis:

Signature redacted

Thesis Reader ....

Ethan Zuckerman
Principal Research Scientist, MIT Media Lab
Director, MIT Center for Civic Media

Signature redacted

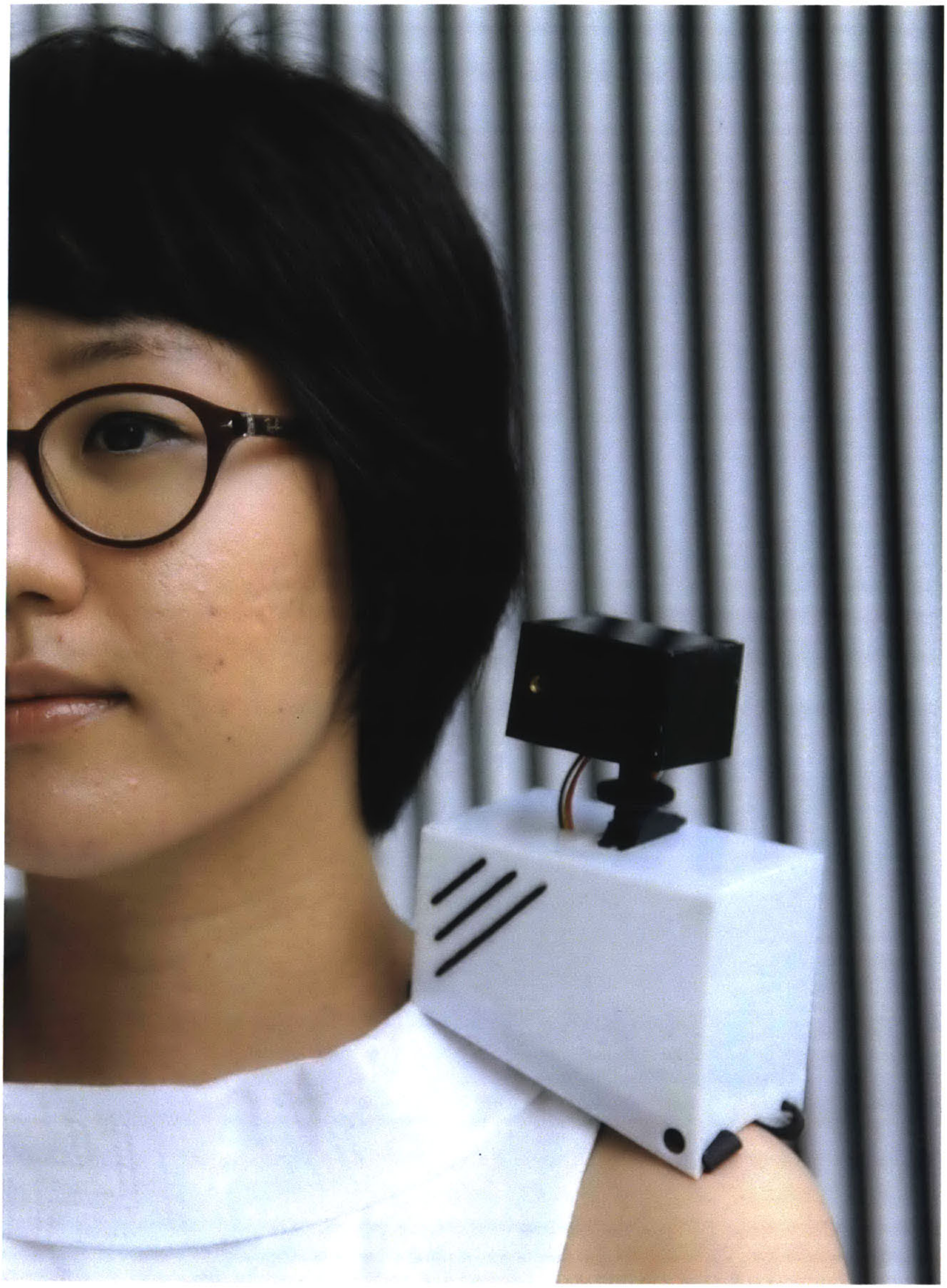
Thesis Reader ....

J. Philipp Schmidt
Research Scientist, MIT Media Lab
Co-founder, Peer2Peer University

Signature redacted

Thesis Reader ....

Jun Rekimoto
Professor, Interfaculty Initiative in Information Studies, University of Tokyo
Director, Interaction Laboratory, Sony Computer Science Laboratories



Acknowledgements

To Pattie Maes and Henry Holtzman. For your vision, your advise, your thoughts, your patience, your kindness and for believing in this boy from nowhere.

To Ethan Zuckerman, Philipp Schmidt, Jun Rekimoto, Hiroshi Ishii, Adrian Cheok and Ravi Poovaiah. For sharing your vision, thoughts and pointing me to the right direction.

To Linda Peterson, Amna Carreiro, Keira Horowitz, William Tilden and Lisa Lieberson. For making my life less of a mess.

To Rob Hemsley, Deepak Jagdish, Linning Yao, Gershon Dublon, Nan Zhao, Nanwei Gong, Jie Qi, Amit Zoran, Jifei Ou, Dan Sawada, Prashant Patil, Jessica Goldfin, Akane Sano, Arlene Ducao, Talia Kaufmann, Matt Hirsch, Kshitij Marwah, Anirudh Sharma, Valentin Heun, Pragun Goyal, Srishti Sethi, Travis Rich, Akito Van Troyer, Xiao Xiao, Hayato Ikoma, Arthur Petron, Tiffany Tseng, Markus Kayser, Rebecca Kleinberger, Sujoy Chowdhury, Andrea Colaco, Santiago Alfaro, Daniel Novy, Akanksha Raina, Mili Tharakan, Yan Yan Cao, Mukund Madhav, Rahul Budhiraja, Nimesha Ranasinghe, Satsheel Shrotriya, Sonal Chinchwadkar, Aditya Sengupta and the good people of 'Funfetti' and 'Katta'. For friendships worth a lifetime.

To the Media Lab. For being the place whose vision transcends life and death.

To Fluids and InfoEcos. For being the most amazing groups of people I've worked with.

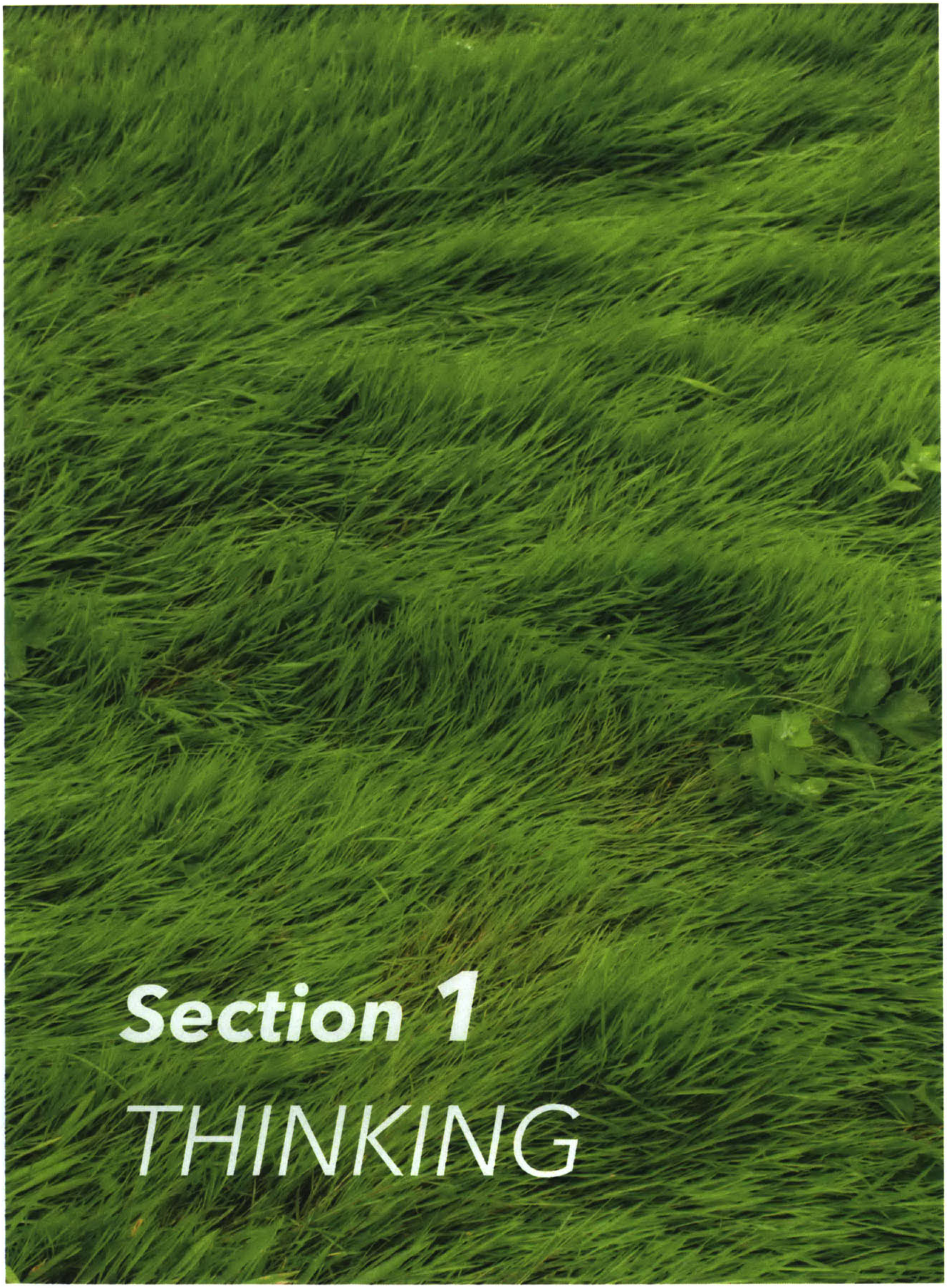
To Sonia Kim. For being there.

To Kasab Shah, Hetal Shah, Janvi Shah, Deepa Ramaswamy, Kartik Shah, Zhanna Zhanabekova and Siddharth Lalka. For being my loving family.

To Mom and Dad. For unintentionally making me a rebel.

To my dearest Uncle. For helping me find me.

CONTENTS



Section 1
THINKING

Chapter 1

The Idea

1.1 The Problem

The Internet has afforded connections across cultures, countries and professions never before possible. We communicate and exchange information, perspectives and experiences through various media - text, images, videos. More specifically we blog our thoughts, capture our moments, share our knowledge, express ourselves through video and connect socially.

The interface and the structure of these platforms obfuscate the fact that most of these transactions are with strangers. We read thoughts of a woman in Pakistan, watch a video made by some youths in Japan or learn to make a pie from a lady in Oxfordshire. These transactions form a larger part of an interesting phenomena - all of us are learning, though not to be masters at a certain craft, skill or technique but enriching our lives byte by byte through these open-minded transactions; unknowingly being a part of a global-scale system.

Unlike traditional teacher-student relationships, we are now learning from multiple teachers in these short transactions while we decide the pace of our learning. In other words, we are strolling through the digital landscape while subtly getting deflected towards new information driven by our natural curiosities, similar to Guy Debord's theory [Debord58] of the dérive which proposes unplanned journeys through a landscape, usually urban, on which the subtle aesthetic contours of the surrounding architecture and geography subconsciously direct the travelers, with the ultimate goal of encountering an entirely new and authentic experience.

Our medium of choice to participate in these exchanges is a computer, tablet or a smartphone - all of which are a slab of glass. Even while wearable computing in the form of Google Glass and other products such as Oculus Rift are coming of age - they propose to insert yet another piece of glass between our vision and our reality. There is a disconnect in the field of learning - the screen and the field of application - the real world. The lack of context awareness - where we consume pre-packaged information and lessons which won't necessarily apply to our reality lends itself to this disconnect. Say I learn to cook Ramen from a YouTube video or a recipe website, however I'm not guaranteed to have access to all the exact tools and ingredients mentioned in the tutorial, whom do I go to for help?

I am motivated by the design enquiry of how such a stranger learning network may be created, where people are connected across distances, where one learns from multiple teachers, charting their own path, where the landscape of this transaction is shifted from behind a slab of glass to the real world, at the same time building empathy and letting others peek into our reality, where we learn by being there.

Some of the design considerations are to create a asymmetric device - where one person would wear the device and be mobile

- broadcasting their environment, while the other person would be able to interact and experience the wearer's reality. Having such an asymmetric experience would avoid confusion in roles. Also, none of the users see each other's faces; they only see their environment - as if they are looking at their world through the other's eyes. This choice was made to avoid subjective judgments of the teacher based on the learner's appearance; especially when the teacher is a stranger, the focus should be the journey. Since teachers would get to see a snapshot of the learners environment and choose where they wish to travel - this incentivizes the learner to go out into the open, discover new places, or simply take a break to be attractive to a potential teacher.

1.2 Research Goals

Medium

An effective medium is one which can transfer intentions without loss in meaning. For the medium our goal is to develop a system of communication which is technically effective, thoughtfully designed and delightful to use.

Learning

We wish to create a learning system that supports soft-learning, learning-on-the-go and cross-cultural exchange. In contrast to a traditional structured learning system, we take a PIY - plan it yourself approach where users can chart their own learning path.

Experience

Learning can happen anywhere and anytime. In this spirit, our system should be designed to work in the real world, staying out of the way of the learners yet equipping them with essential tools.

1.3 icon System

The *icon* system comprises of the *icon* device, the *iconView* app and the *iconHub* platform. The device is worn on the shoulders of the host, it streams a point-of-view video, enables bi-directional audio communication, a guest can tap on the host's shoulder and lights in the device act as an interface to indicate new guests and that the device is live. The iconView app allows the guest to see the video, talk, tap, along with capturing and recording moments in a session with the host. The platform acts as a hub where guests can discover hosts, become hosts and hosts can schedule sessions.

Let's take an example, you're cooking, or you're operating a laser cutter to make that craft object you always wanted. You're doing it yourself. You read the instructions in that blog post multiple times, you saw that video on YouTube but what you made looks different, you're stuck and don't know whom to ask. Maybe you followed all the steps right but don't why it isn't working. You wear the device and someone else, not an expert but someone who just did what you're trying to do - jumps in to your session. They walk you through the steps, they point out and ask your a question about that step, they talk to you. You frustration turns into excitement, you connected with a real person, both of you're better for this interaction. There is often not one right way to do something, but by following someone else's instructions, you don't create something new. Now you do, you take your steps, your guest learns your mistakes or your ingenuity.

1.4 icon Approach

Wearable

The device is wearable allowing the user to be hands-free and move around in their environment. This way the device doesn't dictate the setting for learning but rather assimilates in the learning environment.

Context and Focus

The *icon* system has two stakeholders - hosts and guests, hosts wear the device and guests experience and interact with the host via a computer, tablet, smartphone. Owing to the unique and asymmetric form factors for the host and guest, the guest can set a context for the session while since the host is mobile - they can set the focus for the session.

In Situ, Not in View, In the Real World

Using the *icon* device learning can happen when the host wants, where the host wants. Since the device is worn on the shoulder, no part of the device is in the direct view of the host, leading to a hands-free and eyes-free experience. The augmentations of objects in the host's environment is achieved using a laser pointer.

Platform

The *icon* approach considers everyone a learner and everyone a teacher, that by connecting learners to teachers and vice versa we can facilitate learning by capitalizing on the knowledge in the system. The *icon* system achieves this through the *iconHub* platform.

De-subjectification

Neither the host nor the guest see each others faces, the host never sees a video feed from the guest and the guest only sees a first person point of view. We hope this puts the focus of the interaction on the interaction, while not being biased with the physical appearance and its connotations.

1.5 Thesis Outline

This thesis is divided into four sections : Thinking, Doing, Dreaming and Addendum.

The *Thinking* section reviews the problem area, the research goals of this thesis, the approach taken and its merits. It also reviews related works in the field.

The *Doing* section documents and outlines the system, its implementation and the interactions and applications it facilitates.

The *Dreaming* section concludes this thesis with its defining contributions and future work.

Chapter 2

Related Work

2.1 On Collaboration

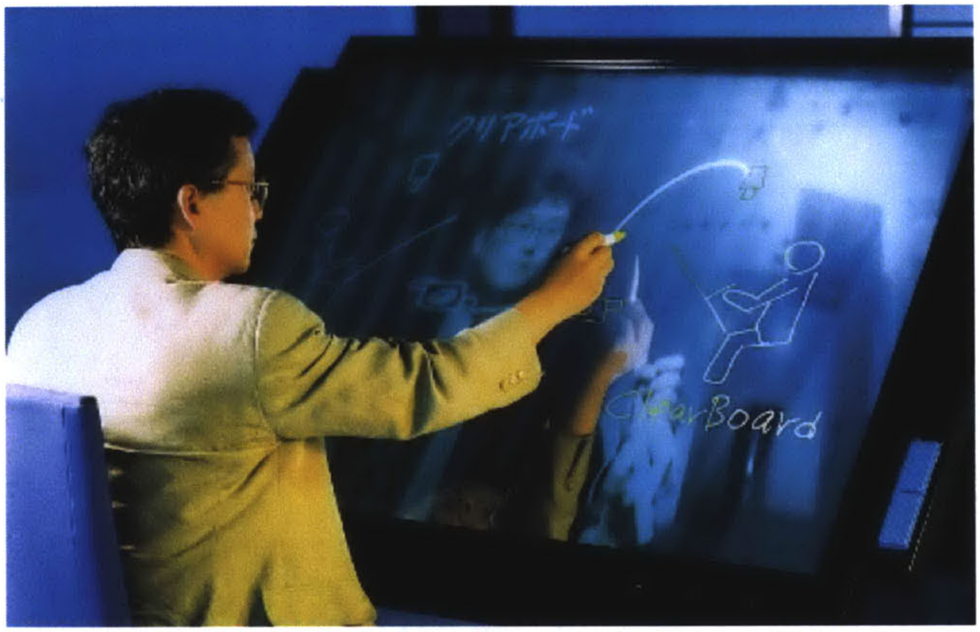


Figure 1: Ishii's TeamWorkStation system

Human collaboration has been a subject of philosophy and social science throughout history. Political economy, sociology, and anthropology are among the many disciplines concerned with how people live and work together. Theories of human collaboration vary widely. Some emphasize cooperation. Others observe the inevitability of conflict. Collaboration, in this context, does not mean happy agreement on ends. All approaches acknowledge, at least implicitly, that the primary activity in the process of collaboration is communication [Johnson86]. As also observed by Hollan in his seminal paper *Beyond being there* [Hollan92] - "In contrast, we argue that a better way to solve the telecommunication is to not focus on the tele- part, but the communication part. That is, to make the new medium satisfy the needs of communication so well that people, whether physically proximate or not, prefer to use it". Further Holan suggests that to create such tools, we should frame the problem in terms of needs, media, and mechanisms. The goal then becomes identifying needs which are not ideally met in the medium of physical proximity, and evolving mechanisms which leverage the strengths of the new medium to meet those needs. Moving from Johnson's and Hollan's metaphysical observations to Ishii's TeamWorkStation [Ishii90] which allows users to keep using their favorite individual tools (in whatever form) while collaborating in a desktop virtual shared workspace, we observe a focus on communication at the fusion of the physical and digital. Recent works such as LuminAR [Maes10] demonstrate collaborations through a robotic augmented reality platform.

2.2 On Actions and Pointing

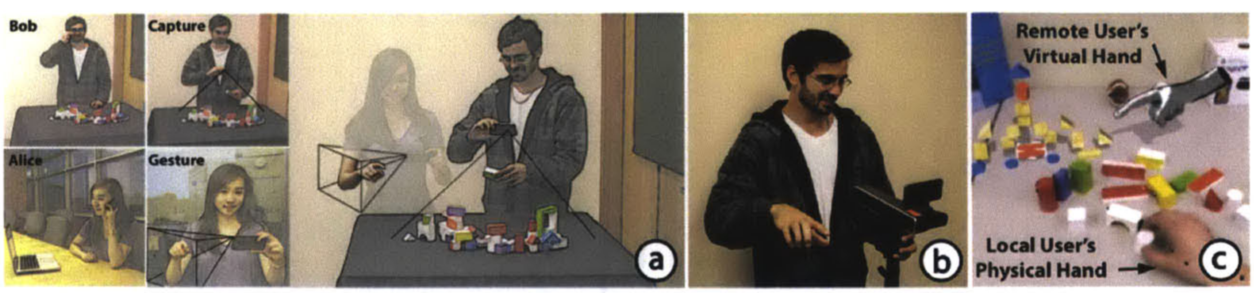


Figure 2: Sodhi's BeThere system

Bauer in his study of remote collaboration [Bauer99] shows that a reality-augmenting tele-pointer is an effective means for enhancing remote collaboration of wearable users. He notes "We see it as an indication of the utility of the reality-augmenting tele-pointer that participants in our study rarely, if ever used phrases like: 'Which hole do you mean?' or 'Where are you pointing at?' Such questions would have indicated serious problems in our design". Gregle concurs in his paper 'Action as Language' [Gregle04] that actions provided a more efficient mechanism for establishing mutual understanding. Recent results from systems such as 'BeThere' [Sodhi13] which use an augmented tele-pointer note "We focused on pointing, because making deictic references meaningful seems to have very high potential impact. Pointing is relatively easy to transduce".

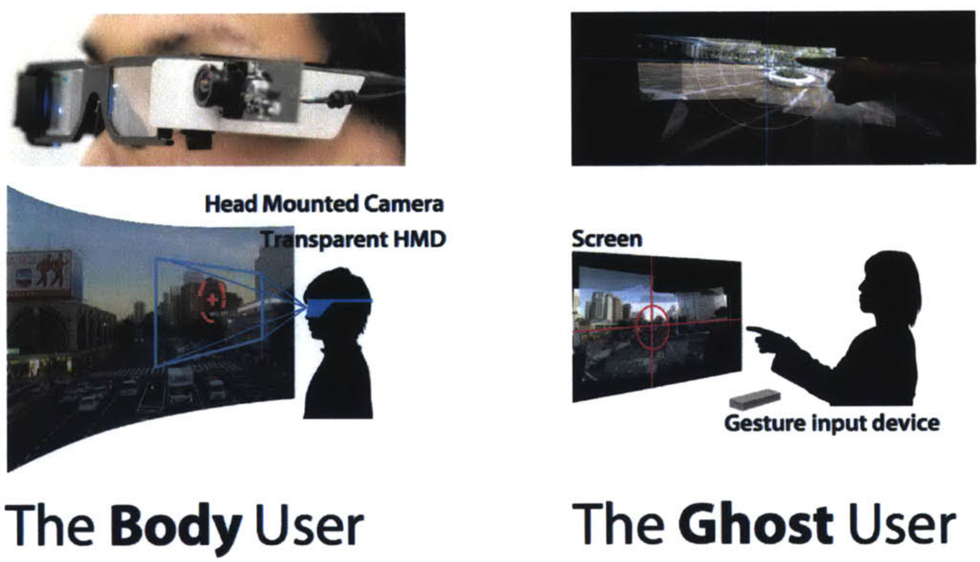


Figure 3: Rekimoto's JackIn system

Rekimoto's JackIn [Rekimoto'14] system uses a HMD (head-mounted display) with a remote user pointing on a screen, the output is shown on the HMD. Steve Mann's Telepointer [Mann00] is a neck worn device used for pointing with a projector.

2.3 On Sharing Vision, Audio or Control

Figure 4: Goldberg's TeleActor system

Several systems have been studied which allow the sharing of vision or audio, one of them profiled in 'Seeing eye to eye' [Tang02] comprises of an HMD which allows wearable computer users to share their views of their current environments with each other. A wearer, by looking around his/her environment, "paints" or "builds" an environment map composed of images from the EyeTap device, along with head tracking information recording the orientation of each image. Other systems such as the TeleActor [Goldberg02] are modeled as collaborative online teleoperation system for distance learning that allows many users to simultaneously share control of a single mobile resource. This is achieved through a "Spatial Dynamic Voting" (SDV) interface that collects, displays, and analyzes a sequence of spatial votes from multiple online operators from their Internet browsers. The votes drive the motion of a single mobile robot or human "Tele-Actor". Much earlier experiments namely HyperVoice [Resnick92] show that it is possible and worthwhile to use the telephone as a platform for input and retrieval of semi-structured information objects. Another notable work [Drugge04] merits the importance of an efficient, context-aware, rich interface for the remote user towards the success of a wearable collaborative system.

2.4 On on-Body Collaborative Devices

Figure 5: Kuzuoka's GestureCam system

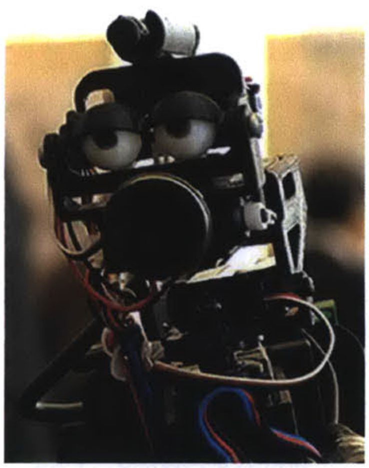


Figure 6: Kashiwabara's Teroos system

The earliest works in this field, GestureCam [Kuzuoka94] where a camera and a pointer is fixed in the user's location while a physical wooden contraption is used on the remote user's side for tele-pointing. In Kuzuoka's user tests involving users speaking different languages, he observed '..the major causes of the long instruction session was the language barrier. She said, however, that the laser pointer and other gesturing functions of the GestureCam did help communication'. Other systems such as WACL [Kurata02] which is a shoulder worn camera and laser system with a HMD for display showed that the WACL is more comfortable to wear, is more eye-friendly, and causes less fatigue to the wearer, although there is no significant difference in task completion time. icon differs from WACL in that it stems from a design approach to learning rather than a technocentric approach. More recent systems such as Teroos [Kashiwabara12] which is a personified shoulder mounted robot and WaistCam

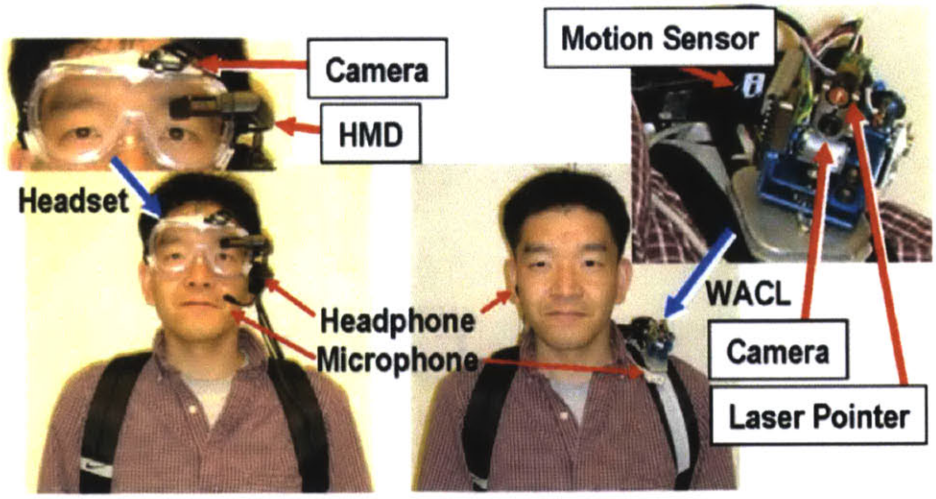


Figure 7: Kurata's WACL system

[Morishima12] which is a waist mounted camera and tele-pointing system show similar results with users preferring the system over a stationary camera.

2.5 On Digital Apprenticeship

In Berlin's paper 'Consultants and Apprentices' [Berlin92] which studied the role of a consultants and apprentices in the context of remote computer supported collaborative work although limited to a sample user space of programmers, showed that 1. enabling apprentices to encapsulate information as they prepare for a meeting, 2. enabling apprentices to switch tasks, 3. facilitating negotiation about the consultant's availability, 4. facilitating the interaction itself, and 5. extending the apprentice's community of consultants via electronic bulletin boards and other information sharing tools, are important design goals if such a system has to succeed.

2.6 On Commercial Products

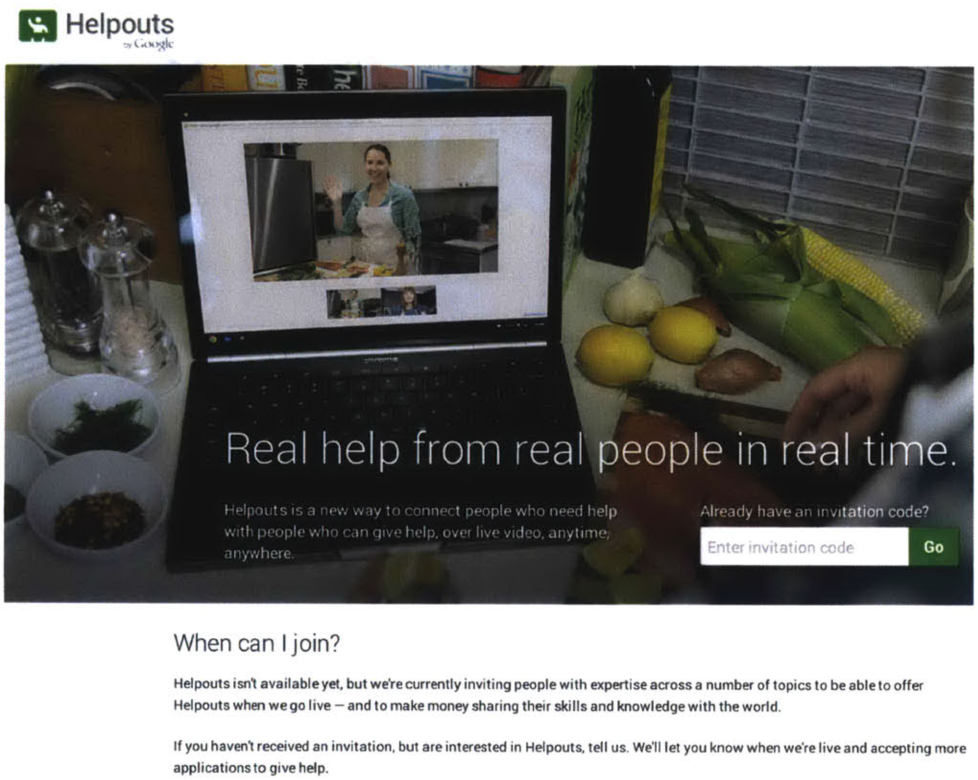


Figure 8: Google's Helpout

Skype, FaceTime, Polycom, Cisco and various other providers facilitate face-to-face video teleconference. Recent products such as DropCam allow for remote monitoring through video and audio channels. Google's Helpouts allow users to seek help from experts through video chats. All these products operate in the realm on video and audio without any physical manifestation of intents.

2.6 On Learning Tools

Tools such as Instructables, DIY.org, wikiHow have democratized the process of learning through exhaustive documentation that is regularly updated and annotated by the community. Other platforms such as Khan Academy, EdX, MIT OCW have allowed the rapid spread of classroom content in an online remote environment. Recent platforms such as UnHangout are reshaping the interactions between peers in an online-first learning environment.

2.7 On Situated Learning

Situated Learning, proposed by Jean Lave and Etienne Wenger [Lave91] is a model of learning in a community of practice. At its simplest, situated learning is learning that takes place in the same context in which it is applied. Situated Learning particularly resonates with adult learners, Kimble and Hildreth [Kimble08] show how adult learners discover, shape and make explicit their own knowledge through situated learning within a community of practice. Halverson [Halverson09] articulately states that the model of learning a skill through technology mimics how individuals learned in the past from a professional in that skill.


Section 2
DOING

Chapter 3

System Design and User Experience

3.1 icon System

Learning happens everywhere. The *icon* system is designed to be wearable in its form and multi-platform in its function. The system has three components operating in tandem to afford a seamless experience between the hosts and the guests. An overall sketch of the system can be seen in the following figure.

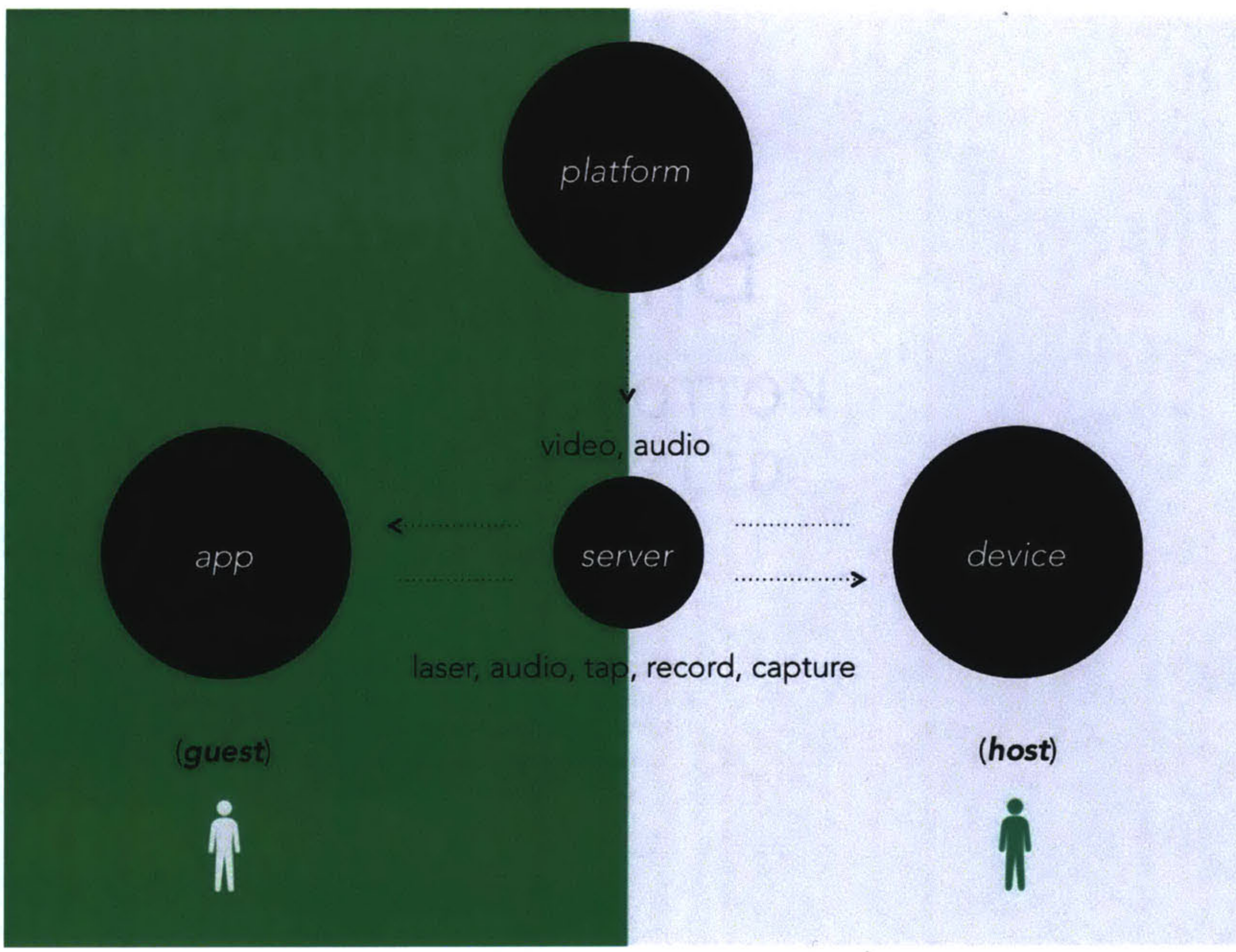


Figure 9: icon System

The system comprises of the device, app and the platform.

The *device* is worn by the host, allowing the guest to view the host's first person point of view, immersing them in the experience. The device makes both the video and audio accessible to the guest. The device also has a laser pointing system that the guest can control with the *iconView* App. The laser pointing system allows the guest to point, draw and call attention to certain parts in the host's environment. The device also consists of the paired smartphone that allows for video and audio streaming.

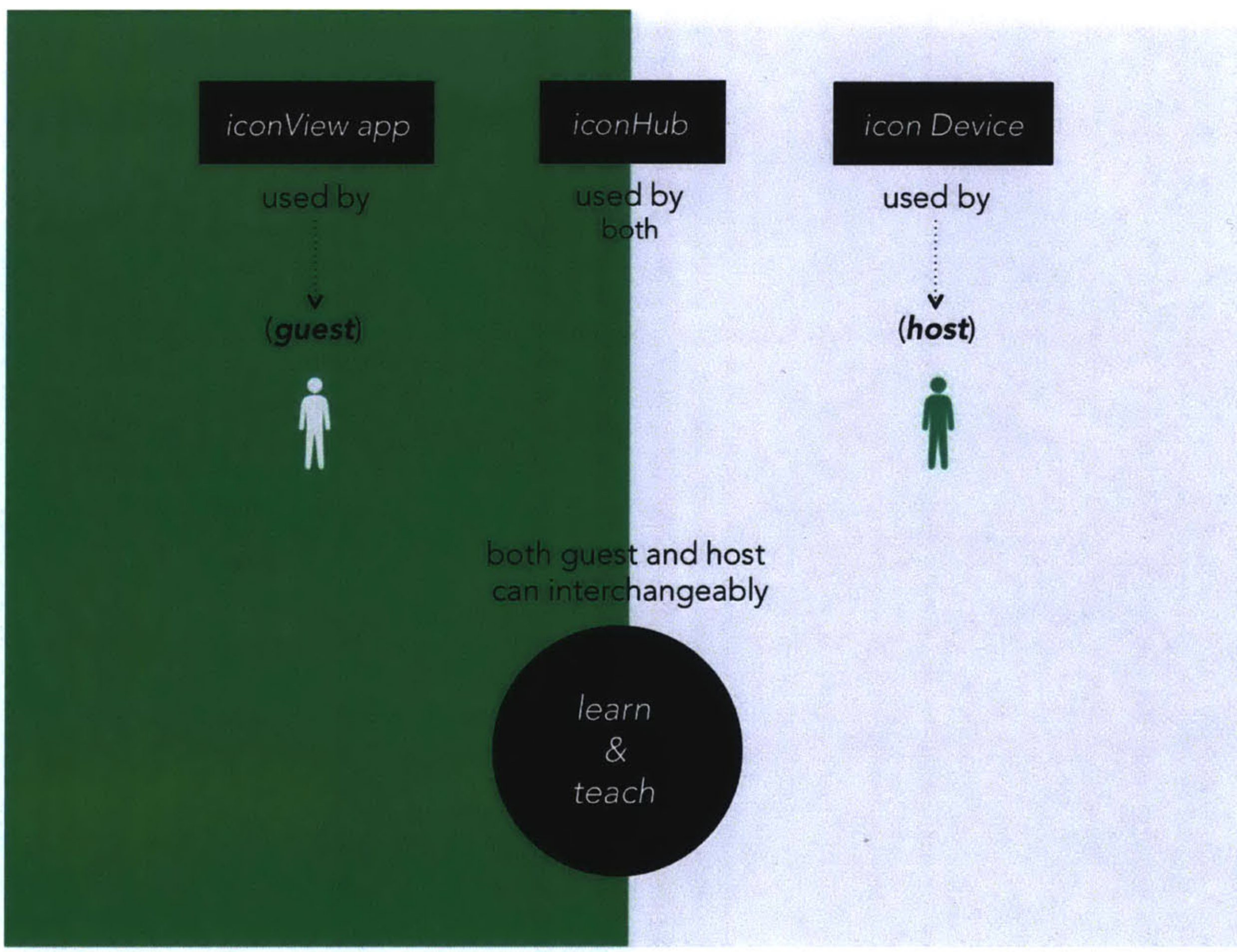


Figure 10: icon Terminology

The *platform* is built for the web, making it accessible on a desktop, tablet as well as a smartphone. The *iconHub* platform is where the guests can find hosts. Once they choose an area of interest they can see the live hosts on a map and what they are offering to teach / experience / show / discuss et cetera. The guests can also themselves become hosts. Hosts can schedule sessions in the future allowing guests to sign up. Both the guests and hosts can edit their profiles.

The *app* is also built on modern web technologies to allow the guest to experience the host's environment. The iconView app can be used on most smartphone platforms and desktop operating system platforms in browsers that support WebRTC. The app allows the guest to see and hear the host's environment along with the ability to talk, point, record the session and get more information on all the participants in the current session.

3.2 icon Device

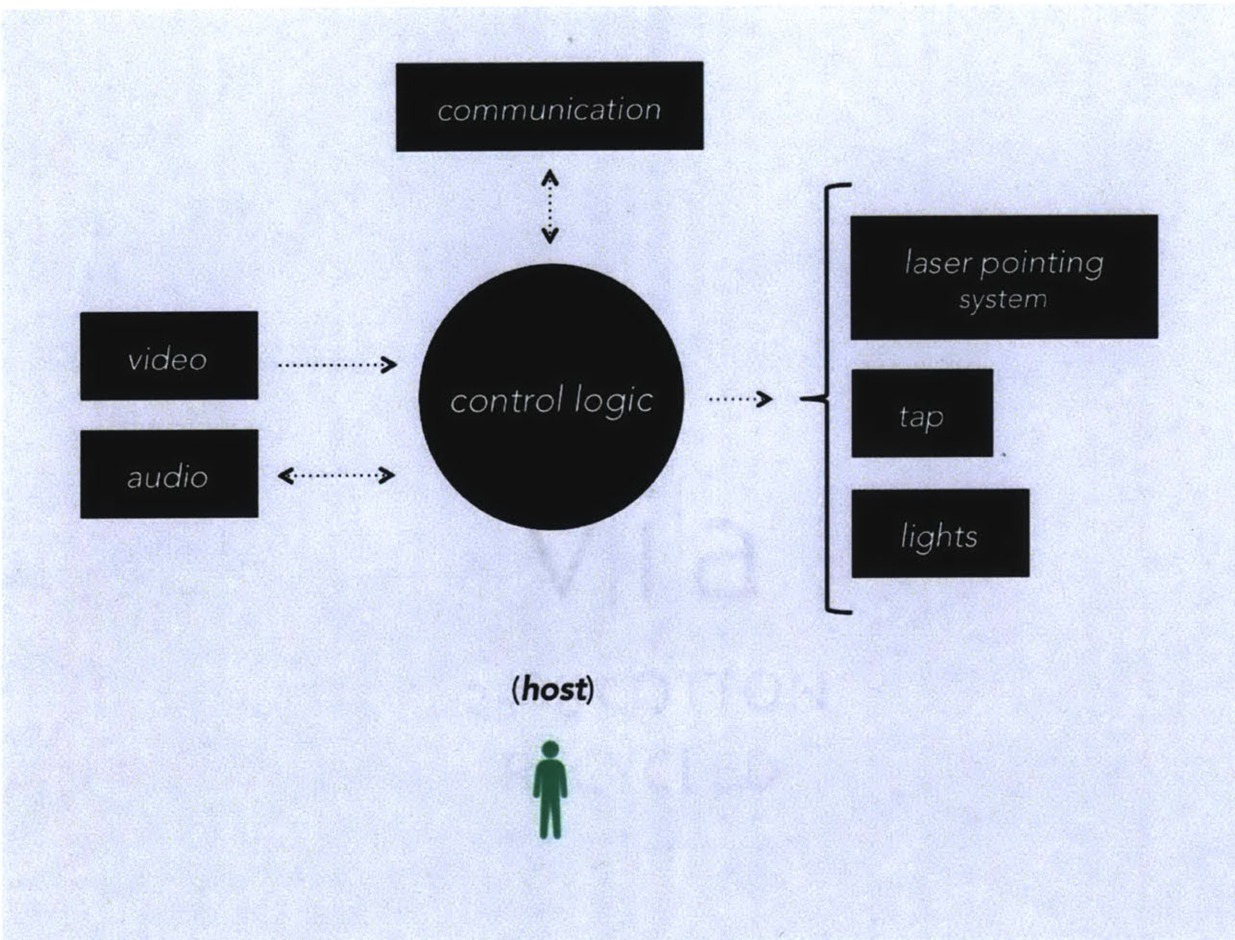


Figure 11: icon Device system diagram

The *icon* device is worn by the host on their shoulder. The device serves as the 'icon' of the remote guest who interacts with the host through the device. The device streams video and mediates two way audio between the host and guest. In addition, the guest can point in the host's environment using the laser pointing system and tap on their shoulders. The connected smartphone is referred to as part of the device in this thesis.

3.2.1 Device Design

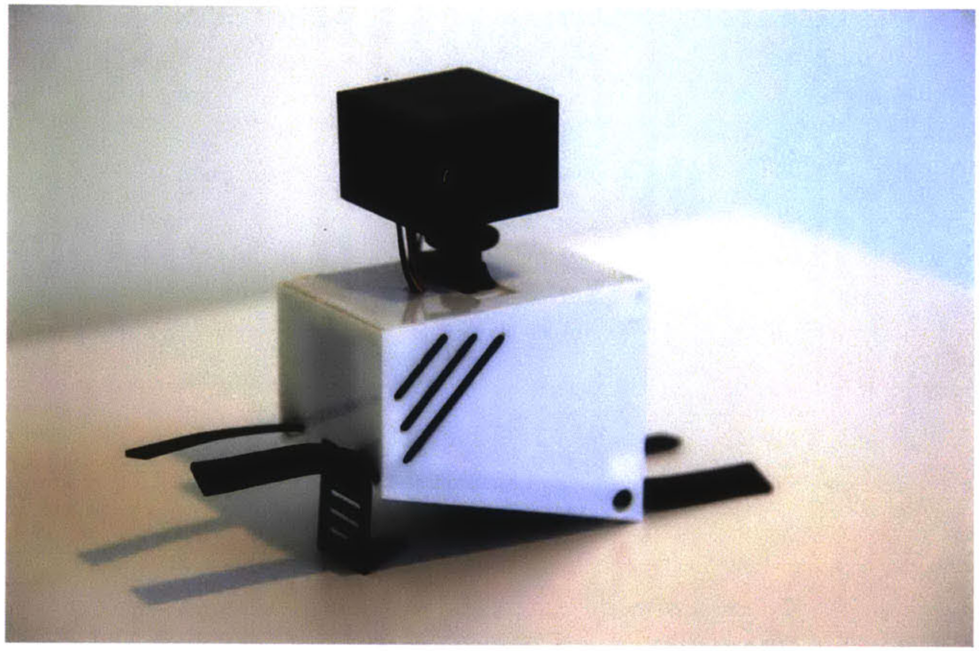


Figure 12: icon Device (mounted on a removable stand)

The device is designed both in its architecture and form to be simple and interfaceless to the host. The shoulder serves as a good mounting place since it offers almost the same point of view as a camera mounted near the eye. The device has no buttons and most of the actions are accomplished using a connected smartphone. The base of the device is padded with anti-slip rubber and snaps in place with magnets. The device pairs with a smartphone to help stream video and audio.

3.2.2 Video, Audio

The associated smartphone captures the video and audio from the host's surroundings and transfers them to the guest. The guest enjoys an approximate point of view while the host does not see the video feed from the guest. The audio channel is two way allowing both the host and guest to talk.

3.2.3 Laser

The laser is mounted on the 2-axis pointing system on the device. The laser is modulated to have a fade-in fade-out effect which makes it visually more pleasing. The laser serves as a metaphorical extension of the guest's hands, which can be used to call out objects, places in the host's environment.

3.2.4 Lights

The embedded white LEDs in the device serve as an indication for a live session and when a guest joins a session. The lights are placed inside the surface of the device and diffuse through giving an even illumination to the device. When a session is live the lights fade-in fade-out giving a breathing effect and giving other people in the host's environment a visual cue that the device is on. When a new guest joins in a session the lights momentarily fade-in fade-out faster giving a cue to the host.

3.2.5 Tap

The guest can tap on the host's shoulder to catch their attention in an otherwise loud environment. The taps serve as a co-located physical extension of the guest's remote presence.

3.3 iconView App

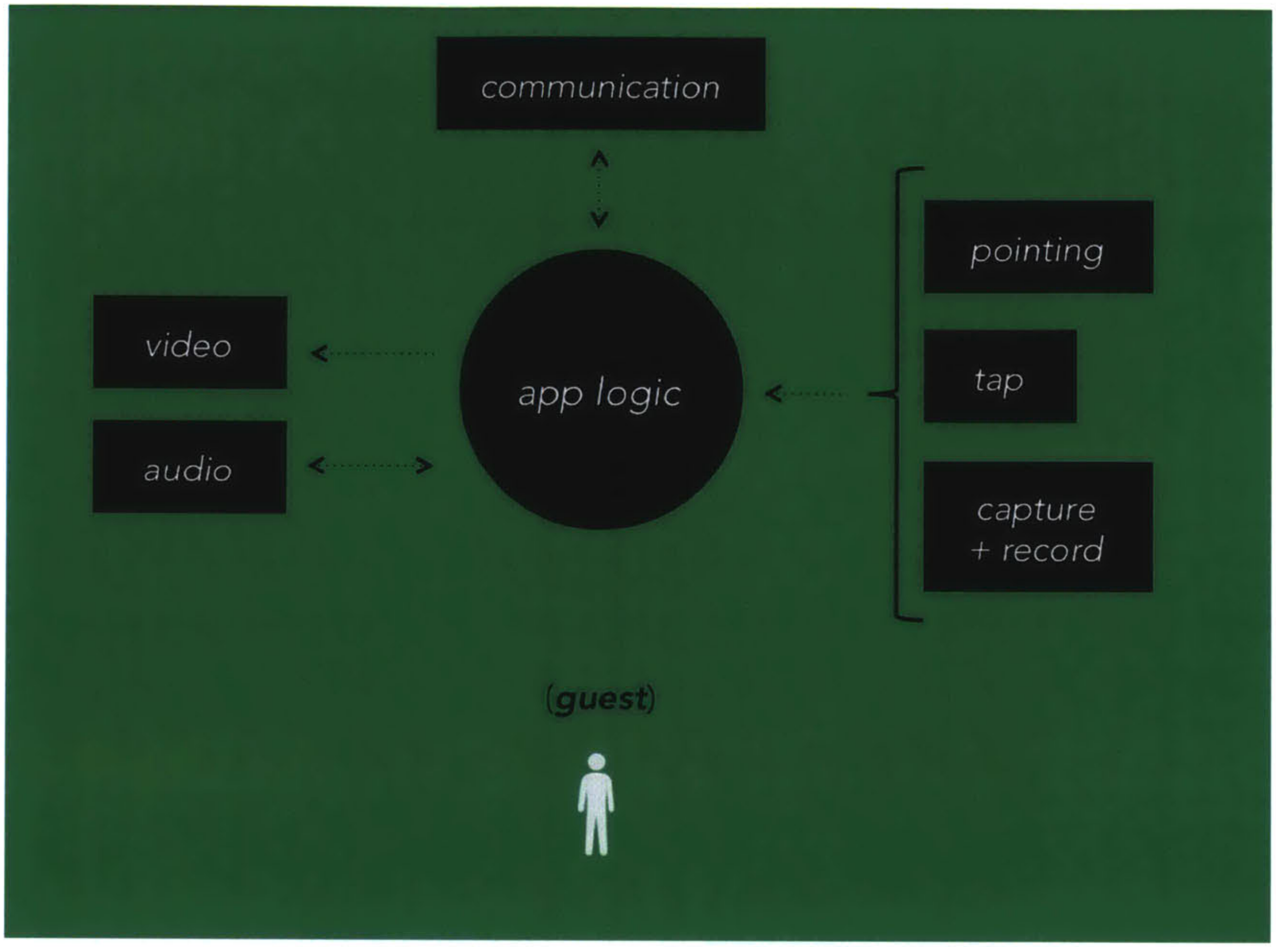


Figure 13: iconView App system diagram

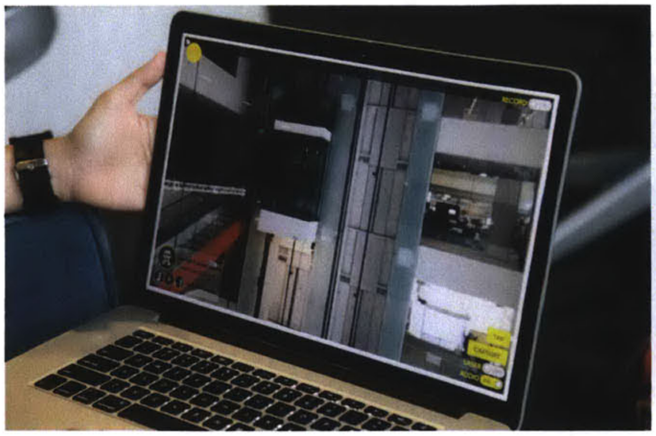
The *iconView* app allows the host to interact with the host's environment. It does this in tandem with the *icon* device worn by the host. The app streams real-time video and audio from the *icon* device and allows to control the laser as well as to talk to the host and tap their shoulder. The guest can capture parts of the session, record the entire session for later viewing as well as see the profiles of other participants.

3.3.1 App Design

Figure 14: iconView App design

The app is designed to show the video full screen at the highest resolution possible for the current network connection. The video and audio streams real-time, the audio can be heard from the device's speakers. All the controls and status information are displayed at the corners of the app screen within the video embed. The focus of the design is to be immersed in the interaction as much as possible. With the accessory controls being within instant reach all the time while still not compromising on the the experience.

3.3.2 Multi-platform



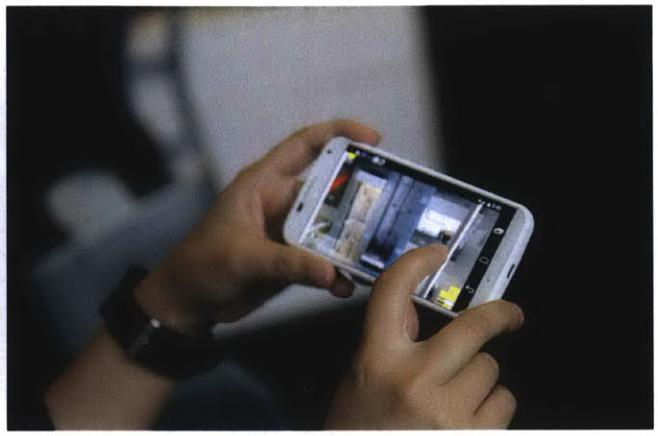


Figure 15: iconView App on a desktop and mobile browser

The *iconView* app has been built using modern web technologies. WebRTC being at the core of these web services. Currently the app has been tested to run on the following platforms-

Desktop (Mac, Debian, Widnows)	iOS	Android
Chrome 33+, Firefox 24+. Opera 20+	Chrome 34+, Bowser 1.2+, Opera 20+	Chrome 34+, Opera 20+

3.3.3 Video and Audio

The app streams real-time video and audio from the *icon* device. The host cannot scrub through the time scale of the video during an ongoing session. The audio is streamed in stereo live through the speakers. The guest can choose to turn off either the video or the audio streaming in a session to improve the experience if they are on a slow network. The guest can talk back to the host using Figure : Real-time video and audio communication

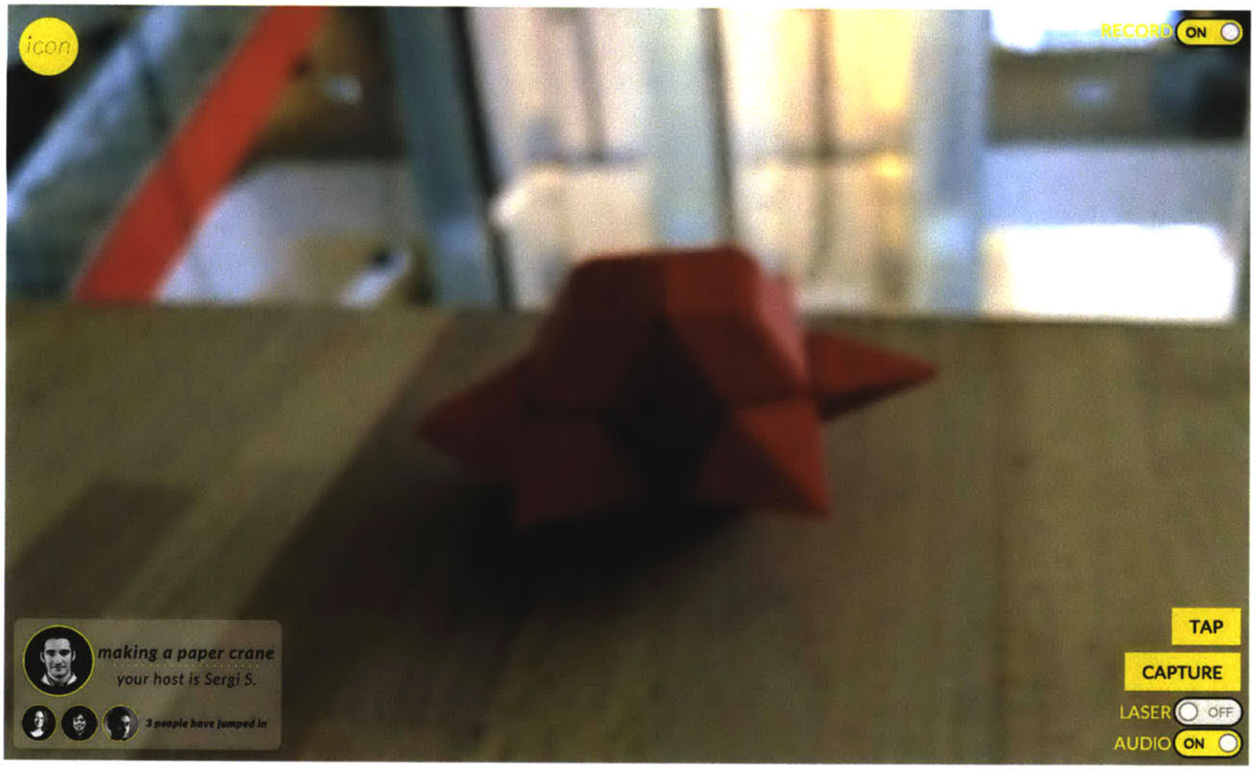


Figure 16: Real-time video and audio communication

the audio channel, making it a 2-way communication channel while the video is only 1-way.

3.3.4 Pointing

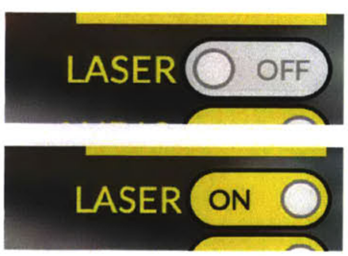


Figure 17: Laser toggle switch

The guest can use the *iconView* app to point in the environment of the host using the laser pointing system on the device. There is a laser ON/OFF toggle switch on the app to give greater control to the guest. The guest can point, draw or write by simply touching / clicking on a specific part of the real-time video on the screen. Thus by directly interacting with a finger onto the screen which is streaming say an object in the host's environment, the guest can point the laser to that object. If the guest continues to hold the touch / click on the video stream, the *icon* device would continue to point the laser in the corresponding path thus mirroring the actions from the screen to the real world. The guest can see the laser's output in the video feed from the host.

3.3.5 Capture

Figure 18: Capture button

The guest can capture instances in the current interaction that they can visit later in the future. The capture feature allows to catalog meaningful moments as the current session progresses.

3.3.6 Tap

Figure 19: Tap button when pressed

There are situations when pointing a laser is not socially acceptable in the host's environment as well as the host might be in an environment with loud ambient noise which would make both pointing and audio based communication impossible. In such situations to capture the attention of the host, the guest can use the tap button to physically relay a tap to the host's shoulder.

3.3.7 Record

Figure 20: Record toggle switch

The guest might wish to record and capture certain parts of the current interaction for later perusal. They can do so by toggling the record switch which stores the current session with the audio, video and pointing information. They can access this recorded session using the *iconHub* platform.

3.3.8 Profile and Participants

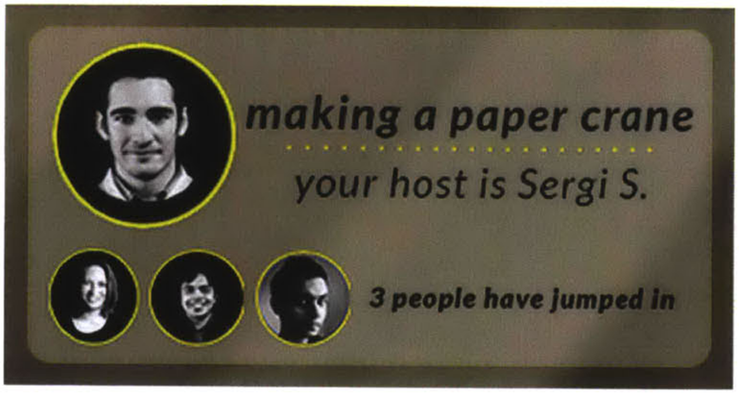


Figure 21: Session information and other participants

While in a session, the guest can see more information of the topic for the current session as well as more information about the host. Specifically to protect privacy in the system, the guest can only see the host's profile picture, screen name, city and interests. In sessions where multiple participants (guests) join in, each of them can see the profile information of the other live guests.

3.4 iconHub Platform

The *iconHub* platform acts as a hub for interaction between the community. Guest can choose what they wish to do, example - learn / teach / help / travel et cetera. They can find hosts in their interest area around the world and join in. They could also search for hosts or post a request for a particular session. Guests can also use the platform to schedule sessions in the future with hosts.

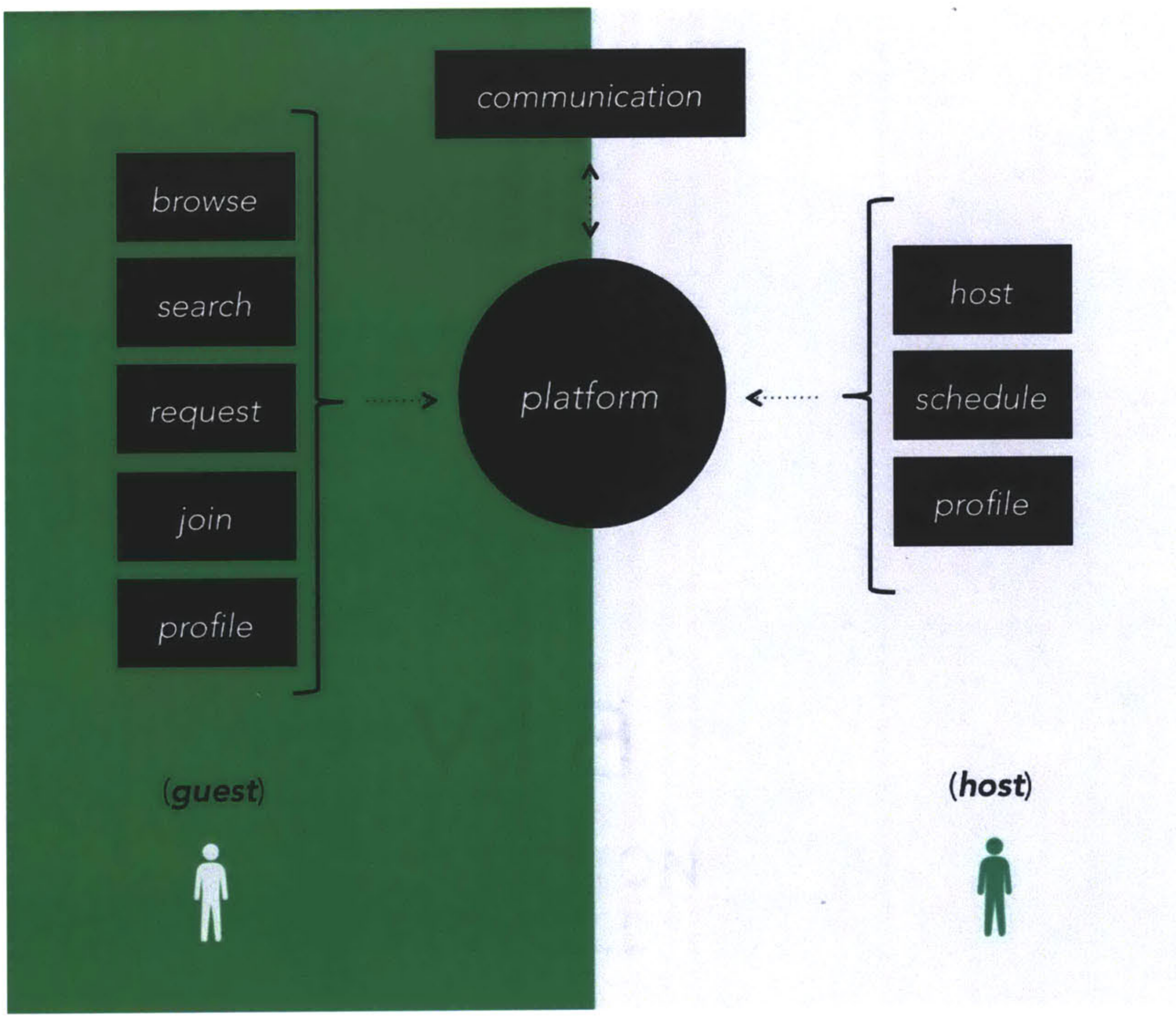


Figure 22: iconHub system diagram

3.4.1 Activities

The *iconHub* platform connects guests to hosts who are interested in performing various activities. The platform accommodates the following activities:

Figure 23: iconHub : choose an activity

Learn	*Learn skills, sports, cooking, crafts or just about anything that a wearer wants to teach.*
Teach	*You can teach something you know to a wearer and explore their surroundings, culture and ways.*
Travel	*See the world through the eyes of a local, live a day in the life of ...*
Witness	*A first person view of the events happening around the world.*
Speak	*Learn or practice a new language with a native speaker in their own country*
Help	*Help someone fix something.*
Play	*Coach someone playing a sport, see the action through their eyes.*
Observe	*Simply observe someone, be mindful, appreciate.*

Hire *You can hire wearers local or global to do a specific task for you. Shopping? Flashmob?*

3.4.2 Profile

Once signed up for the *iconHub* platform, the user has to create a profile with their screen name, profile picture, location and interests. These profiles are public and serve as point of interaction and identification of common interests between guests and host.

3.4.3 Hub

Figure 24: iconHub

Once the guest is logged in and chooses an activity, they can see a world map with markers representing hosts which are live around the world with the *icon* device. On clicking the marker *icon*, the guest sees a popup. This popup displays the title of the session which the host wants to engage in (either teach, or a place they offer to take the guest around: depending on the activity chosen). The popup also shows a live image capture from the *icon* device. This gives the guest a preview, a feel of the host's environment, to set expectations right before the guest chooses to jump into a session with the host.

3.4.4 Search, Request and Schedule

Figure 25: Search for wearers

Apart from visually browsing the available live hosts, the guest can also perform a keyword search for interesting sessions that are being offered live.

Figure 26: Leave a request

If the guest does not find a relevant live session, they can put a request on the platform for a specific session they are looking for. These serve as suggestions for hosts when they want to start a new session, the guest is also notified as soon as their requested session is available.

Some hosts may choose to specify their free time slots and the guests could then sign up for these sessions. This scheduling feature allows the host and guests to better co-ordinate the sessions for timing and critical mass.

3.4.5 Host a Session

The *iconHub* is also the place where people can choose to host a session once they have an *icon* device. To host a session, the user simply has to supply a title for their intended session, they could then be live with the device or schedule a session at a particular time. The device automatically extracts the location as well as an image capture of the host's environment when they are live. These go in enriching the live map view where guests can find them.

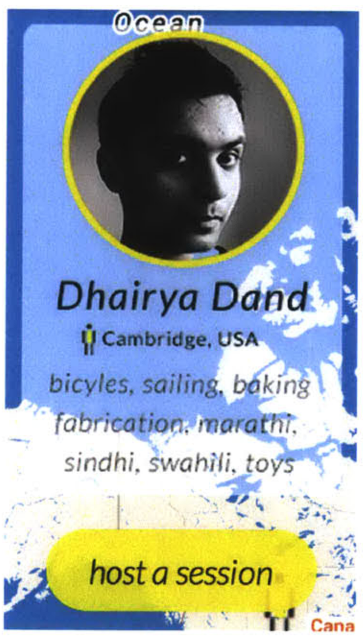


Figure 27: Host a session

3.5 Interactions and Experience

The goals of the *icon* system's experience is a subtle seamless interactions and focus on the exchange at hand. The system should work in tandem to allow the users to experience and immerse in a far away reality while still giving them features and functionality that would enrich their exchange.

The following diagram encapsulates the various actors in the system and the interaction flow between them.

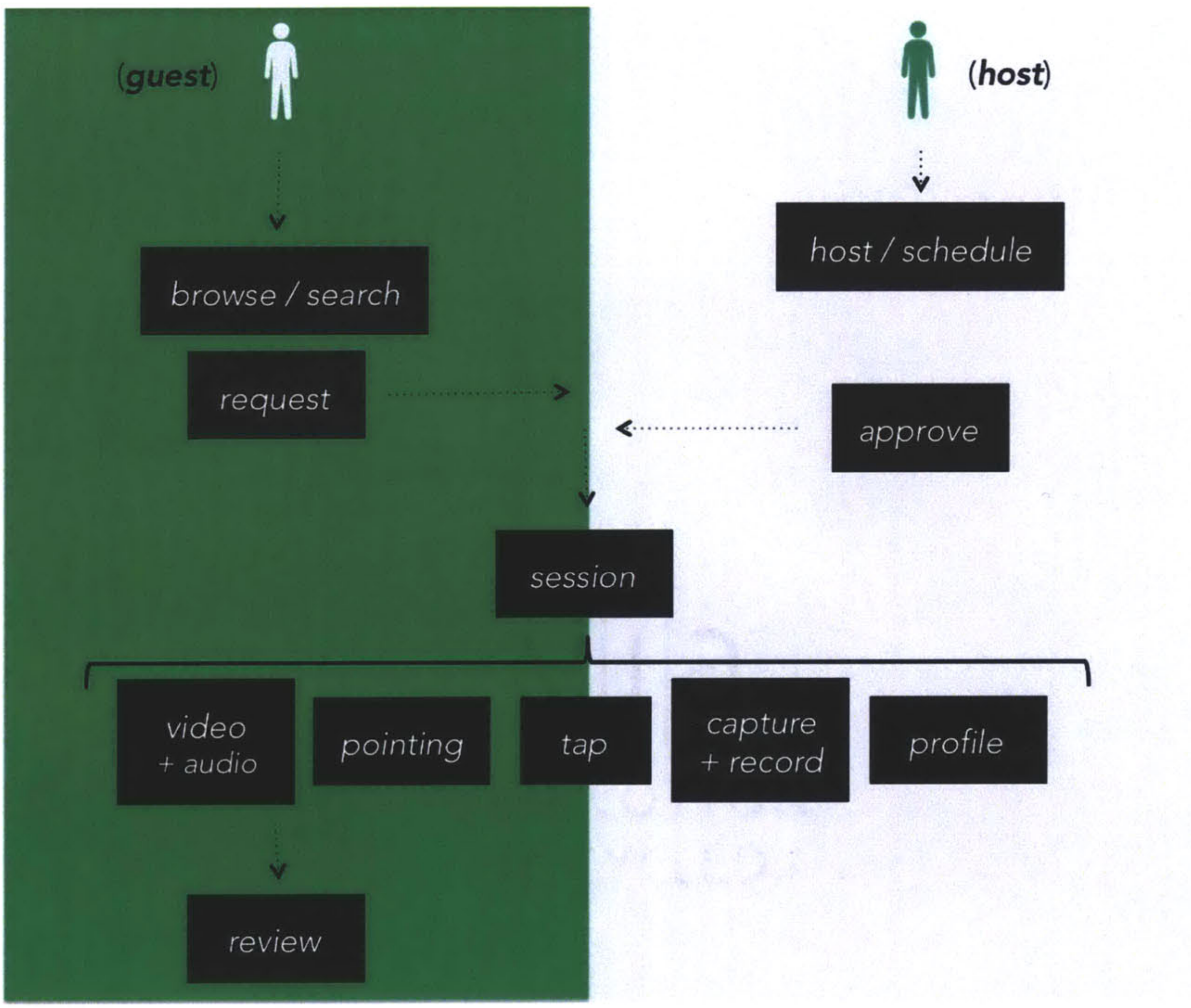


Figure 28: icon System activity flow diagram

Step 1. Available to Host

The host inputs their session title using the *iconHub* platform on their smartphone. The app also records their location (city).

Step 2. Wearing the Device

The host wears the device on their shoulder. The device sends image snapshots from the host's environment to the platform.

Step 3. Guest discovers Host

The guest logs in to the *iconHub* platform on their tablet / desktop / smartphone, chooses an activity and is presented with live hosts which match their interests. After visually exploring or using the textual search bar the guest finds an interesting session they would like to join in.

Step 4. Request and Approval

The guest then sends a request to the host to join in, the host hears the request from their device. They can approve their request after seeing their profile, or if the session is open - the guest automatically joins the session - in which case the white diffused light on the device blinks signaling that a new guest has joined the session.

Step 5. Session

The hosts and guests can do any of the following actions once in a session:

Video, Audio

Once connected the guest receive a real-time video stream from the device providing them with a first person point of view of the host. They can also audio chat with the host through a two-way audio channel.

Pointing

The guest can use the laser pointing system on the device to point the laser in the host's environment. They do this by touching / clicking the screen at the video and the laser points at the corresponding location in the host's reality.

Tapping

If the host is in a loud or public environment where audio communication and laser pointing becomes difficult, the guest can use the 'Tap' button to relay a physical tap on the host's shoulder.

Capture, Record

While in the session, the guest can choose to capture moments by clicking the 'Capture' button. They can also record the entire interaction using the 'Record' button.

Participants

The guest can see the profiles of the current participants in the session. The host can also choose to get notifications if new participants join in or existing ones leave.

Step 6. Review

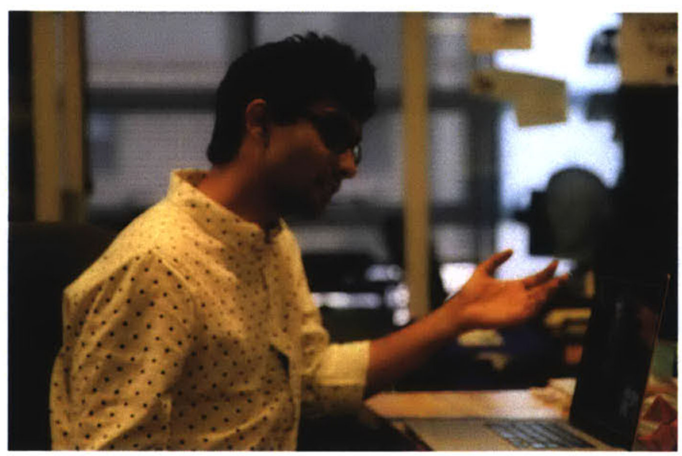
Once the session is complete, the guests can review their captured or recorded interactions. The hosts can choose to make their entire session public for other users to access.

Step 1. Available to Host

Step 2. Wearing the device

Step 3. Guest discovers Host

Step 4. Request and Approval



Step 5 Session : Video, Audio

Step 5 Session : Pointing

Step 5 Session : Tapping

Step 5 Session : Capture Image

3.6 Multiple Hosts, Multiple Guests

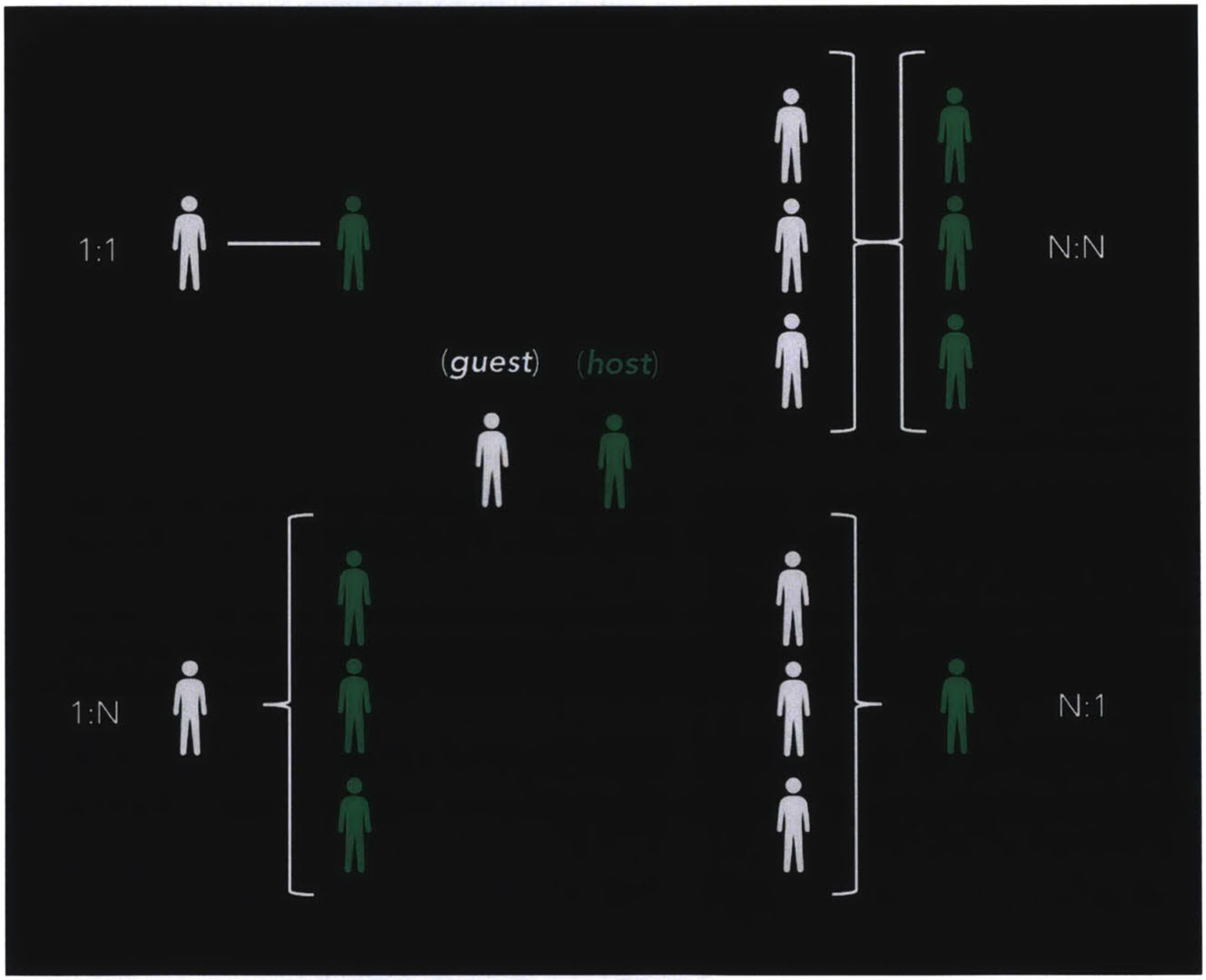


Figure 29: Multiple hosts, Multiple

The above four configurations are possible with the number of hosts and guests in a session.

Each of them lead to interesting use cases and interaction challenges, some of them not fully implemented in the scope of this thesis.

Chapter 4

Implementation

The *icon* system is an interplay between hardware, software, fabrication and an online platform. This chapter dives deeper into the implementation details of each of these parts that make the system a whole.

4.1 Initial Prototypes

Early on in the process of conceiving the idea and testing parts of it, I built a few prototypes; learning's from which have in some form found their way into the current system that exists today.

4.1.1 Exploratory Prototype : Smartphone +Pebble

Prototype: In this prototype, the host carries a smartphone paired with a bluetooth audio earpiece and wears a Pebble smart watch on their wrist. The guest has a smartphone with a web app that can send short texts to the host's Pebble. The host's Pebble has a custom app installed which displays these texts received from the guest. The accompanying Pebble app on the host's smartphone initially receives these texts which are then pushed to the paired Pebble.

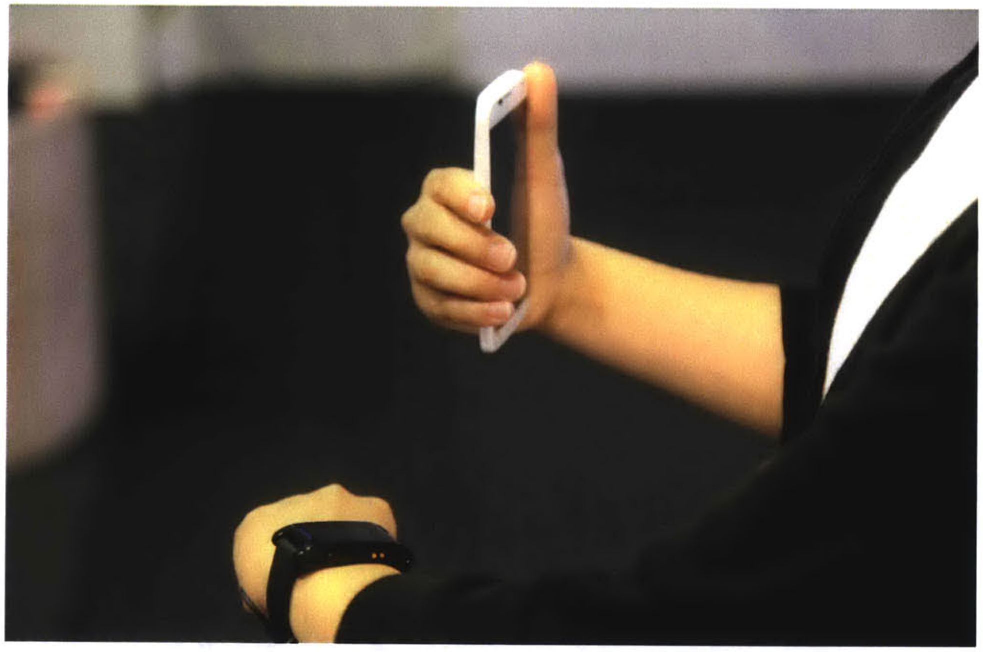


Figure 30: Smartphone back camera streams video, audio, pebble notifies

Lessons learnt: The audio communication was effective to convey details but since there was no way for the guest to see the host's environment - most of the communication involved an extra layer of the host having to graphically describe what they saw. The Pebble served limited purpose, mainly in transliterating words which were difficult for the host to pronounce just by listening.

4.1.2 Exploratory Prototype : Laser-pointer

Prototype: The aim of these exploration was to experiment the effectiveness of using a laser pointer as a substitute to pointing with fingers in a conversation. One user would engage in a conversation with another user, instead of pointing fingers or using any other body cue they were only allowed to use a laser pointer.

Lessons learnt: The laser turned out to be extremely accurate and unambiguous in its use for pointing.

4.1.3 Exploratory Prototype : Google Glass

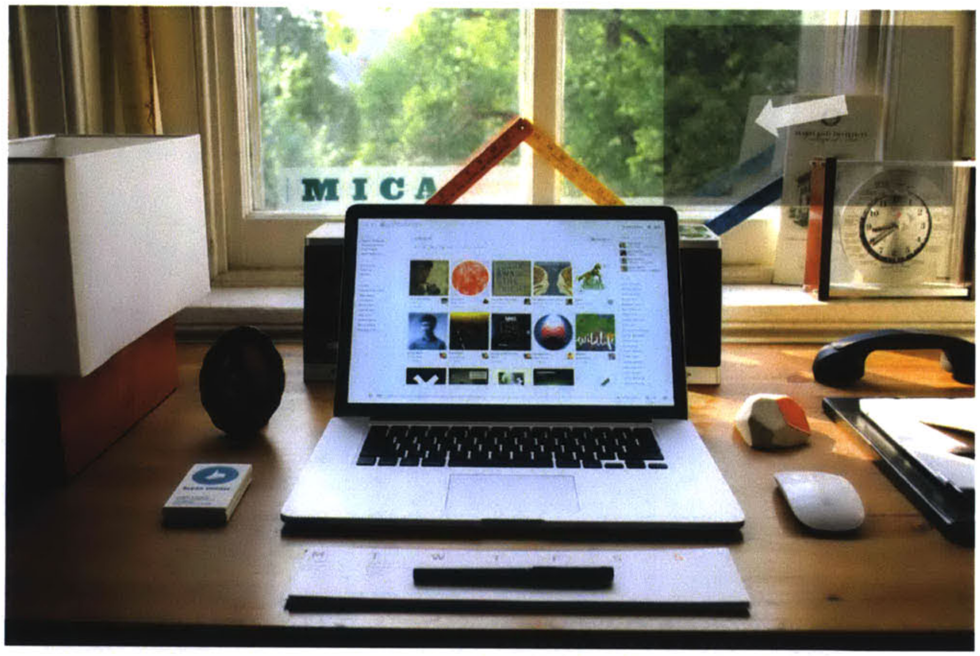


Figure 31: Google Glass point of view, arrow indicates object pointed

Prototype: This exploration is pretty close to the current one in the sense that it facilitates comparable augmentations. The guest can view a live feed from the host's Glass. The guest can then touch / click on the object to be pointed and an associated crosshair appears on the host's prism display on the Glass.

Lessons learnt: However with the limited area that the Glass's display offers, it required further visual interpretation and steps before the host could direct their vision to the object being pointed at.

4.1.4 Exploratory Prototype : Raspberry Pi

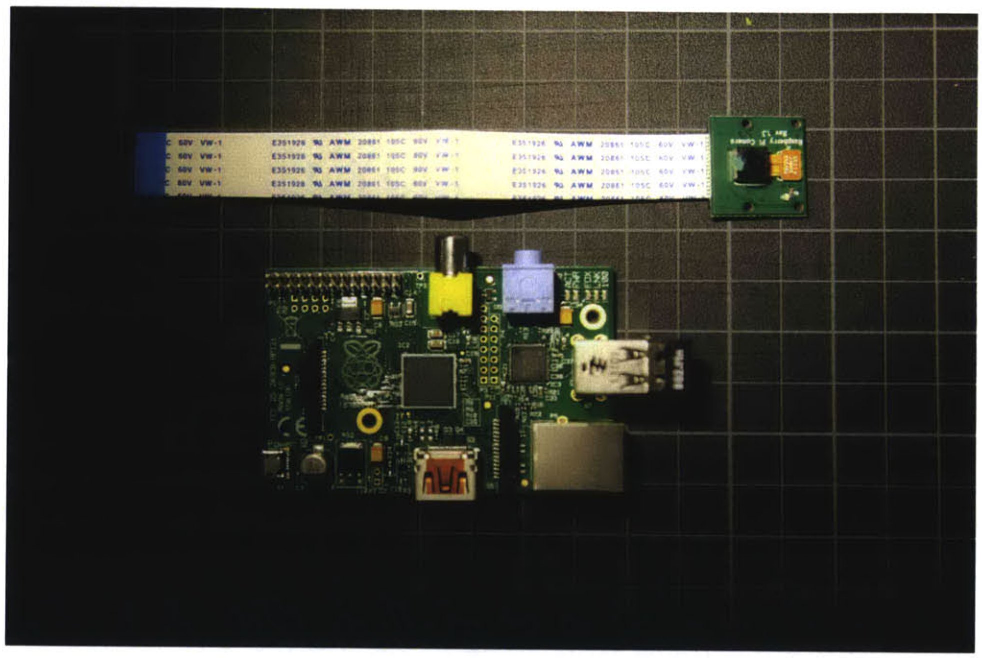


Figure 32: Raspberry Pi with the RasCam module

Prototype: This implementation driven exploration involved deploying the entire device-software on the Raspberry Pi in an effort to make *icon* a standalone device. The Pi was loaded with Raspbian - a Debian flavor for the Pi. A combination of the mjpegstreamer, raspvid, ffmpeg were used to capture video as JPEG frames to minimize file size, were compressed and streamed over the network as UDP packets.

Lessons learnt: The major drawback of this implementation was the limited platform compatibility of the resulting stream. The guest needed to install a special native app to decode the stream and view it. Viewing on a smartphone or tablet wouldn't be possible without a special app.

4.2 Implementation Guidelines

Taking into consideration the end user experience, functionalities, we selected the following guidelines for the implementation strategy.

a. Interfaceless Interface.

Since the device would be worn on the body, I wanted it to have a usability similar to a jewelry or a piece of clothing that is also worn on the body. There should be no visible interface layer. Most of the actions should be accomplished with minimal direct interaction with the wearable.

b. Cross - Platform Compatibility.

I wanted to ensure that the guest can use the *icon* system from any device or platform, be it a smartphone, tablet or a desktop. Building web-based technologies even for the video-audio communication seemed like the best way to ensure cross-platform compatibility.

c. Reusability.

The average smartphone is a robust piece of engineering with all the required processing power, sensors and communication technology built into its small form factor. So it seemed obvious to reuse a currently existing gadget to power the *icon* device rather than building something from scratch as was explored with the Raspberry Pi prototype.

4.3 Hardware

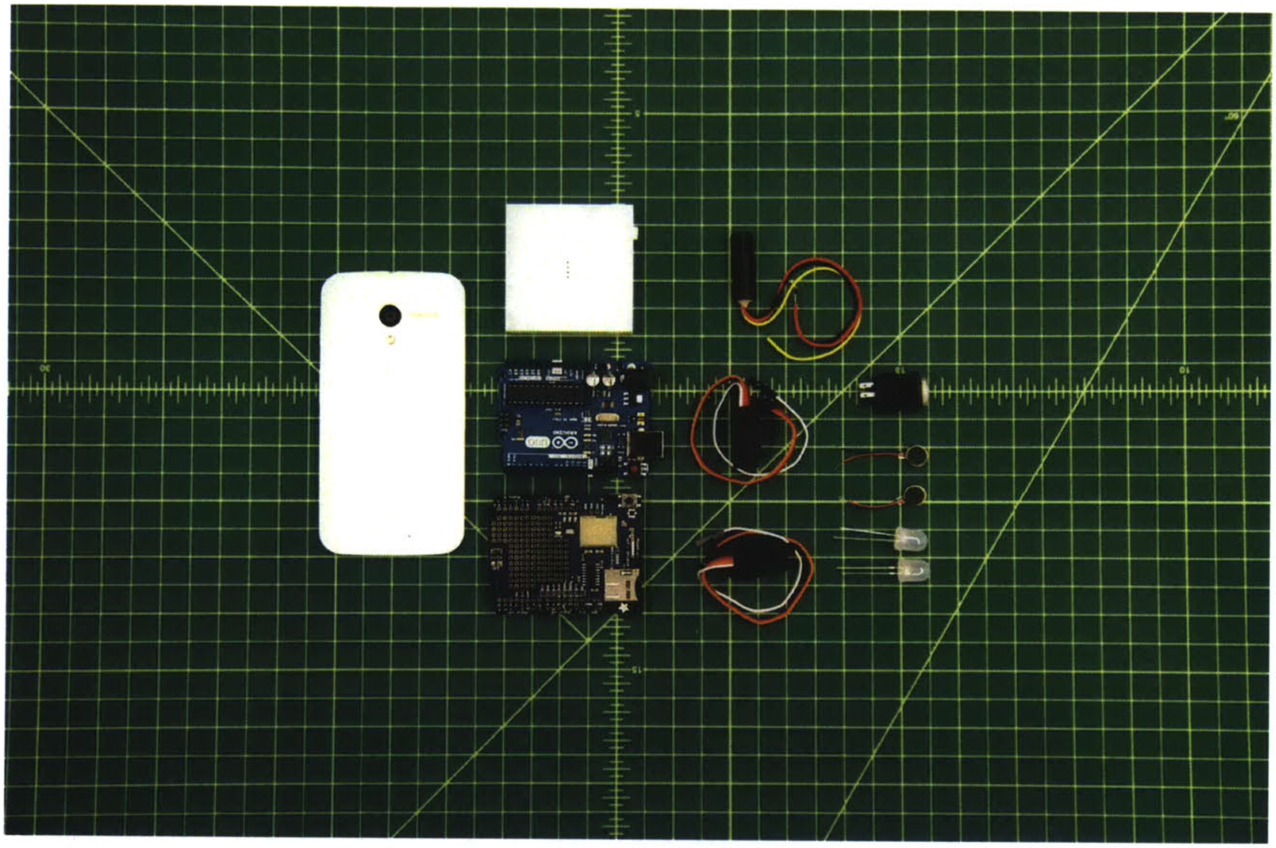


Figure 33: Hardware components in the icon Device

4.3.1 Device

The wearable device has the following hardware components:

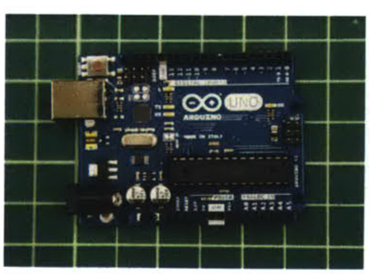


Figure 34: ATmega328 based Uno

Microcontroller

We use the ATmega328 based Uno micro controller from the Arduino family. This resourceful controller has easily interfaceable digital and analog pins. Specifically it has 6 PWM pins which are used modulating the duration of the pulse. This allows us to precisely control the servo motors in the laser pointing system.

We program the Uno using the Arduino IDE which is forked out of the Processing IDE. The programming language we use is based of C and C++ which the IDE compiles into a series of machine code.

Wireless Communication

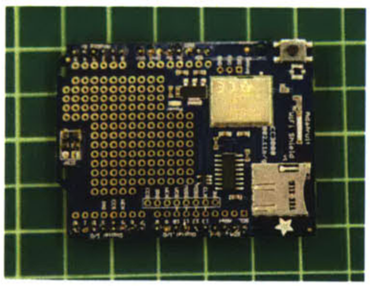


Figure 35: TI CC3000 wireless chip

We use the CC3000 board to enable wireless connectivity to our Uno board. It uses SPI for communication which allows us to push data as fast as we want or as slow as we want. It has a proper interrupt system with IRQ pin so we can have asynchronous connections. It supports 802.11b/g, open/WEP/WPA/WPA2 security, TKIP & AES. A built in TCP/IP stack with a "BSD socket" interface. TCP and UDP in both client and server mode, up to 4 concurrent sockets.

It has an onboard 3.3V regulator that can handle the 350mA peak current, and a level shifter to allow 3 or 5V logic level. It has an onboard ceramic antennae that gives us a good range in a smaller footprint.

We use the CC3000 library which handles connections seamlessly. On boot up the board either searches for an open network if none is supplied or joins the network specified.

```
// Set up CC3000 and get connected to the wireless network.
if (!cc3000.begin())
{
  while(1);
}
if (!cc3000.connectToAP(WLAN_SSID, WLAN_PASS, WLAN_SECURITY)) {
  while(1);
}
while (!cc3000.checkDHCP())
{
  delay(100);
}
Serial.println();
```

Figure 36: Connecting to the network

Once the board is on the wireless network, it acquired a dynamic IP. However for our purposes to communicate with the board from the *iconView* app we need a fixed address to avoid updating the code each time the micro controller joins a wireless network and is assigned a different IP address. We solve this by using the mDNS library along with the Bonjour service which implements a multicast DNS query support for the Uno.

```
// Start multicast DNS responder
if (!mdns.begin("iconyun", cc3000)) {
  Serial.println(F("Error setting up MDNS responder!"));
  while(1);
}
```

Figure 37: Setting up mDNS

Now we can address our board without having to query the IP address each and every time.

REST API

Now that we have our micro controller wirelessly connected to the internet, we need to establish a protocol to send commands and receive feedback from it. One option is to use a streaming serial data connection, however it becomes cryptic and messy to handle as the complexity of the project grows.

For this purpose, we write our own implementation of REST API which runs on the micro controller. The REST API (Representational state transfer) is a software architecture style that leads to much cleaner URIs and is scalable, visible and portable. The REST API architecture is defined by a configuration of components, connectors, and data.

An example of using a REST API to execute a command on the digital pin would be:

```
$.get( "http://iconyun.local/arduino/digital/3/1" );
```

Figure 38: REST call to set digital pin 3 to HIGH on the icon device

And now since, we can make these API calls to control the logic of the board we don't necessarily need to program the board beforehand, we can send it command logic on the fly. Our REST library implementation supports the mode, digital and analog commands.

In response to the commands, we program the micro controller to respond its execution status in the JSON format.

```
{"Message": "Pin D3 set to HIGH", "id": "001", "name": "iconyun", "connected": true}
```

Figure 39: JSON response to our REST query

We also implement a REST API call to access state variables which can give us real-time information of the processes and status on the board. For example :

```
$.get( "http://iconyun.local/arduino/motorXposition/" );
```

```
{"Message": "50", "id": "002", "name": "iconyun", "connected": true}
```

Figure 40: REST call to find the value of variable motorXposition and the JSON response to our query

We extend this to allow calling functions with REST calls.

```
$.get( "http://iconyun.local/arduino/tap/" );
```

```
{"Message": "Function Tap Executed", "id": "001", "name": "iconyun", "connected": true}
```

Figure 41: REST call to execute the function 'tap' and the JSON response to our query

We can also pass parameters with these function calls, this is where the portability of our REST architecture shines.

```
$.get( "http://iconyun.local/arduino/tap/10" );
```

```
{"Message": "Function Tap Executed (Duration : 10sec)", "id": "001", "name": "iconyun", "connected": true}
```

Figure 42: REST call to execute the function 'tap' for a duration of 10 seconds and the JSON response to our query

Laser

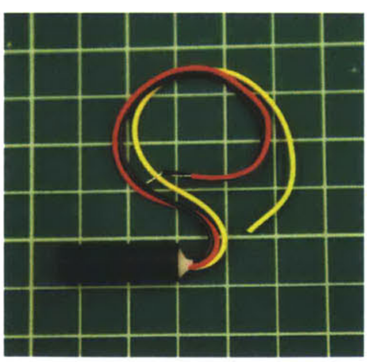


Figure 43: 5mW 650nm Class IIIa modulated laser

We use a TTL laser diode which produces a 650nm wavelength laser while rated Class IIIa and safe for everyday use at 5mW. This module in particular is designed to be modulated/pulsed via the trip wire. This extends its functionality much more than a mere ON/OFF laser. With pulse width modulation we can program the laser to have smoother fades which makes them comfortable on the eye. There is of course the danger of someone pointing the laser in someone's eye, this can be overcome by implementing a simple face detection algorithm in the video feed. This however was not implemented in the scope of this thesis.

Pointing System

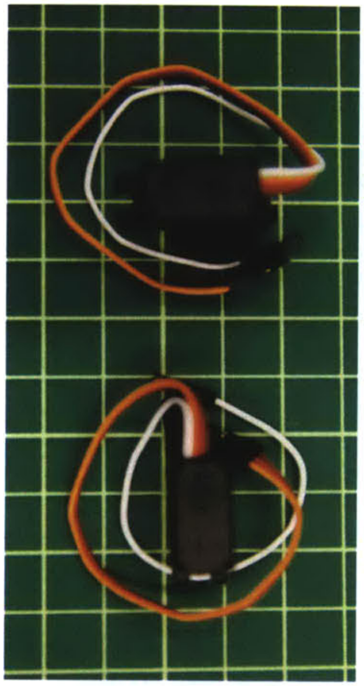


Figure 44: Analog feedback servo motor

We use two custom made analog feedback servo motors as part of the 2-axis pointing system.

Being an analog feedback servo allows us to get information back from the motor readable with a pin header regarding the current location of the motor head. This feedback loop allows us to monitor and auto correct the position instead of simply relying on the motor to execute the rotation correctly every time.

To make a 2-axis pointing system, we mount the motors on top of each other. By specifying the angle of rotations in the two planes, we can point the laser to any given position in the hemisphere in front of the host.

The motors have a 180 degree coverage and using the REST API the *iconView* app can send precise commands wirelessly to the pointing system.

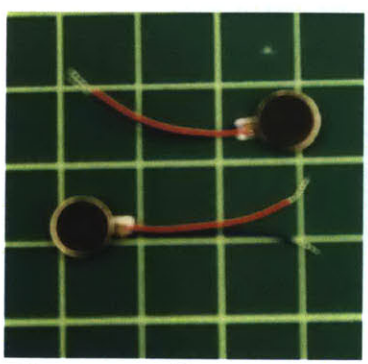


Figure 45: Vibrotactile motors

Tapping and Lights

Three vibrotactile motors are mounted on the base of the pointing system above the shoulder. These are programmed to emulate the effect of tapping on the shoulder. The *iconView* app can send a REST call to the micro controller to activate the motors.

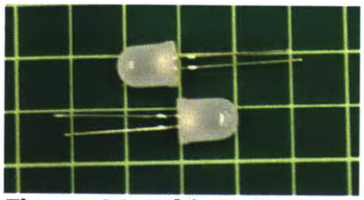


Figure 46: White LED

When a new guest jumps in to the session, the device gives out a whit breathing glow. We use bright white LEDs placed underneath the shell of the device. These 5mm LEDs give out 20,000mCd brightness and are programmed to emulate a breathing fade-in fade-out pattern.

Figure 47: MotoX

Smartphone

As discussed earlier, we use the smartphone for its embedded camera, mic, speaker and its ability to execute web-based code to stream this video and audio data with the *iconView* app. The smartphone doesn't interfere with the other hardware on the device and hence is modular, any modern smartphone would work. We have implemented a webRTC based video and audio streaming mechanism which is discussed in greater detail in the software sub-section of this chapter. The host could wear a bluetooth headset for the audio or use the smartphone speakers.

Power

Most of the components in our circuit are low powered, we use a 5V alkaline (MnO2) battery which is rated 500mAh at a discharge of 50mA.

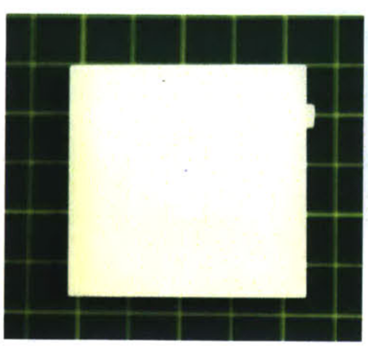


Figure 48: 5V DC-18 battery enclosed in casing

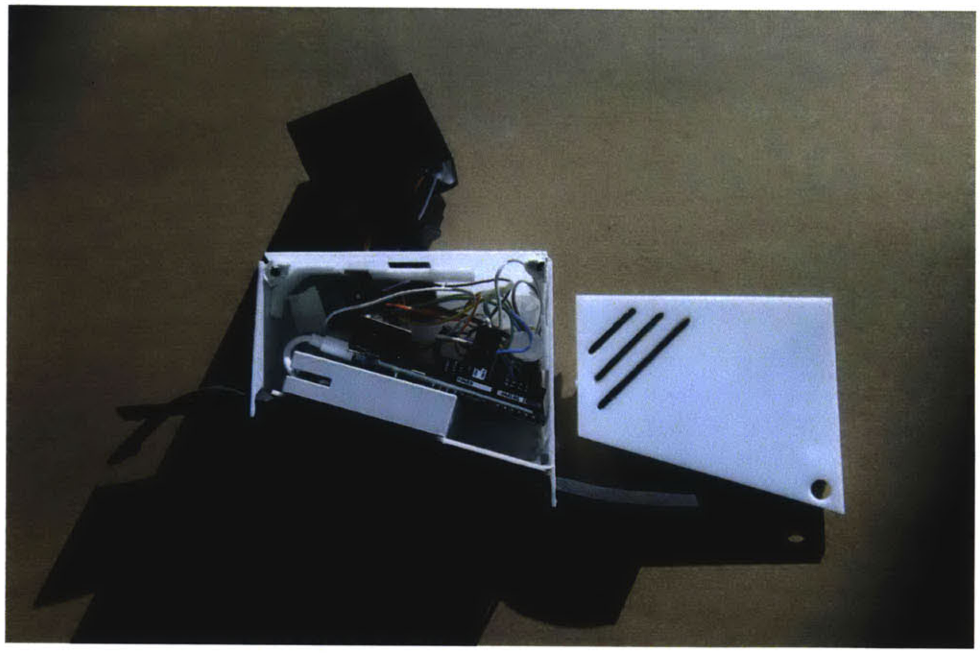


Figure 49: Assembled Hardware

4.3.2 Fabrication and Aesthetics

Design Considerations:

a. Placement

The body has a limited real estate. Since the device streams a point of view video, the device had to be in the upper part of the body above the chest. We decided against placing the device on the head with the amount of hardware we have. The shoulder came in as a great place where the device could stream a point of view video and had enough space on the shoulder to house the hardware. Anatomically the shoulder serves as a connector between the hands and the upper body, we wouldn't be compromising or intervening any critical body functions by designing the device around the shoulder.

b. Weight Distribution

Since the device was to be placed on the shoulder which sees a lot of motion, it was critical to have an equal weight distribution in the front, top and behind portions of the shoulder. Also a snapping mechanism was to be in place to keep the device intact on the shoulder in case the host bends over or turns to the side.

c. Forgiving

To accommodate the various postures and motions of the shoulder, the device had to be designed as a one functioning whole comprised of three ergonomic parts that would be connected together. These connections had to be of a material that was much forgiving and that could blend in to the dynamic shape of the shoulder.

Materials

The exterior casing for the micro controller, X-motor and battery were laser cut from a thin white plexiglass. The casing for the Y-motor and the laser were cut from a translucent black plexiglass. This material is durable and doesn't degrade with time.

Snaps

The device was snapped to the host's shoulder using magnets that were attached to an anti-slip rubber surface which allow for easy snap on, snap off without much inconvenience.

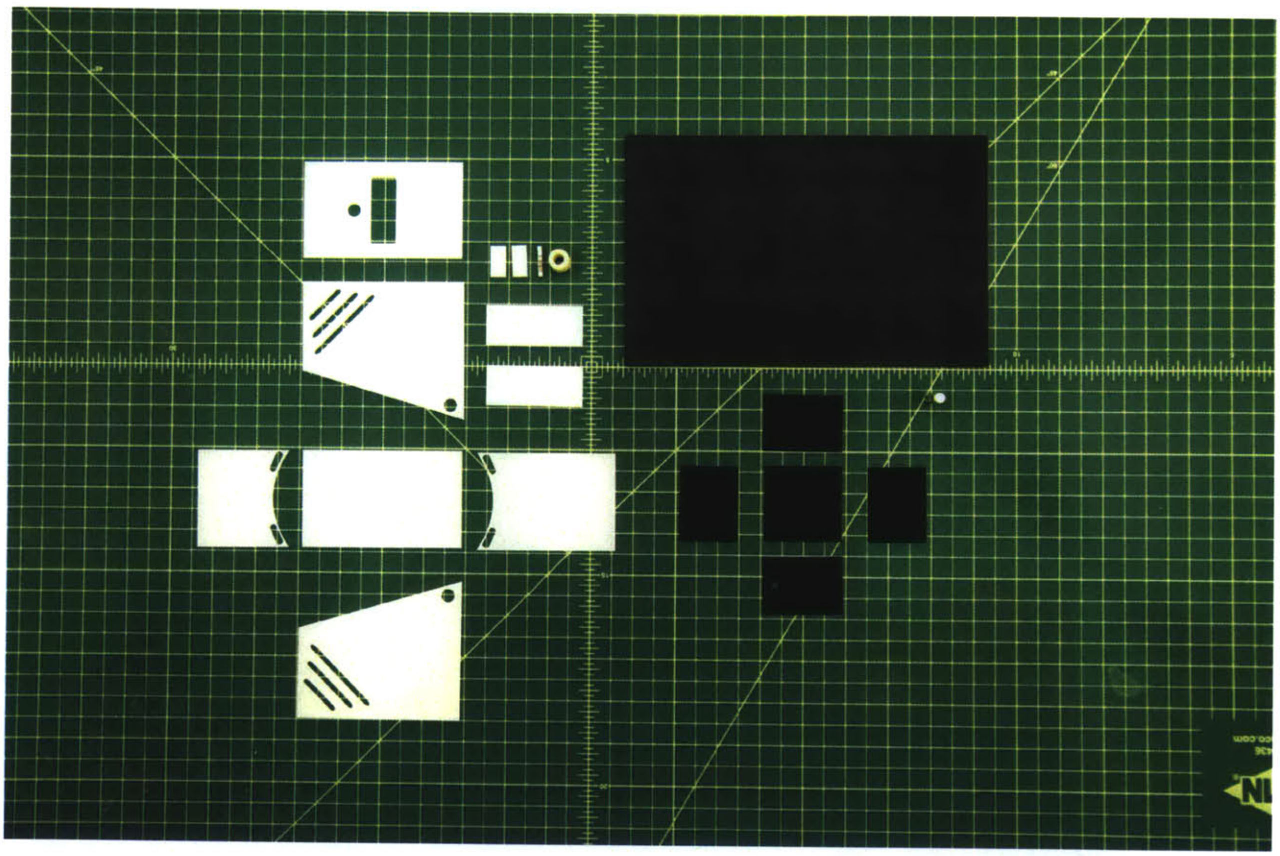


Figure 50: Plexiglass pieces, rubber, magnets before assembly

Aesthetics and Ergonomics

The device aesthetics are metaphorically inspired and geometrically rendered. Metaphorically, they remind us of friendly pets such as parrots, cats sitting on the shoulders of their owners or of mythological representations of angels and devils

Figure 51: Metamorphosis to an icon

that whisper in our ears. The *icon* system poetically serve the same purpose, as a portal to a friend far away and one from

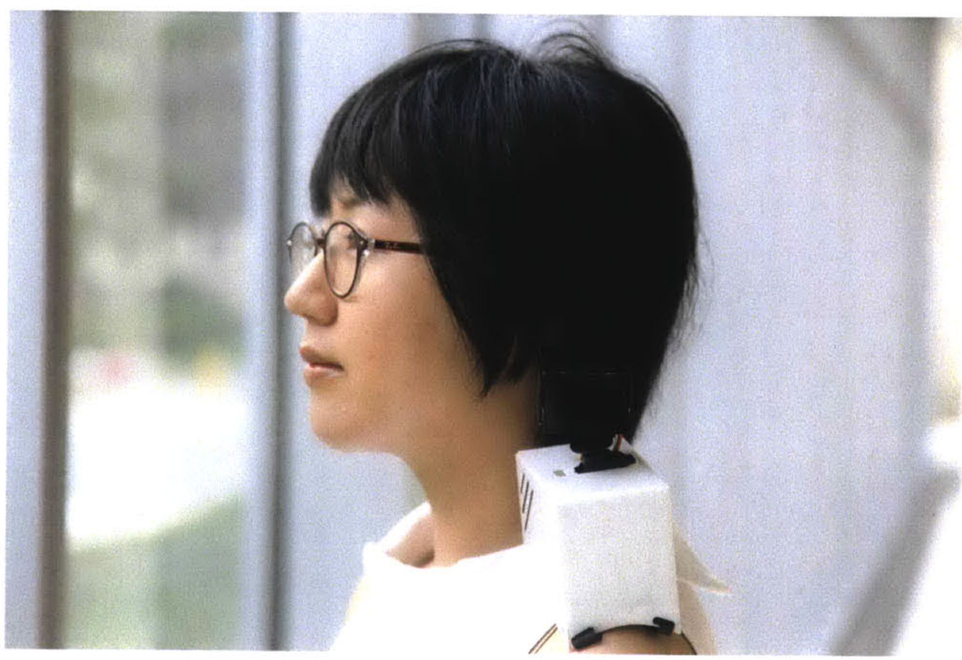


Figure 52: icon + User

whom we learn and take advice. The geometrical renderings of the shape are designed to optimally wrap around the hardware conserving space. The angular cuts also follow the natural folds of the shoulder and neck allowing them to flow into the device rather than the device obstructing natural motion.

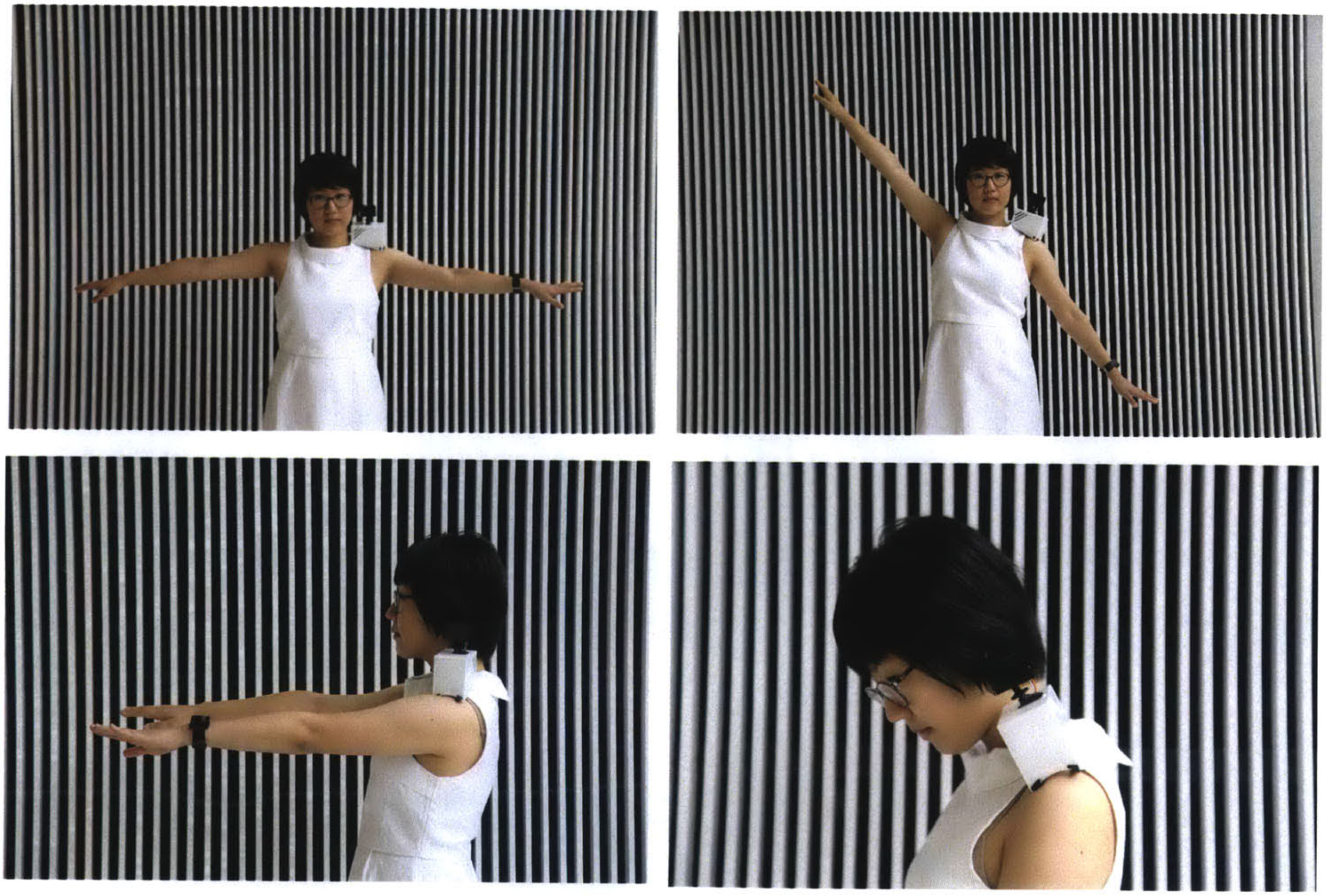


Figure 53: Ergonomics : Device stability for various body postures

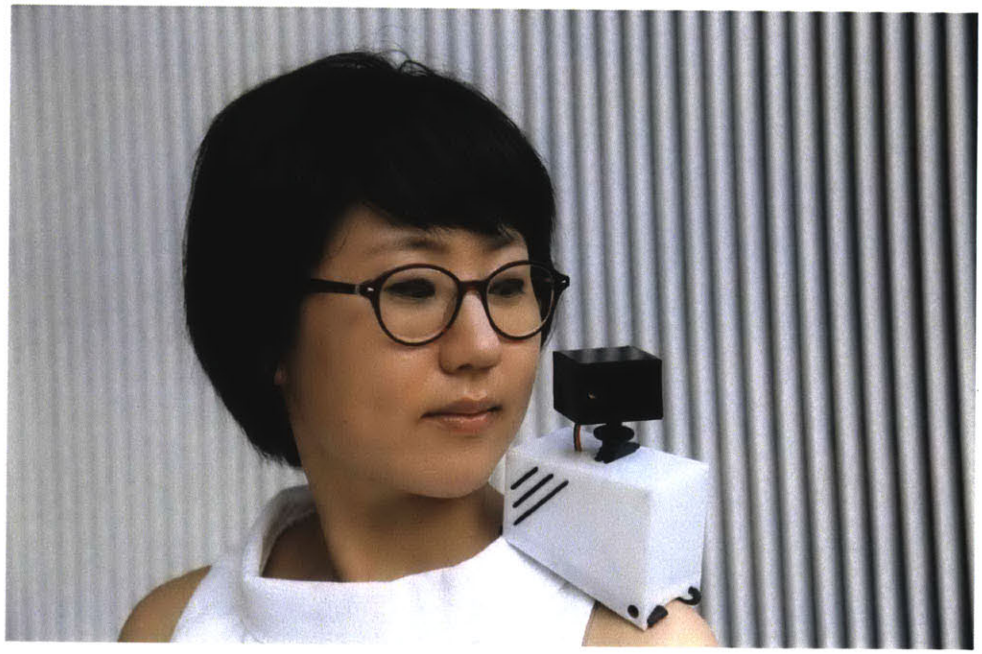


Figure 54: Ergonomics : Turning head to the left

4.4 Software

4.4.1 iconView app

The *iconView* app is used by the guest to communicate with the host and experience their reality. This app allows video and audio streaming, tapping, capturing, recording as well as makes real-time calculations for the laser pointing system and handles asynchronous requests sent to the device. The app is written purely in javascript, html and css3. This allows the app to run cross-platform on most browsers and device types including smartphones, tablets and desktop computers.

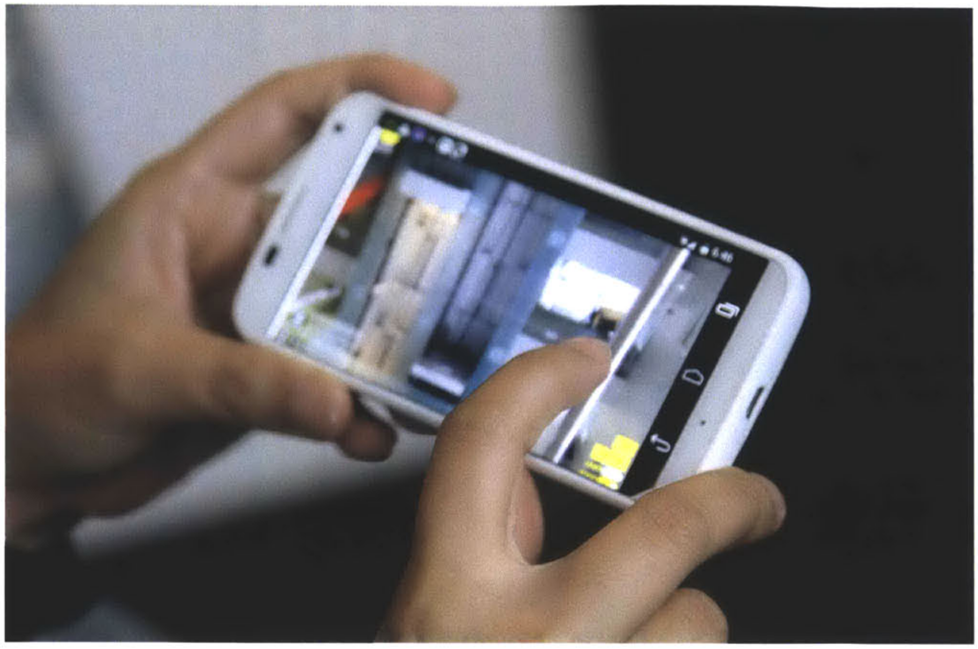


Figure 55: iconView App

WebRTC

```
navigator.getUserMedia = navigator.getUserMedia || navigator.webkitGetUserMedia || navigator.mozGetUserMedia;
navigator.getUserMedia({video: true, audio: true}, function(stream) {
  var call = peer.call('another-peers-id', stream);
  call.on('stream', function(remoteStream) {
    // Show stream in some <video> element.
  });
}, function(err) {
  console.log('Failed to get local stream' ,err);
});
```

```
navigator.getUserMedia = navigator.getUserMedia || navigator.webkitGetUserMedia || navigator.mozGetUserMedia;
peer.on('call', function(call) {
  navigator.getUserMedia({video: true, audio: true}, function(stream) {
    call.answer(stream); // Answer the call with an A/V stream.
    call.on('stream', function(remoteStream) {
      // Show stream in some <video> element.
    });
  }, function(err) {
    console.log('Failed to get local stream' ,err);
  });
});
```

Figure 56: Initiating and answering a call through WebRTC

WebRTC (Web Real Time Communication) is a set of standards that support browser-to-browser applications for voice calling, video chat, and P2P file sharing. The major advantages of using

WebRTC is that it runs natively in most modern browsers without external plugins. We have implemented a peer-to-peer media connection between the device and *iconView* app to stream two-way audio and one-way video. Node.js serves as the backbone for the service, we run a PeerJS server which brokers connections between the peers and the PeerJS library serves as a wrapper on top of the browser's WebRTC implementation to ensure consistency.

Perspetive estimations

We calculate the input values to the laser pointing system from the guest input on the screen. The guest touches / clicks on the screen are captured by the following code.

```
var x = new Number();
var y = new Number();
var canvas = document.getElementById("subscriber");
console.log("getPosition");
if (event.x != undefined && event.y != undefined)
{
  x = event.x;
  y = event.y;
}
else // Firefox method to get the position
{
  x = event.clientX + document.body.scrollLeft +
      document.documentElement.scrollLeft;
  y = event.clientY + document.body.scrollTop +
      document.documentElement.scrollTop;
}

x -= canvas.offsetLeft;
y -= canvas.offsetTop;
```

Figure 57: Capturing x, y coordinates

These values are in cartesian co-ordinates which we convert to spherical co-ordinates using this formulae.

$$\alpha=\arctan(x/d)$$
$$\beta=\arctan(y/d)$$

where d is the distance of the surface from the device.

However these values in spherical co-ordinates correspond to the location of the camera. Our hardware pointing system is offset from the camera on the y-axis. In order to shift the origin of our spherical co-ordinates to that of the pointing system we use this formulae.

$$\beta_n = \arctan(d/(c+\arctan(\alpha)*d))$$

where c is the distance of the laser pointer from the smartphone camera.

These values in radians once transformed to degrees are ready to be sent over to our laser pointing system using the REST API previously discussed.

We do have an approximation error in these calculations since we assume a constant distance range between the host and the objects (unless of course the focus is at infinity), however as it turns out - from a usability point of view the error margin is trivialized since the guest and host are also communicating through audio and the fact that the resolution required in most of our learning applications is at the level of centimeters.

Record, Capture

The guest can capture or record a session to review them at a later point. When capturing a session, the client side javascript simply stores the frame information and sends this to the server which successfully extracts the associated frame since all sessions are also written to disk on the server.

The record feature allows the guest to capture the entire session and review or share it a later point. The server writes all sessions to disk while concurrently streaming them. These sessions are made accessible on a web-embedded player to users who chose to record the sessions.

Communication

When the guest points to an object on the screen, taps, captures or a new user joins in; the app needs to send these to the hardware device to execute the associated actions of pointing the laser, actuating the vibrotactile motors, fading the lights. This communication between the app and the device is asynchronous

```
$.get( "http://iconyun.local/arduino/point/30/72/" );
```

Figure 58: An example XHR request to point the laser at 30^{rad} and 72^{rad}

in nature such that the video and audio streaming are not affected. We use XMLHttpRequest (XHR) APIs in javascript to send these requests. Following is a snippet of an XHR API request sent to the device.

4.4.2 Platform

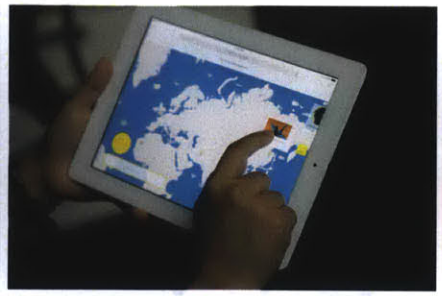


Figure 59: iconHub platform

iconHub

The *iconHub* is the platform where guest can discover or request for sessions and the hosts can create new sessions. The front end of the platform is written in javascript, html and css3. The backend uses a MAMP stack with PHP, MySQL and Apache.

Creating Sessions

The host can create their profiles with their interests, profile picture, screen name and location. These get stored in the database in the user table. The hosts can then decide to create a

id	uname	access	title	description	startTime	endTime	...
1	soniak	open	Teach...	We'll learn h...	21:03:01...	22:03:01...	...
43	lizim	closed	Fixin...	I'm going to...	12:13:45...	12:43:45...	...
...	...	...	...	...	...	...	...

Figure : Database schema for sessions

new session, where they supply the name and a short description of the session. These get stored in a the session table. The hosts can either schedule the session for the future or start it right away.

They can also mark the sessions as open or closed depending on if the guests can simply join in with or without approval from the host.

Jumping in Sessions

Once the guests choose an activity, they see all the live sessions in a map view. This is rendered with a stylized Google Map API v3 where they can explore the sessions, their titles and description. These are retrieved from the session table from the database. The guests can also do a text based search for particular sessions or if they can't find sessions they wish to join, they can request for sessions in the future. These requests are stored in the request table in the database. Depending on if the session is open or closed the guest may or may not require approval from the host to join. Once in an active session they can also see the profiles of other guests in the session.

Chapter 5
Applications and Evaluation

The *icon* system is used in activities which require hands-on immersive experiences and it qualifies itself in facilitating such activities at a distance where traditionally they would be co-located. It believes that everyone is a teacher and a life long learner and by providing a medium which can seamlessly help transfer intentions across distances we can all enrich our lives at every step. Owing to the design choices in this thesis, the system is comprised of a stationary guest and a mobile host. This chapter describes some applications of this system. As with everything human, someone's question is another one's answer. I hope *icon* is an answer to someone's question that I haven't yet come across.

5.1 Outdoors and Crafts

5.1.1 Outdoors

Synopsis:

The *icon* system can be used to learn various things, having a first person point of view facilitates a much more immersive learning experience. Depending on the particular situation, either the host or guest can be the learner.

Scenario:

Say you want to learn about the different flower species in your surroundings. You wear the device, find a guest who wishes to teach you or request for a guest who can. You're now in a session

Figure 60: Learning about flowers from a guest

together with the guest or maybe multiple guests. They can be a part of your reality while being far away and interact with you, you walk around your surroundings talking to them while they point out things in your reality. As they point out they describe what you see and through your interactions you immerse in a deep enriching conversation. You can go back and review your session with them or share it with others.

5.1.2 Crafts

Synopsis:

Owing to the pointing system, the *icon* system facilitates teaching over distance. Either the guest or host can be the teacher depending on the particular situation.

Scenario:

You have something to teach, say you want to share your love of origami with someone. You create a session, wear the device, guests who want to learn join in. You show them your reality, your desk with origami paper, where you take them through how to make a paper crane. They follow your lead, they interrupt you, they point at a fold or a tool, ask questions, you talk to them, you show them. If you're too engrossed or in a crowded area, they tap on your shoulder. They can capture or record the session and review it at a later point in their learning process. You may teach one guest or many.

Figure 61: Teaching origami to a guest

5.2 Civic and Cultural Exchange

5.2.1 Witness

Figure 62: Witnessing and remote interacting with a crowd at Kendall

Synopsis:

The point of view camera and the pointing system along with a host who is a local, can allow multiple guests to witness and interact with the people in the host's environment.

Scenario:

We live in an aware world, a world connected more than ever, technology has accelerated collective empathy. You see the protest in downtown Beijing on YouTube, you wonder what it would be like to be there, what are the people thinking, you want to peek deeper than what the biased media shows you, you want to be with the people. You find a host who is in the middle of the protest, you jump in the session, now you see a first person point of view, you hear their environment. You can ask them to go and talk to people, the host becomes your voice, your body. Or you could be the person in the middle of the protest and you host guests from around the world, tens or hundreds of them, you become their eyes and voice. It may not be a protest, but just a unique cultural phenomenon.

5.2.2 Travel

Figure 63: Traveling around Boston

Synopsis:

Hosts can take the guests through a journey of their surroundings, guests can travel and interact with the surroundings of the host.

Scenario:

We all yearn to travel, to wander and see beyond, to taste, to talk, to immerse. Or we all are hospitable guests inside, maybe its not our virtue but a human need to share and to find roots. You find a host in Bhutan, she takes you around the markets of Paro, you see her world, you point at streets, objects, lanes - ask them while curiously pondering, maybe they could take that detour. Or you are the host wearing the device, you walk your way to work, nothing special, its just another sunny day. But your guests seem enthralled with the subway, that graffiti on the wall, the chaos in the streets, you talk about it, inside, you smile, you discovered a new city in the same old one.

5.3 Physical Tasks and Remote Help

5.3.1 Physical Tasks

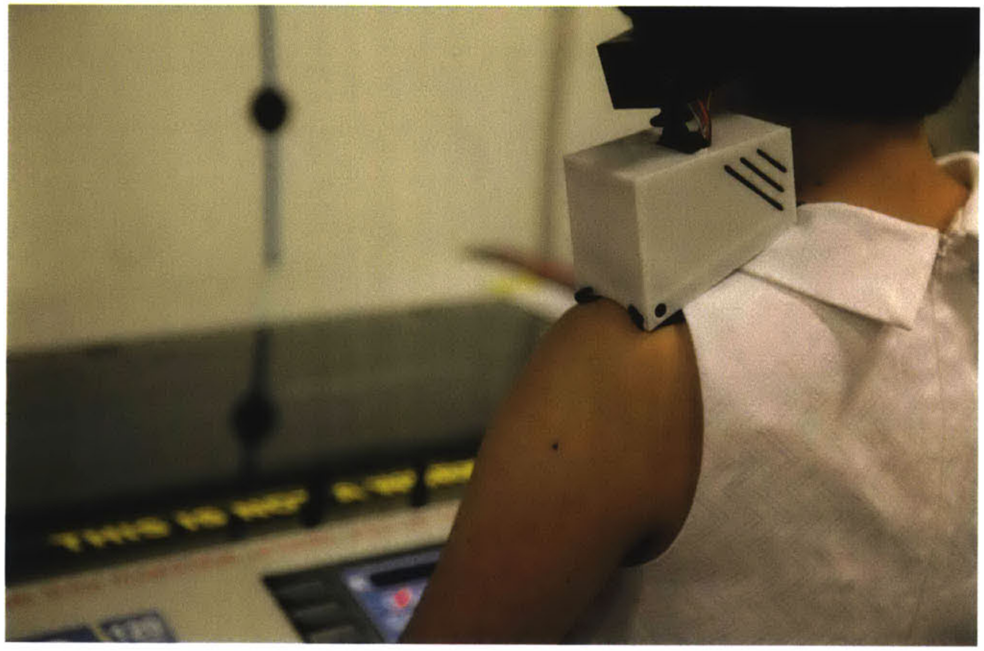


Figure 64: Creating a laser-cut jewelry with instructions from a guest

Synopsis:

The *icon* system with its distributed hosts and guests across cultures and the unique communication medium can be used to learn crafts, cooking and a plethora of other subtle cultural nooks.

Scenario:

You're cooking, or you're operating a laser cutter to make that craft object you always wanted. You're doing it yourself. You read the instructions in that blog post multiple times, you saw that video on YouTube but what you made looks different, you're stuck and don't know whom to ask. Maybe you followed all the steps right but don't why it isn't working. You wear the device and someone else, not an expert but someone who just did what you're trying to do - jumps in to your session. They walk you through the steps, they point out and ask your a question about that step, they talk to you. You frustration turns into excitement, you connected with a real person, both of you're better for this interaction. There is often not one right way to do something, but by following someone else's instructions, you don't create something new. Now you do, you take your steps, your guest learns your mistakes or your ingenuity.

5.3.2 Help

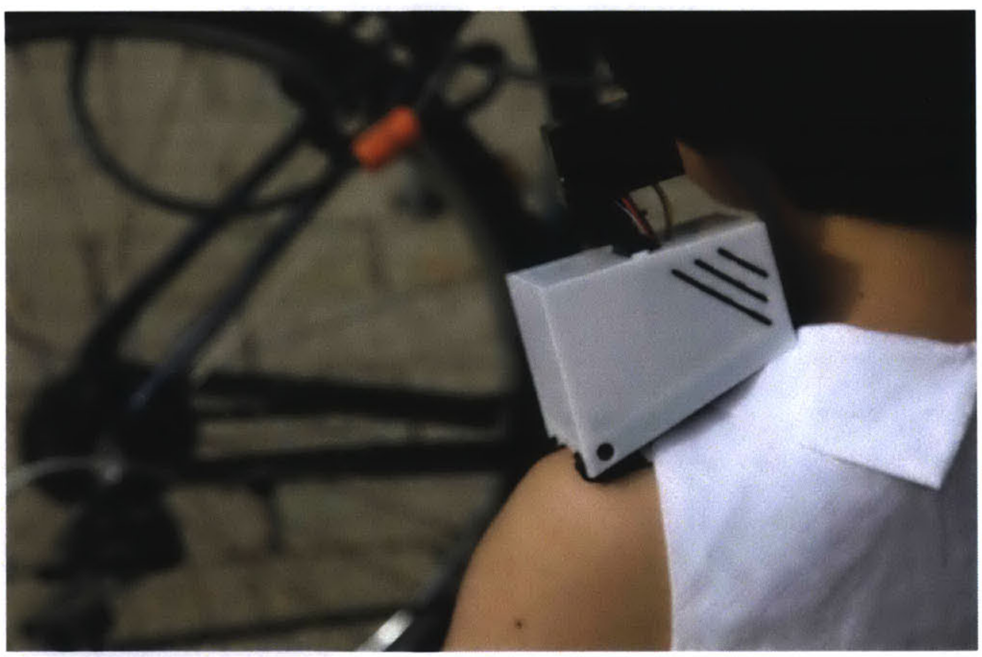


Figure 65: Fixing a bike chain with help from a guest

Synopsis:

With the *icon* device, hosts can now get tailored help from customer service reps or from helpful guests.

Scenario:

You're on the road, your chain slipped off the gear. You have your device on, you ask for help. Someone, anyone. This guest joins in, they could be a bike expert, but chances are they are as bike-literate as you. But now you're no more alone on that road, there is someone. You hear their opinions, while you try and fix your bike, they point out things you missed, you overlooked.

Maybe you want to check out three apartments in Seattle where you're moving. Sitting in your home in Boston, you hire three hosts for help, they go to the three apartment at the same time. You see their feed in a split screen. You check out all three apartments, at the same time. Your hosts are your body.

5.4 Other Application Areas

5.4.1 Sports, Music

Synopsis:

Coaches can monitor and help players distributed across the field using the pointing system. Likewise masters can teach students across the world.

Scenario:

You're a mountaineer, each rock you put your hand on, each step you take could be a step upward or a fall downwards. You're

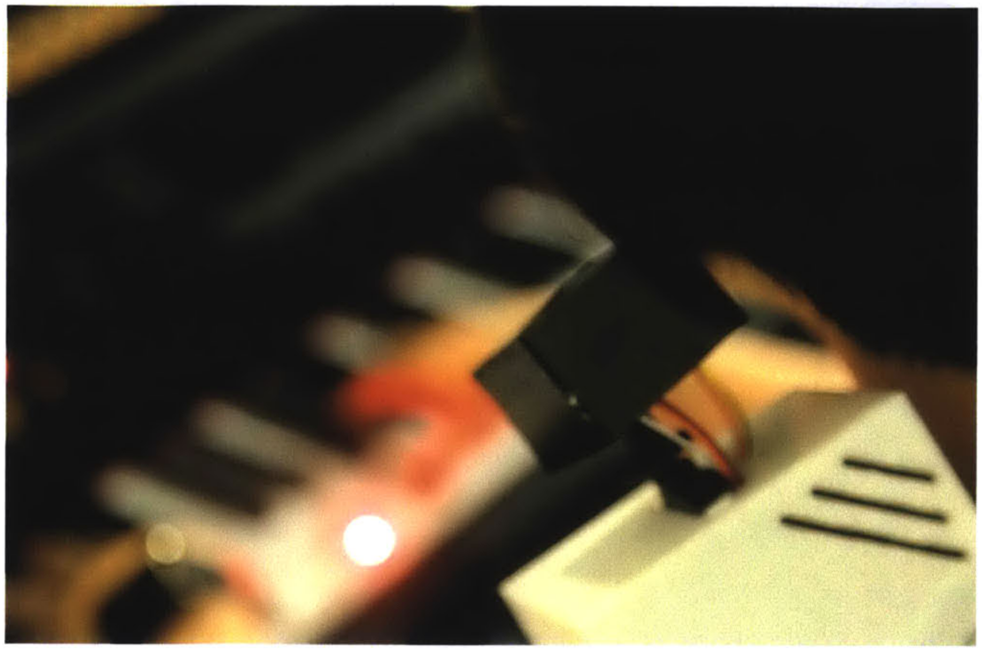


Figure 66: A master (guest) monitoring one of his several students

wearing the device, and so are your fellow climbers. Your coach cycles through each of your point of view, seeing what you see. The point at that crack you missed, or warn you against that advance. You could be playing football, or the piano, or making a pot. Your teacher is looking over your shoulder, from a far away distance. They don't distract you by their presence, but they're there.

5.4.2 Meet

Synopsis:

Hosts and guests can communicate much more effectively than a video chat software. Even without an active intention to learn, teach et cetera; it can be enriching to meet people from around the world.

Figure 67: Meeting and chatting with guests

Scenario:

We're humans, we're a community. Sometimes the only thing we want is the company of other humans. The *icon* system connects you to others. You can chat, talk, meet new people pretty much like ChatRoulette. It doesn't compromise on your identity, your guest never sees your face, the camera is pointed outward, you never see them either, its a one way video stream. You spend your time with a friendly stranger.

5.5 Evaluation Sessions

5.5.1 Feature Evaluation

The first set of evaluations tested each of the individual features of the *icon* system in isolation. These evaluations were carried with 5 users who had never used the system before.

iconHub Platform

Task: Users were given the task to navigate through the interface without instructions and explain the features as well as describe the experience of finding a host.

Figure 68: Hosts (Human icons)

Feedback: Uses found the interface delightful. Overall users liked the introduction page where once they hover the different activities - they can see a dynamic text describing the activity. Users said they could imaging the possible use cases with the device after reading the short descriptions. Once on the map, all the users mentioned that they like the overall design, organization and color palette on the page. However it wasn't clear to the users that the human icons on the map were of live device wearers. Some suggested that it would be nice to even have the inactive wearers and denote them with a different color.

iconView App

Task: Users were given the task to use the interface during a video call with another user. They were asked to describe the features as they used them.

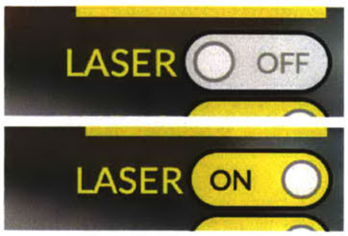


Figure 69: Laser toggle switch

Feedback: The users appreciated the minimal and single color design used in the interface. They were slightly skeptical of trying out the Laser and Record toggle switch but readily tried the Audio toggle. They had questions "Where would the recorded video be stored?", "Will I damage the laser if I turn it ON/OFF frequently?".

Device Ergonomics

Task: Users wore the device and were encouraged to move around and go about their daily tasks. They were asked to report their experience wearing it.

Feedback: Users were amused by the shape of the device, they mentioned it resembled a parrot or a robot. Since some users were wearing stiff clothing, the device couldn't be easily snapped on the shoulder side. After wearing it, they didn't notice any significant weight on their shoulder. Users wearing loose clothing felt that the device was unstable and would occasionally slip forward or backward as they moved. Users were able to move their head normally without the device interrupting their motion. Taking the device off was challenging for the users due to the strong magnets.

Video+Audio Experience

Task: Users were asked to report on the video streaming experience from an *icon* device which was mobile, as well as the 2-way audio streaming.

a. Guest

Feedback: The guest using the *iconView* app commented that the resolution was fairly good to get a sense of the host's environment. They reported unease over the shakes in the video owing to the host walking around, however after a minute of two of adjusting to the moving frame they reported that they had quickly adapted to the shaky video. Users however still preferred the video experience when the host was not in motion. One user reported motion sickness by constantly looking at the first person point of view. The audio experience was clear and real-time for most of the users.

b. Host

Feedback: The host wearing the *icon* device communicated with the guest via audio. The hosts reported that they would prefer wearing a headset than to have the audio loud on the speaker. Users were happy with the audio as a means of communication while they were walking.

Pointing Experience

Task: The guests were asked to use the laser pointer to point at objects in the host's environment. They were asked to report on their experiences.

a. Guest

Feedback: The guests found it a very seamless process to click and see the laser point at the object in the next two seconds (They saw the laser point through the video stream). Users were fascinated at the possibilities, and started pointing at various things in the host's surroundings. One guest asked the host "Can you go by that door?" and pointed the laser before the host asked the follow up question about which door he was talking about. Overall guests quickly started using the laser along with their video and audio communication.

b. Host

Feedback: The hosts reported that not all times the laser shoots exactly where the guest intended, since the hosts are mobile. However since they are also talking to them on audio, an offset error in pointing the laser didn't lead to any misunderstanding. Some hosts mentioned that the laser was much brighter in a darker room and wasn't pleasant to look at. On the other side, when outside in bright sunlight, the laser tends to not be bright enough. However the hosts pointed out that sometimes just by looking at the device motors and the direction in which it was pointing was enough for them to know where the guests were pointing the laser. Some hosts asked if they could turn off the laser in case they run into a irritating guest.

Tapping Experience

Task: The guests tapped the hosts and were asked about their experience.

a. Guest

Feedback: The guests appreciated the tap feature at times when they couldn't communicate, some of them also expressed that it made them feel as if they were physically co-present with the host.

b. Host

Feedback: The hosts mentioned that the tap feature allowed them to reassess the situation and talk to the guest. One host commented that the tap sensation felt too much of a shock, and that hosts should be able to calibrate the feedback. Hosts also mentioned that the tap function should be time-limited to avoid tap-friendly guests.

5.5.2 Activity-based Evaluation

Activity: This evaluation was performed by two users - one guest and one host in separate rooms. The guest was given a pre-constructed structure with LEGO blocks, and the host was given the loose LEGO blocks required to complete the structure. The guest would be helping the host through *icon* to construct the LEGO structure. The same task was performed with the host having a pre-constructed LEGO structure and helping the guest constructing it from loose LEGO blocks.

a. Guest helping the Host

Feedback: Guests started communicating about their LEGo structure via audio to the host. That conveyed a general sense of scale to the host and since the LEGO pieces supplied were all of the same color, the guest couldn't simply refer to them by color. Guests asked the hosts to get the LEGO piece close to the camera to have a look at it. Once they knew which one was where they started pointing at it. In the later stages one guest stopped pointing and started talking about them, "Pick that small one to your left", "Now take the one next to it". A very premature observation at this point, but between guests and hosts who spoke different mother tongues - the use of the pointer was higher. Some hosts found it difficult to keep up with their guests in the later stage, especially because the resolution of pointing while building the structure was in millimeters compared to in centimeters when they were pointing to help them pick up blocks. This was caused due to the inaccuracy in pointing the laser when the wearer would move in the very instant the guest would point. "Ok, don't move now, I'll be pointing", was something heard frequently at this stage. 4 of 5 guest-host pairs completed the task successfully and they all mentioned it was possible due to both the video+audio and the pointer. The 5th pair was able to complete the task with a tutorial on how best to use the system.

b. Host helping the Guest

Feedback: Most guests started with asking the host to get the structure close to the camera rather than asking them to describe it. They rather asked specific questions by pointing at them. Some of them asked the host to remove certain blocks so that they can have a better understanding of its construction. The hosts were often heard asking questions like "So how does it look now", "Oh right I can't see your view". This task was completed in almost half the time it took for the same tasks when the guests were helping the host. The frequency of use of the pointer was similar to that in the previous task, however the use of video and audio increased significantly. All the 5 pairs successfully completed their tasks.

5.5.3 Public Perception

Task: One host was wearing the device and I accompanied them to the Kendall Square area. The host was acting as if they were talking to someone, I was observing people around, some people approached curious about the device, while at other times I confronted passers by asking them what they thought of the object the host was wearing. I later explained them what it did and asked them again for their feedback.

a. Initial Reaction (Without knowing what the device did):

Feedback: The device attracted odd stairs, smiles and questions from passers by. One person came forward and asked if it was a pet. Another was curious to know what it did. Yet another asked if it was a "crazy artificial intelligence thingy that talked with the brain". When asked about its looks, everyone liked the aesthetics, they found it friendly, especially the fact that it moved seemed to lend it life.

b. Later Reaction (After knowing what it did):

Feedback: Seeing the laser working, their first reaction was a smile and later caution as it pointed at their bodies. When the motor moved again, some of the people who were talking to us tried to cover their eyes fearing it might hit them. After knowing the *icon* concept and its potential usage, they were excited at the possibilities. Some raised concerns at the lack of physical interactions with other guests. While others mentioned they wouldn't be comfortable having their kids talk to strangers. Some possible scenarios were more appealing than the others, traveling around other cities was a crowd favorite. Some noted that having a green laser would make them feel more comfortable.

5.6 Analysis and Reflections

This section reflects on the evaluations to suggest possible changes to the system described in this thesis to make it more effective.

Concept

Users welcomed the interaction concept of a host and guest. However some users suggested use of a better terminology to avoid confusion between their interchangeable roles as learners and teachers and as wearers and remote users of the system.

App and Platform Design

Users found the design and use of the software very visually pleasing and intuitive. It'd be interesting to see interaction design challenges as the system scales up to thousands of users and their representation on the platform. Users appreciated the fact that it was web-based and hence multi-platform.

Device Design

Users likened the device design to that of a friendly pet sitting on their shoulder. The form factor of the device can be further greatly reduced by using MEMS based laser mirror switches and an embedded chip. Once this is achieved, more subtle device designs can be explored, such as a patch-like or a jewelry-like design. It'd be interesting to see the user reactions when the miniaturization takes away its voluminous pet-like resemblance.

Video and Audio

Users mentioned that a more stable video stream would greatly improve their experience. This could be achieved by a continuous video stitching algorithm and then motion stabilizing it. Elimination of the smartphone is possible by adding a high-power system-on-chip and a faster video encoding stack along with an embedded camera.

Pointing System

Users found the pointing system novel and useful in their interactions. A green laser would reduce the apprehension that comes with a red laser. A face detection algorithm to avoid mis-pointing the laser to eyes would be a great safety addition. A faster reaction time would eliminate the need for the host to be stable for the 2 seconds after the guest points using the app.

Learning

This system was found most effective for learning applications which are more hands-on, situated and action based. It'd be interesting to examine learning with this system on a longer term, how do learners chart their own paths through interactions with multiple remote teachers. This system excels at soft-learning applications and is more suitable for life-long adult learners.

Section 3
DREAMING

Interlude : A haiku in 2050

In 2050,
learning,
not education.

Holding a guitar,
I never played,
it plays itself.

I'm in Tibet,
at immigration,
I speak Tibetan.

I pick up a paper,
thinking of a crane,
it folds.

A bouncing ball,
I stare,
I learn physics.

Reading a book,
I don't understand,
the author explains.

At the bird park,
I spot a bird,
someone whispers.

I never swam,
I can swim,
swim.

Cntrl+C, Cntrl+V,
Cntrl+Z, Cntrl+Alt+Del,
in the physical.

I see students,
in a school,
in the museum.

Why? Answered.
How? Answered.
What? Answer.

Knowledge,
then it was about questions,
now it's about choices.

Chapter 6
Conclusions and Future Work

6.1 Contributions

The novelty and merits of the *icon* system include:

Wearable, eyes-free, hands-free

Being a wearable device worn on the shoulder, *icon* integrate with the natural environment of the learner. The device allows a hands-free and eyes-free interaction focusing the attention of the learner on the task at hand.

Pointing

Unlike traditional telepresence systems which primarily use video and audio, *icon* allows the remote user to manifest the pointing capabilities of their hands in the wearer's environment.

Hosts, Guests

Owing to the two distinct setups of the wearer and the remote user, the system embodies in its experience the idea of a host and a guest. This asymmetry allows the guest to provide the context and the host to provide focus to the session. The system also introduces the concept of multiple hosts and multiple guests to interact leading to interesting use cases and applications.

Soft-learning

While most learning systems are built to support a structured learning approach, *icon* enables seamless communication between learners and teachers in the hope that this unique medium would facilitate exchange of ideas, thoughts, experiences, skills in a more unstructured yet enriching way.

6.2 Future Work

As *icon* addresses a novel area of wearable tools for learning across distances, only a small set of applications and possibilities have been explored in this thesis.

Integrated Device

Implementing a moveable camera as part of the core hardware instead of depending on the smartphone.

Haptic Accessories

Wearable haptic accessories to sense and actuate fingers and hand motions allow a more natural way of learning for certain skills.

Scalable Phone-only Implementation

One way that the central idea of this thesis can be deployed with the current smartphone hardwares is by making an app where the host sees the world through the smartphone (showing a camera feed on the screen) while the guest's pointing gestures are overlaid on the host's screen.

iconHub

At this point the *iconHub* exists as a fully functional deployment, however to scale this platform we need to figure out the incentive aspect of building such an online community.

Bibliography

[Debord58] Theory of the Dérive by Guy Debord http://www.cddc.vt.edu/sionline/si/theory.html June 1958

[Johnson86] Bonnie Johnson, Geraldine Weaver, Margrethe H. Olson, Robert Dunham, and Grady McGonagill. 1986. Using a computer-based tool to support collaboration: a field experiment. In Proceedings of the 1986 ACM conference on Computer-supported cooperative work (CSCW '86). ACM, New York, NY, USA, 343-352. DOI=10.1145/637069.637114 http://doi.acm.org/10.1145/637069.637114

[Ishii90] H. Ishii. 1990. TeamWorkStation: towards a seamless shared workspace. In Proceedings of the 1990 ACM conference on Computer-supported cooperative work (CSCW '90). ACM, New York, NY, USA, 13-26. DOI=10.1145/99332.99337 http://doi.acm.org/10.1145/99332.99337

[Berlin92] Lucy M. Berlin and Robin Jeffries. 1992. Consultants and apprentices: observations about learning and collaborative problem solving. In Proceedings of the 1992 ACM conference on Computer-supported cooperative work (CSCW '92). ACM, New York, NY, USA, 130-137. DOI=10.1145/143457.143471 http://doi.acm.org/10.1145/143457.143471

[Hollan92] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92), Penny Bauersfeld, John Bennett, and Gene Lynch (Eds.). ACM, New York, NY, USA, 119-125. DOI=10.1145/142750.142769 http://doi.acm.org/10.1145/142750.142769

[Resnick92] Paul Resnick. 1992. HyperVoice: a phone-based CSCW platform. In Proceedings of the 1992 ACM conference on Computer-supported cooperative work (CSCW '92). ACM, New York, NY, USA, 218-225. DOI=10.1145/143457.143482 http://doi.acm.org/10.1145/143457.143482

[Kuzuoka94] Hideaki Kuzuoka, Toshio Kosuge, and Masatomo Tanaka. 1994. GestureCam: a video communication system for sympathetic remote collaboration. In Proceedings of the 1994 ACM conference on Computer supported cooperative work (CSCW '94). ACM, New York, NY, USA, 35-43. DOI=10.1145/192844.192866 http://doi.acm.org/10.1145/192844.192866

[Bauer99] Martin Bauer, Gerd Kortuem, and Zary Segall. 1999. "Where Are You Pointing At?" A Study of Remote Collaboration in a Wearable Videoconference System. In Proceedings of the 3rd IEEE International Symposium on Wearable Computers (ISWC '99). IEEE Computer Society, Washington, DC, USA, 151-.

[Tang02] Felix Tang, Chris Aimone, James Fung, Andrej Marjan, and Steve Mann. 2002. Seeing Eye to Eye: A Shared Mediated Reality Using EyeTap Devices and the VideoOrbits Gyroscopic Head Tracker. In Proceedings of the 1st International Symposium on Mixed and Augmented Reality (ISMAR '02). IEEE Computer Society, Washington, DC, USA, 267-.

[Goldberg02] Collaborative Online Teleoperation with Spatial Dynamic Voting and a Human "Tele-Actor" K. Goldberg, D. Song, Y. Khor, D. Pescovitz, A. Levandowski, J. Himmelstein, J. Shih, A. Ho, E. Paulos, J. Donath, IEEE International Conference on Robotics and Automation, May 2002.

[Drugge04] Mikael Drugge 'Wearable Computer Interaction Issues in Mediated Human to Human Communication' PhD thesis submitted to the Department of Computer Science and Electrical Engineering Luleå University of Technology, November 2004

[Gergle04] Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2004. Action as language in a shared visual space. In Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW '04). ACM, New York, NY, USA, 487-496. DOI=10.1145/1031607.1031687 http://doi.acm.org/10.1145/1031607.1031687

[Kurata04] Takeshi Kurata, Nobuchika Sakata, Masakatsu Kourogi, Hideaki Kuzuoka, and Mark Billinghurst. 2004. Remote Collaboration using a Shoulder-Worn Active Camera/Laser. In Proceedings of the Eighth International Symposium on Wearable Computers (ISWC '04). IEEE Computer Society, Washington, DC, USA, 62-69. DOI=10.1109/ISWC.2004.37 http://dx.doi.org/10.1109/ISWC.2004.37

[Morishima12] Morishima, S ;Mashita, T. ; Kiyokawa, K. ; Takemura, H. 'A waist-mounted ProCam system for remote collaboration'. in Proceedings of IEEE International Symposium on Mixed and Augmented Reality 2012 Science and Technology Proceedings
5 - 8 November 2012, Atlanta, Georgia. DOI=10.1109/ISMAR.2012.6402584

[Kashiwabara12] Tadakazu Kashiwabara, Hirotaka Osawa, Kazuhiko Shinozawa, and Michita Imai. 2012. TEROOS: a wearable avatar to enhance joint activities. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, NY, USA, 2001-2004. DOI=10.1145/2207676.2208345 http://doi.acm.org/10.1145/2207676.2208345

[Sodhi13] Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Giuliano Maciocci. 2013. BeThere: 3D mobile collaboration with spatial input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 179-188. DOI=10.1145/2470654.2470679 http://doi.acm.org/10.1145/2470654.2470679

[Rekimoto14] Shunichi Kasahara and Jun Rekimoto. 2014. JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation. In Proceedings of the 5th Augmented Human International Conference (AH '14). ACM, New York, NY, USA, , Article 46 , 8 pages. DOI=10.1145/2582051.2582097 http://doi.acm.org/10.1145/2582051.2582097

[Maes10] Natan Linder and Pattie Maes. 2010. LuminAR: portable robotic augmented reality interface design and prototype. In Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology (UIST '10). ACM, New York, NY, USA, 395-396. DOI=10.1145/1866218.1866237 http://doi.acm.org/10.1145/1866218.1866237

[Mann2000] Steve Mann. 2000. Telepointer: Hands-Free Completely Self Contained Wearable Visual Augmented Reality without Headwear and without any Infrastructural Reliance. In Proceedings of the 4th IEEE International Symposium on Wearable Computers (ISWC '00). IEEE Computer Society, Washington, DC, USA, 177-.

[Lave91] Jean Lave and Etienne Wenger. 1991. Situated Learning : Legitimate peripherial participation, Cambridge: University of Cambridge Press.

[Kimble08] Chris Kimble and Paul Hildreth. 2008. Communities of Practice : Creating Learning Environments for Educators. Information Age Publishing. ISBN: 1-59311-865-3

[Halverson09] Halverson A. C. 2009. Rethinking education in the age if technology. New York: Teachers College Press.

Skype http://skype.com

Polycom http://polycom.com

Cisco http://cisco.com

Dropcam http://dropcam.com

Google Helpouts https://helpouts.google.com/home

Oculus Rift http://oculusvr.com

Instructables http://instructables.com

DIY http://diy.org

wikiHow http://wikihow.com

Khan Academy http://khanacademy.org

EdX http://edx.org

MIT OCW http://ocw.mit.edu

UnHangout http://unhangout.media.mit.edu

ChatRoulette http://chatroulette.com

Colophon

This document was authored in iA Writer and designed in Pages 5.2 using the following typefaces - Avenir Next by Adrian Frutiger and Akira Kobayashi, FF Din by Albert-Jan Pool. All photographs were captured by the author on a Canon 5D Mark II.

August 2014, Cambridge