

# Augmented Reality Visor Concept



Björn Svensson & Mattias Wozniak



Master Thesis

Department of design sciences

Lund university

ISRN: LUTMDN/TMAT-5037-SE

EAT 2011



Concept video available at:  
<http://www.youtube.com/watch?v=uWWD5oOY4sQ>

# Abstract

Imagine what can be done with an AR (Augmented Reality) visor and the proper technology that supports tracking of the human body and identification of the user's surroundings, an AR system that augments our mundane duties and aids us in understanding the world in a better way. The technology requires new, intuitive and innovative interaction methods set apart from traditional human – computer interaction. This master thesis will inject new ideas into computer interaction design and guide developers of AR systems to develop interaction techniques that provide good user experience in these types of systems.

Today's AR systems focuses a lot on how to present information to the user in a visually impressive way, which only provides good user experience to some extent. The interaction is often basic and consists of traditional input methods that aren't adapted to AR.

By using focus groups, expert reviews and surveys this master thesis concludes that social acceptance is the single most important factor to take into consideration. Among other topics discussed is the hands busy aspect, collaborative work, intuitiveness and consistency.

## Keywords

Augmented Reality, head mounted display, social acceptance, interaction design, information visualization

# Sammanfattning

Föreställ dig vad som kan göras med ett par AR-glasögon (Augmented Reality) och den rätta tekniken, tekniken som stöder tracking av den mänskliga kroppen och kan identifiera användarens omgivning. Ett AR-system som hjälper oss i våra vardagliga sysslor och underlättar vår tolkning och vårt förstående av omvärlden. Teknologin kräver nya, intuitiva och innovativa interaktionssätt som skiljer sig från traditionell människa – dator-interaktion. Detta paper tillför datorinteraktionsområdet nya idéer och tanken är att arbetet ska kunna vägleda utvecklare av nya AR-system att ta fram nya interaktionstekniker som ger en bra användarupplevelse.

Dagens ARsystem fokuserar mest på hur man på bästa sätt imponerar på användaren genom snygga visuella presentationer. Detta bidrar visserligen till en förbättrad användarupplevelse men det räcker inte hela vägen. Interaktionen i dessa system är ofta grundläggande och består av traditionell input som är dåligt anpassad för AR. Genom fokusgrupper, expertutvärderingar och användarundersökningar har resultatet att social acceptans är av största betydelse uppnåtts. Andra rubriker som diskuteras är handfri användning, kollaborativt arbete, intuitivitet och enhetlighet

# Acknowledgements

We would like to thank our supervisor Klas Hermodsson at Sony Ericsson for the constant support with invaluable knowledge and ideas throughout the project and for giving us the opportunity to conduct the work at Sony Ericsson. We would also like to thank Mattias Wallergård, our supervisor at Department of Design Sciences, Lund University, for his much appreciated guidance and help during the project. Thank you to Ola Thörn at Sony Ericsson for his passionate interest in innovative techniques from which a lot of input has been gathered and used, especially in the design process. A general thank you to all of those who have followed our work and to those who have contributed to our work in any kind of way. Last but not least a big thank you to Paulina Nossborn for her invaluable contribution to the concept film.

*Mattias Wozniak & Björn Svensson*

Lund, December 2010

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Scenario	1
1.2	About AR	1
1.3	Why AR Visor	3
1.4	Applications/Usage Areas	4
1.5	Goal	4
1.6	Limiting factors	5
<b>2</b>	<b>Theoretical Background</b>	<b>7</b>
2.1	Related work	7
2.2	User experience	8
2.3	Context awareness	8
2.4	Tangible interaction	9
2.5	Gaze tracking	10
2.6	Gestures	11
2.7	Social acceptance	12
2.8	Output	13
<b>3</b>	<b>Method</b>	<b>15</b>
3.1	LUCID	15
3.2	Social Acceptance Survey	19
<b>4</b>	<b>Results and discussion</b>	<b>21</b>
4.1	Requirements/paradigm	21
4.2	Results of Social Acceptance Survey	23
4.3	Concept Video	26
4.4	Services	27
4.5	General System Interaction	28
4.6	Shortcuts	32
4.7	Notifications	33
4.8	Wayfinding	34
4.9	Mirroring	35
<b>5</b>	<b>Conclusions</b>	<b>37</b>
5.1	Design choices	37
5.2	Obstacles Along the Way	39
5.3	Future work	39
	<b>References</b>	<b>41</b>
	<b>Appendix A, Personas</b>	<b>45</b>
	<b>Appendix B, User survey</b>	<b>46</b>
	<b>Appendix C, Storyboards</b>	<b>59</b>

# 1 Introduction

## 1.1 Scenario

Imagine the following scenario:

*“You’re wandering around downtown on your own. You’ve got an appointment in 45 minutes and want to beguile the time and decide to sit down at a cafe. The system highlights nearby cafes in the real world and indicates a friend sitting down at a cafe only a block away from you, whom you decide to join. When arriving at the cafe you get a personal recommendation of coffee based on your preferences. To clear your mind before the appointment you come to think of a funny blog post a mutual friend posted the other day, which you want to show your friend. You show your friend the blog post without using extra peripherals, the coffee table is your display which is seen only through your and your friend’s visor. To keep you from missing your appointment, you’re shown a reminder 13 minutes before the appointment, 13 minutes based on the distance to the appointment location. You leave the cafe as your friend wishes you good luck. Outside, the system guides you to the location of the appointment via a guiding line on the sidewalk.”*

A way of achieving this is by using augmented reality (AR) seen through a pair of glasses.

## 1.2 About AR

Azuma’s definition of what an AR system should be capable of [1]:

- Combining real and virtual
- Interaction in real time
- Registering in 3-D

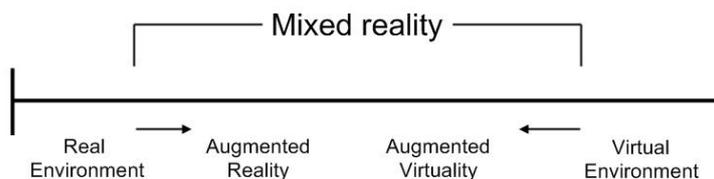


Figure 1 Milgrams Continuum [37]

Milgram's Continuum defines the differences between real and virtual environments. Virtual Environments (VE) immerses a user inside a virtual world. In opposition to VE, AR still resides in the real world providing overlaid virtual information. To summarize you could say that users in a VE are a part of the computers world while AR aims at making the computers being a part of our world.

The first Augmented Reality (AR) interface was envisioned as early as 1965 by Sutherland [17]. Only three years later, in 1968, he developed what is seen as the first head-mounted display (HMD). Its display was optical see-through working in stereo, generating virtual images onto the real world and using tracking [16]. These early ideas in all glory but it's not until the two latest decades companies and industries have drawn more attention to the subject. Even with this in mind still little work has been done in the areas of interaction design and visualization [18].

Weiser wrote [19]:

*"There is more information available at our fingertips during a walk in the woods than in any computer system, yet people find a walk among trees relaxing and computers frustrating. Machines that fit the human environment instead of forcing humans to enter theirs will make using a computer as refreshing as taking a walk in the woods."*

As AR aims at moving the computer world into the physical world, this changes the boundaries of available interaction space from traditionally being very limited to an almost infinite amount of interaction space. The ideal vision of AR is almost an utopia. Today's AR experience is rather limited due to the formats that serve as a basis for the system. The most common and popular platform right now are cell phones due to reasonable pricing of the hardware in relation to the perceived experience as well as mobility. Various, mostly homebrew, applications can be downloaded from the stores of each respective operating manufacturer (Android, iOS). The standard way of accomplishing AR in cell phones is to use the phone's camera to view the real world through the display and overlay additional information on top of this. The overlaid data is gathered using the built in sensors (gyro, magnetometer) to measure head position and orientation while the geographical location is tracked with GPS in combination with data gathered from an online hub.

When building visual AR systems technologies often referred to are *tracking*, *registration* and *display* [11]. Tracking refers to aligning virtual objects at a location in the real world, making them look like actual physical objects that are detached to the real world. Registration means to attach

virtual objects to the real world, making it look realistic. Display refers to the physical display used to view the AR information.

An important note is that usually when speaking of AR it's often referred to the visual representation of the world, which facilitates the visual sense. Although AR is not only limited to this, as its goal is to "improve reality" an AR system could also be extended with the help of for example sound, providing 3d sound. It can be used on all our senses. There are examples of implementations enhancing our smell and taste [20].

Most examples of AR glasses available today are more like entertainment glasses where the user can browse his music, watch movies etc. They usually don't provide much interaction. The interaction is instead performed on the device connected to the glasses (e.g. phone, mp3-player, computer etc.).

## 1.3 Why AR Visor

AR in its current form is very limited when it comes to user experience compared to what it actually could provide. The mobile AR available to customers today is mostly based around the cell phone to augment the world. Hermodsson refers to this as "browsing the world through a keyhole" as the user's cell phone is the tiny window in which the augmented world can be experienced [21]. The interaction also has to be performed on the device which leads to both non-dynamic and intrusive interaction, the cell phone has to be at hand each time the user wants to enable AR. The common way of experiencing AR today is to hold the cell phone in front of you with your arms extended. This is both unergonomic and the user is constantly reminded of the social acceptance factor when using it in public. The experience isn't very immersive and neither are the interaction techniques, which often consist of pressing buttons on the phone. Due to the nature of the format, it also lacks functionality that would become available when using a more integrated and immersive system.

By using a visor (HMD, glasses), or even contact lenses [22], more or less full peripheral view can be offered and the full potential of AR could be utilized in a more dynamic way. Instead of having to pick up your cell phone from your pocket on demand the system would always be available as long as the visor is carried (which in the ideal case is just about all the time). Some effort would have to be put in to developing the interaction system for this kind of system, but once done the hands busy problem can maybe be completely eliminated. Complete hands-free interaction has its limitations today (mainly due to technology and lack of user experience with these types of systems) and the suggested system in this report will instead be focusing on how to develop a mixed system with somewhat traditional

interaction techniques but supporting the hands busy aspect far better than systems available today. The usage of AR could be better integrated into everyday tasks in a more natural way.

## 1.4 Applications/Usage Areas

The usage areas of AR are many. With the ongoing development there are already applications used in a variety of industries [1].

In the medical sector AR can be used as an aid during surgery or training. Some car manufactures are doing research on using AR with an HMD as an aid during repairs. The system identifies recognizable objects and indicates what to do at what point during the repair.

While the above areas are more or less exploited in the private sector the biggest and most growing area of use of AR is the one focusing on the public market. Applications that act like tourist guides, for example Layar<sup>1</sup> and Wikitude<sup>2</sup>, showing points of interests and further information about given locations are publicly and easily available to anyone. As for the entertainment part there are applications based on finding virtual treasures, so called geocaching.

A couple of marker based applications like AR-postcards, LEGO (displaying assembled product) and furniture solutions by IKEA where the user can furnish his/her apartment.

## 1.5 Goal

The overall purpose of this master thesis is to *investigate the possibilities of AR systems of the future*. By using a full-peripheral view visor the user experience is heightened since the user is fully engulfed in the system in a way that today's technology can't achieve. More closely, the thesis aims at providing:

- Functional, intuitive and cognitively sound interaction methods.
- Information presentation that facilitates everyday duties.
- Socially acceptable interaction.

---

<sup>1</sup> Layar: <http://www.layar.com/>

<sup>2</sup> Wikitude: <http://www.wikitude.org/en>

The above deliverables combined provides an intuitive, ergonomic and easy to use system which changes the way people look upon man-machine interaction that seamlessly assist the user in everyday tasks.

This idea raises some important questions:

- What would the information look like?
- How would the user interact with it?
- What would the information consist of?

There are a set of examples of other AR visor concept systems that have been produced. All of them have been focusing on either the technology behind them or making a cool presentation to the user. None has truly gained an impact or reached critical acclaim and none have focused on usability and user experience.

## 1.6 Limiting factors

Hardware matters aren't taken much into consideration in this paper since the ideal technology is still a couple of years away and much of the ideas presented could also be applied to more traditional types of systems.

Today there are no visors that provide full peripheral view (180 degrees) which would offer "unlimited" visualization space. Applicable hardware on the market today are optical see-through HMDs and video see-through HMDs. Optical see-through displays works through partially transmissive and reflective combiners. This lets the user look through the display and see the real world mixed with virtual overlaid objects, like a pair of traditional glasses with display technology inside. A problem with these displays is the amount of light that emits through the glasses. It requires good spatial tracking precision in order to align virtual and real information [23]. Studies have also shown that there is a risk of symptoms varying from eye strain to headache [23]. Another issue is that in order to take advantage of object recognition the user is in need of an additional camera.

Thomas et. al. have concluded that commonly used AR glasses in outdoor environments perform badly, text and icons gets washed out and the usability is strongly affected [36].

Video see-through displays combine the display with one or two cameras. The cameras provide the view of the real world while a scene generator adds graphical images, blending the real world with the virtual. This guarantees spatial alignment of virtual and real information [23]. The big drawback with these types of systems is that it's far from as responsive as optical see-through glasses which quickly ruin

the immersion for the user as well as making eye strain very apparent.

The idea with optical see-through HMDs which also makes it the best way to experience AR is that the user is still looking at the real world but with overlaid data. This stands in contrast to video see-through HMDs where the user sees a digitally represented image of the real world, which imposes great strain on the technology to accommodate the way the user expects to see the real world. If for example the user were to turn his head and the image seen in the HMD was displayed with even a few milliseconds delay the user would start to develop motion sickness. Motion sickness and eye strain is the most common issues with video see-through HMDs [38].

To utilize HMDs in AR systems the display would need to identify what the user sees. This is quite simple in a video see-through system where at least one camera already exists and can monitor and recognize what the user sees. In the case of the optical see-through system an additional camera would need to be implemented to monitor this.

Because of the technological limits this master thesis work will result in a proof-of-concept rather than a final product. The ideas presented here will be prototyped in a virtual environment as well as presented in a concept video showing what our ideas would look like if the technology supported it.

Ideally the Field-Of-View (FOV) limitations, along with the above mentioned limitations, will soon be solved by the industry providing an unobtrusive user interaction with virtual objects superimposed anywhere in the real environment. This project is therefore carried out with the belief that adequate equipment will be available to end users at a sufficient cost in the near future.

# 2 Theoretical Background

The purpose of this chapter is to give a brief overview of some of the areas involved in AR. The overview is included to give the reader a better understanding of the different aspects that come into play when discussing interaction in AR systems.

## 2.1 Related work

A lot of studies have been performed relevant to this work since it covers so many areas. Petersen et. al discusses the use of continuous natural interfaces, trying to reduce the gap between real and digital world [7]. They state it should be easy for content to switch domains from a virtual instance in AR to a physical instance in the real world, since switching between devices interrupts work flow. Their main contribution is the methodology for an intuitive interface.

Further studies on how to integrate virtual reality into the real world is done by Holman et. al. A prototype windowing environment is presented that simulates the use of digital paper displays, since, according to the paper; monitor input is indirect and heavily dependent on visual cues [8]. The user should notice no difference when transitioning between domains. A set of interaction methods are also presented to transfer display from screen to paper and back, flip, point, collocate, collate etc.

Gestures are a common way to interact with AR systems. Gestures can be intuitive and doesn't require any input device. Nielsen et. al presents a procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction. Using this procedure assures that the gestures are intuitive, tailored for the purpose of the application and user group, that they are culturally independent, not too many, ergonomically efficient, easy to learn and easy to remember [9].

Benko et. al presents a set of cross-dimensional interaction techniques for use in a hybrid user interface that integrates existing 2D and 3D visualization and interaction devices. What gestures to use when translating 2D virtual objects into 3D and back using pull and push is discussed, as well as gestures to connect/disconnect to get 2D representation of 3D object and pin/unpin actions [10].

Another issue in augmented reality is interaction in 3D space. Interaction with virtual representations of objects that are out of your physical reach becomes a problem when you can't actually reach it. Bowman et. al presents a couple of

different techniques attempting to solve this issue. Selection techniques include simple ray-casting technique, two-handed pointing, flashlight technique, aperture techniques, image-plane technique and fishing-reel technique. Solutions for direct manipulation are also provided in form of *Simple Virtual Hand* and *Go-go Interaction Technique* [11].

Looser et. al presents a solution with a lense as an interaction tool in AR environments. Their idea offers users a natural, tangible interface for zooming in and out offering semantic information filtering in the real world providing a partial solution for the distance problem [13].

## 2.2 User experience

When designing products for consumers the aspect of consumer satisfaction is unavoidable. So what's the reason for buyers becoming permanent users of a product? User experience was first stipulated by Donald Norman. Today it's become so common and widespread that it's defined in ISO 9241-210 as "... a person's perceptions and responses that result from the use or anticipated use of a product, system or service" [35]. User experience is a broad term. It offers insight into a lot of different domains. When talking about user experience one does often encounter the questions of usefulness, user friendliness and usability [34].

N. Bevan formulates the following pragmatic goals for user experience [34]:

- Acceptable perceived experience of use (pragmatic aspects including efficiency).
- Acceptable perceived results of use (including effectiveness).
- Acceptable perceived consequences of use (including safety).

## 2.3 Context awareness

A majority of the systems that resides today are "static" and not automatically adapting. They require user input to do an operation like starting an application of the user's choice. This is time consuming and may also interfere with the user's situation and therefore be interpreted as intrusive. Imagine a system where the context is situation based, and even situation aware. Different kinds of information are shown with different contexts and situations.

There exist several definitions of context. What's the definition that is best suited for our case? Dey et. al defines

context “as any information that can be used to characterize the situation of an entity” [14]. An entity is an object or place that is considered relevant to the interaction between a user and an application. They further discuss and adapt this to form the definition of a context-aware system which can be seen as a system that provides relevant information and/or services to the user. Relevant information depends on the situation the user is in. By combining this together with for example a camera or a gyro the system can make smart decisions differentiating between for an example a driver and a passenger in a car. This could be done using the camera trying to identify and locate the steering wheel. If the steering wheel is positioned in front of the person, this person is considered to be the driver. The driver can then be shown appropriate and adequate information for the given situation while the passenger may have access to a different kind of information. By applying the context in this way the system makes the driver less liable to information intake while the passenger is not as limited in his receptivity.

Context awareness in a traditional manner focused on *what* the user does. Today one often speaks of *who*, *when* and *where* as added factors to *what* when talking about context awareness. The answers to these questions determine *why* a situation is occurring [14] [15]. By taking advantage of this a context aware system can enrich and improve the user experience. The main challenge with such systems is how to make it natural for the user at all times and the question of how well the user responds to the self-adapting information.

M. Weiser stated the following principles about UbiComp [19]:

- The purpose of a computer is to help you do something else.
- The best computer is a quite, invisible servant.
- The computer should extend your unconscious.
- Technology should create calm.

Parts of these principles is what context awareness tries to envision.

## 2.4 Tangible interaction

Human Computer Interaction (HCI) has more or less always been equal to Graphical User Interfaces (GUI). Most of these GUI:s have then be based on interaction by Windows, Icons, Menus and Pointers (WIMP). Interaction with this kind of GUI always relies on accessories as keyboards, mice and a computer screen. Through the latest

technical progresses this approach has been expanded and evolved into interaction with smaller computational devices as cellphones and tablets. The accessories are not of as much of importance than just a decade ago. This is a clear indication of where modern HCI is evolving. Tangible User Interaction (TUI) was defined by Ishii et. al as an extent to Weisers vision of ubiquitous computing [25]. The idea is to “*augment the real physical world by coupling digital information to everyday physical objects and environments.*” Computational devices are more and more becoming a part of our lives and the old paradigm of us being in the computers world is evolving, computers are instead becoming a part of our world. This opens up a variety of new, exciting and dynamic ways to interact with digital information. One could argue that the uncertainty with such an approach is of a too big concern but the main idea with tangible interaction is to build upon our view and familiarity with real world objects and our physical skills, providing a natural and intuitive way of interacting [15].

Norman expanded the term *affordance* and defined it in the world of HCI as “*perceived action possibilities*” [26]. Affordance in this context refers to what the system offers in its visual appearance combined with the actor’s goal, past experiences, plans, values and beliefs.

By exploiting our natural familiarity with physical objects the same user would not have to think “how do I open this file?” when the file in an AR system could be a book that slides open compared to a file hidden in several directories on a desktop.

## 2.5 Gaze tracking

The idea of implementing gaze tracking in a system like this is mainly due to its “mild” interaction compared to gesture based tracking, and as stated before social acceptance is of great importance. As with all things, pros come with cons. Barakonyi et. al concludes that even though gaze tracking has been proved to be faster than manual selection techniques users usually prefer the latter due to several addressed problems with gaze [27].

**Jittery eye motion:** The human eye performs saccades at both controlled and uncontrolled times.

**Tracking inaccuracy:** The tools and equipment used today for gaze tracking often requires a calibration profile for each user. This does of course lead to tracking inaccuracy if the desired accuracy is to be at a couple of pixels. We assume that this is not as large of a problem within a couple of years when a new generation of AR products are introduced.

**Midas touch** is the term referred to when talking about the

difficulty of selecting objects. If a selectable object would be activated/selected by a look at it, this inevitably results in a poor user interface.

**Staring/dwell times:** Dwell times were introduced as a proposed solution to the Midas touch problem. By using a dwell time, a certain amount of “staring”, objects would not be selected if not desired. Different dwell times have been experimented with and it’s difficult to find a threshold which suits different contexts.

As the eye is an indicator of the current point of interest, using gaze tracking as assistance to decide what the user is looking at could simplify the process of analyzing where the user’s looking for the moment.

## 2.6 Gestures

Gestures and body language are natural parts of our everyday lives. They are used to reinforce, clarify and superimpose the meaning of our actions and reactions. This naturally evolves into a need for clear understanding of social acceptance. A gesture that represents some kind of action or reaction in one culture may imply something completely different in other cultures. Gesture interfaces have lately increased in popularity, just as AR. The Nintendo Wii has had great sales figures and the launch of Microsoft Kinect<sup>3</sup> and Playstation Move<sup>4</sup> indicates where the game markets current point of interest is. The main idea with these consoles and accessories was to develop a game interface which is more intuitive instead of using a gamepad as indirect manipulation. They exploit our sense of familiarity with the physical world. A tennis ball in the game is hit with a swing, just like in the real world. Nielsen conducted an approach of how to plan and design new gestures for HCI. His steps were all based on the usability engineering’s five principles [28]:

- Learnability. The time and effort required to reach a specific level of use performance.
- Efficiency. Steady-state performance of expert users.
- Memorability. Ease of system intermittently for casual users.
- Errors. Error rate for minor and catastrophic errors.

---

<sup>3</sup> <http://www.xbox.com/en-US/kinect/>

<sup>4</sup> <http://us.playstation.com/ps3/playstation-move/>

- Coverage. The amount of operators discovered vs. the total operators.

With these in mind he stated the following when designing a set of gestures to fit the corresponding action [9]:

1. Easy to perform and remember
2. Intuitive
3. Metaphorically and iconically logical towards functionality
4. Ergonomic; not physically stressing when used often

Something that ties in with “Metaphorically and iconically logical towards functionality” is that a gesture can only have one function across systems. You can’t tie a gesture to a function that ties to another function in another system that users already are very familiar to.

*“A gesture interface is not universally the best interface for any application. The objective is “to develop a more efficient interface” to a given application” [9].*

Charlotte Magnusson mentions several good things about gestures in her lecture about “Touch and gesture controlled interfaces” [29]. If designed well and with care, gestures are usually; simple, intuitive, they can be shared in a social context, our previous personal experience is huge since we practice them in real life at all times, complex actions can be simplified and relative gestures are easy to use. When not designed with care gestures tend to be non-intuitive, a sense of vacuity can be sensed because of lack of knowledge, they can be embarrassing and absolute gestures are in contrast to relative gestures difficult to use.

## 2.7 Social acceptance

As mentioned in section 2.6 gestures are used as a part of our everyday language. The introduction of new and innovative interaction techniques becomes more apparent every day. By manipulating the information indirectly (keyboard, mouse) the interaction methods have evolved into direct manipulative utilizing interaction with no “middle hands”, e.g. using your hands (old consoles vs. Kinect e.g.). This brings a need for understanding the gestures meaning. Montero et. al. conducts that “...an important factor in determining social acceptance of gesture-based interaction techniques the user’s perception of others ability to interpret the potential effect of manipulation” [30]. To understand

social acceptance there is a need of a clear definition of the concept. Montero et. al points at the common misinterpretation of the difference between social acceptance and user acceptance. They differentiate user acceptance in two parts: User's social acceptance and Spectator's social acceptance.

***“User’s social acceptance:** for every task a user performs, they will be left with an impression – did they feel comfortable or uncomfortable, awkward or natural, relaxed or embarrassed? This will lead to an overall positive or negative impression of the task or technology.*

***Spectator’s social acceptance:** user actions are performed in a range of public and private situations, i.e. contexts. The spectator’s social acceptance is a measure of their impressions of these actions. Does the audience understand what the user is doing? Do they think the action is ‘weird’ or ‘normal’? The spectator quickly builds a positive or negative impression of the user’s actions.”*

Montero et. al further discusses manipulation vs. effect. As mentioned above, a performed gesture is both interpreted by the user himself and the surrounding spectators. If a gesture is obtrusive and loud a negative impression will form [30].

With this definition a gesture can seem acceptable if both the user performing it and nearby spectators consider the gesture appropriate in the given context. Even with a clear definition as clarification this area is still very difficult to work with since the acceptance of a gesture also depends on appearance, social factors and culture [12]. Gesture based interfaces are, as earlier stated, a new developing paradigm. As Rico et. al further states even though the entry of screen-based gesture interfaces (tablets, touch-screen cell phones) has become widely accepted not much work has been concluded in the area of screen-free gesture interfaces [12].

Social acceptance can of course be seen from another point of view. The need for unobtrusive and discrete accessories which preferably do not distinguish itself too much from regular clothing and accessories.

## 2.8 Output

Output is used to convey the result of an action performed by the user:

- Visual output
- Auditive output
- Haptic output

A visual device in a regular desktop environment can be seen as the computer screen. In comparison our visual device would be the glasses. A computational device, either integrated with the glasses or some external device would be used to calculate and position the augmented information while the glasses would be used to actually show the result.

Auditive feedback could be used to ease the cognitively visual load as long as the audio itself is kept on a non-intrusive level. One kind of auditive feedback is 3D sound which can be used in for example directional wayfinding.

Haptic devices on the other hand provide a sense of touch like vibrations. This adds an extra level of realism to the system. By using the approach of TUI the physical environment is used as input which also means that the physical objects also work as feedback givers.

All of the above can of course be combined into a multimodal system. By having a button changing color when pressed together with a vibration that indicates that it's pressed as well as a press sound gives the user valuable feedback. All of these together mimic the reality quite well, which is what we're trying to achieve. A multimodal system exploits more of our senses and our general impression of the system is clearly affected. When designing user interfaces (UI) it is important to have the human body's perceptual and cognitive limitations in mind, which means that there are many important aspects that have to be taken into consideration. Aspects ought to be taken into consideration are information overload, change blindness, perceptual tunneling and cognitive capture which are phenomenon that have to be taken under consideration during development of HCI [6].

Information overload refers to the state when too much information is presented to the user. The user does not manage to process all of the information at once and the user attention may be addressed wrong. It is for this reason important to plan and consider the amount of output information streamed to the user at one time. Show only what is necessary in an onerous context. Change blindness is another cognitive matter. State changes in the interface can be missed and so called "change blindness" occurs. This generally happens when the user is putting his or her focus at one space and is therefore distracted and not open and available to actions happening elsewhere in the visual field.

# 3 Method

## 3.1 LUCID

The development method chosen for this project was decided to LUCID - Logical User Centered Interaction Design [3]. The approach has been altered somewhat to fit this particular project, since it's not of traditional interaction and in order to fit the workflow of all involved participants but the basis remains the same. Some iteration of the different steps has also been introduced. The definition of LUCID according to the Cognetics Corporation [3]:

1. **Envision**  
Develop UI Roadmap which defines the product concept, rationale, constraints and design objectives.
2. **Analyze**  
Analyze the user needs and develop requirements.
3. **Design**  
Create a design concept and implement a key screen prototype.
4. **Refine**  
Test the prototype for design problems and iteratively refine and expand the design.
5. **Implement**  
Support implementation of the product making late stage design changes where required. Develop user support components.
6. **(Support)**  
Provide roll-out support as the product is deployed and gather data for next version.

### 3.1.1 Envision

During the start-up phase of the project product concept, rationale, constraints and design objectives were all discussed and defined. This phase was revisited a couple of times later on during the project to update mostly the constraints to fit the user needs and what we wanted to do with the system. The two main ideas and constraints were that no interaction should be performed using cognitively straining movements or socially unacceptable interaction.

Many before us have designed AR systems but not many have implemented or focused on good interaction. Most of it is based on touching virtual objects in the air which provides no feedback, isn't really socially acceptable and is quite strenuous to perform in the long run.

### **3.1.2 Analyze**

To come up with ideas for what type of system the user needs, a focus group was put together. The idea of this was to gather information through brainstorming of what the perception of a system like this had to be able to do. Asking yourself the questions "Who? When? What and How?" resulted in a great mix of ideas which were all written down using card sorting method, later to be expanded upon.

To get a better overview of the system features the ideas were sorted into categories according to their nature. This forced a revisit of the constraints previously defined to update them to fit the updated user needs.

The constraints were further developed through whiteboard sessions where the cards from the card sorting served as a basis for each different case. During these sessions some base interaction techniques were discussed to get an idea of how each task would be carried out. With these at hand, some general interaction requirements started taking form.

In order to decide on a target audience personas were developed [see appendix A]. A typical family with two kids was pictured to convey the possibilities that most people will be able to use the system. By focusing on a regular family, different personalities, ages and personal preferences are covered. Developing personas also eliminates the risk of focusing too much on your personal needs as a developer.

The personas went hand in hand with the scenarios later developed, which followed the family during a typical weekday. The scenario shows how people would use the system to enhance the mundane chores that is done every day.

### **3.1.3 Design**

The next step was to define and create a design concept of the interaction. This was done using the scenarios to create story boards [see appendix C]. Story boards are a great way to introduce people to something they are unfamiliar to as well as keeping the developers up to date with requirements and what functionality to include. It provides examples of situations as well as illustrations which ensures that the message and the vision is conveyed.

The basis of an augmented reality system like the one aimed to develop revolves around mobility, which means that it's important that the system is usable everywhere and

in almost every situation. This leads to three other important aspects, social acceptance, ergonomics and safety. Since the introduction of hands-free phone calls it has slowly become more socially accepted to “talk to yourself”, which points towards extended exposure to something makes it more acceptable. Still, social acceptance has to be taken into consideration when designing a system that should function in all places.

The typical demonstration of augmented reality consists of the famous computer interaction that Tom Cruise performs in the motion picture *Minority Report*. He navigates the computer system by using his arms stretched out in front of him to perform various gestures which are recognized by a system that tracks his hands. This way of interacting, while looking cool, is neither socially acceptable nor very ergonomic. Better ways to navigate computer systems are crucial since the example used tends to get very tiresome in the long run and if the interaction is too cumbersome, the user won't bother.

The traditional way of interacting with computer systems is to sit down by a table with the system on the surface in front of the user. Since the development of mobile phones computer usage has evolved to be used while on the move. Still today, mobile phone usage isn't well adjusted to on the move usage. It is possible because of the light weight and small size of the unit, but it isn't recommended since it often risks the safety of the user.

For the system to be mobile the user needs to have his interaction device and surface within reach at all times. As Shneiderman's first golden rule for interaction design suggests, consistency is important [4]. The interaction has to be based around the same principles at all times. Too many different ways of interacting with a system produces unnecessary amounts of load on our working memory and the user has a difficult time remembering the interaction. Wallergård's guidelines for designing new interaction techniques [5]:

- Use natural mapping
- Use interaction analysis
- Study efficiency
- Have physical ergonomics in mind

Another important aspect is what Shneiderman calls *Universal Usability* [4], to utilize the knowledge the user already has. This makes the interaction feel more intuitive to the user which in turn provides a better user experience. It is important to not use interaction that conflicts with interaction in other systems.

With the above mentioned aspects in mind, the base interaction in the system was developed. The fundamental idea of the interaction is based around using the user's hands as *familiar fixed points*. Familiar fixed points provide the user comfort in knowing how to interact and how the system will respond when they do. It can also provide great mobility if the fixed points are chosen accordingly, which makes them ideal for developing new innovative interaction techniques. The hands are familiar in that we know how they function and how they respond, and fixed in that we always know where they are and that they are within reach. We are aware of their abilities and limitations and we are comfortable and used to interacting with them. The hands in this thesis functions as a starting point for the entire system. They function as the main interaction device as well as a potential display surface. Everything starts with your hands.

When developing the story boards some ideas of gestures for interaction sprang to mind. This is also largely how they were developed, even though refinements have been made along the road.

#### 3.1.4 Refine

Early refinement work consisted of several reviews where experts on different areas got to look at the system and gave comments according to their area of expertise. Early prototyping was performed using a pico projector, displaying images "behind the curtain" when the user performed different actions. The size of the projector allowed for on the move use and use in different situations, called body storming. Body storming allows testers to live through different potential use cases in real time. The difference of actually being in the situation and just pretending of being in it allows for a better experience and usually many thoughtful ideas. The purpose of the body storming was to discover if the interaction was suitable in areas and situations where it's plausible that it will be used. This way of prototyping, while revealing weaknesses, is fully rigged and doesn't quite allow everything to be tried out.



Figure 2 Pico projector prototyping

The expert reviews was iterated throughout the project during all phases and assured quality on all areas. They also served as an input on the general system, meaning that the experts have been giving input on all areas, not just their area of expertise. This provided varied input which served as a kind of user test.

## 3.2 Social Acceptance Survey

Another important aspect that was kept in mind was social acceptance. Inspired by Ronkainen et. al. [24] a market survey was performed to test this out in a quantitative manner, rather than a qualitative. A quantitative approach was chosen to get the best overview on the general public's opinion, which for this project is more important than the opinion of a select few. The survey was sent per e-mail to the participants who were selected with variety and relevance in mind. In order to get accurate answers the user needs to be somewhat familiar to modern technology and have an understanding of newer interaction interfaces. Aware of this, the survey was sent to people from the cognitive science department and engineering students at Lund University as well as to random contacts.

The purpose apart from social acceptance was to test the intuitiveness of the gestures. A couple of alternatives for each action was presented to the user through video clips showing what the gesture would look like, without showing the result of the action. The gestures included in the survey were alternatives brought up earlier during the project, both good and bad, with some already in use in systems today. The variety of gestures were selected to mirror that a well used gesture in today's systems doesn't have to implicate a good gesture. Today's gestures are often limited because of the technology used to detect them.

The purpose of not showing the result was to avoid revealing too much to the user which could easily lead the user away from the important. The social acceptance factor is also more pronounced when the user isn't aware of the full purpose of the gesture.

The survey takers first answered questions about his/her sex, age and line of work. The participants then answered a set of questions corresponding to each gesture shown in the video clips. More exactly, each gesture was followed by three questions:

- How comfortable would you feel performing this gesture in front of: (Not at all - 1/Very much - 5)
  - Strangers
  - Colleagues
  - Friends

- Family
  - Alone
- I don't think it would be strange seeing the following persons performing this gesture: (Agree - 1/Disagree - 5)
  - Strangers
  - Colleagues
  - Friends
  - Family
- How intuitive do you think this gesture is? (Not at all - 1/Very much - 5)

On the last page of the survey the participants were asked how comfortable they would feel performing a task on different surfaces, namely:

- How comfortable would you feel reading and interacting with your mail (which only you can see) in your palm? (Not at all - 1/Very much - 5)
  - In public
  - With friends
  - At work
  - At home
- How comfortable would you feel reading and interacting with your mail (which only you can see) on a wall? (Not at all - 1/Very much - 5)
  - In public
  - With friends
  - At work
  - At home

After each section of questions the participant were allowed to write his/hers thoughts or comments in a free-form text field. For more information about the survey, see [Appendix A].

# 4 Results and discussion

## 4.1 Requirements/paradigm

The requirements for a system as wide as the one envisioned are numerous and hard to write down with precision. In this chapter the important aspects for this project will be discussed. This includes basic features but also ones that are more specific and the ones that make our AR visor interpretation stand out from others.

A general requirement when developing AR systems is to try to utilize augmented reality as much as possible, and to use less traditional computer interaction. This can be a challenge since we're so used to computer interaction and most interaction today works in a similar manner, whether it is indirect (e.g. traditional input methods, gamepad, mouse, keyboard etc.) or direct (e.g. newer input methods, touch UIs, gaze tracking etc.) manipulation. In some cases it might be an idea to look at existing innovative UIs to get ideas, but the developer should always strive to utilize the possibilities of AR as much as possible.

The developer also needs to look at social acceptance and take the current situation into account. Limiting yourself too much because of the social acceptance factor isn't good since user acceptance is raised over time when exposed to the unfamiliar and also because the system needs to be innovative to be successful. Being too liberal on the other hand leads to users not wanting to use the system, which prevents the actions from ever becoming socially accepted.

A more functional requirement that was put on the system was to not use a wide array of different input techniques. One way of performing an action is usually enough, with the exception of an alternative "advanced user" way of doing things. As Shneiderman's second golden rule mention, allowing for advanced users to perform actions that are quicker to them is efficient, but often too advanced and not very intuitive to casual users [4].

Keeping the interaction techniques down is a way to make it easier for the user to learn how to navigate the system which in turn provides a better user experience. Combining interaction techniques to perform an action may be a viable option, especially if it includes gaze tracking. As an example, if the user wants to perform an action on an object they see it might be a good idea to combine hand interaction with gaze tracking. The hand could do a gesture and the corresponding action is then applied to the object the user is looking at. This provides a natural way of interacting with a certain object without physically touching it

to activate it. Gaze tracking is also a good way to interact with objects at a distance, out of physical reach to the user.

It is also good practice to allow for one-hand interaction in most cases concerning everyday use, since the user often is on the move or is doing something with his hands. Since the user doesn't need to hold a device to perform an action, both hands are free to do something else or to interact with. The trend lately with touch UIs on mobile phones has been that we're going towards holding the phone with one hand and interacting with the information displayed on the phone with the other. Interaction with one hand facilitates ergonomical aspects as well as use of the system in more situations.

One could argue that when using a visor based system with gaze tracking and speech recognition the system could be controlled entirely using these techniques and while being novel, still has its limitations. Full control using gaze tracking has yet to be implemented in a functional and useful way. The main problem, as mentioned, being midas touch. Implementation needs some kind of activation mechanism in order to trigger what's gazed upon, which means that gaze tracking in most cases require another interaction method in combination.

The biggest limiting factor to using full speech recognition is the technological aspects. Speech recognition requires the user to memorize certain phrases or keywords to be used in the correct combination. This puts a huge load on the users working memory which really isn't viable. Speech will be a good interaction method only when a proper artificial intelligence (AI) exist that can interpret what the user wants to do based on the sentence he/she forms. This would remove the need to use keywords and the user could speak to the system as he would a person.

A great idea when designing AR systems is to integrate the things people do every day. To check your mail, take a photo or check a map today you need to take your phone up from the pocket, turn off screen lock and then start the appropriate service. The step to bring up the phone from the pocket is the biggest advantage using an AR visor. If the application is integrated in a good way the user will use it a lot more. A good example of this is geo-tagging applications (tag real world locations with virtual data, text message or similar). These exist in phone applications today, but they haven't become popular since no one bothers taking the phone from the pocket to check for tags. If this was integrated into the visor the user would see them without having to request to see them, which is a lot more natural and follows the flow of how we like to do things. If the user doesn't want to see them he can easily turn them off and never again have to see them. It's more of a push than a pull approach, making the user aware of things instead of the user having to find things out for themselves.

The push vs pull methods are also something the developer needs to consider. Push forces information onto

the user, while pull forces the user to request the information himself. Push by default can be experienced as a rough and messy way to introduce new stuff to the user, but can be enhanced by making the approach context sensitive. This means that the system “knows” the user and can decide when and in what situations the user needs to see the information. For example, if a user is walking in town looking for a place to eat the system can aid him and show him diners that he hasn’t discovered and ones that fits his taste. It is however necessary to have a pretty perfect context sensitive system which can predict the right things. As soon as the system forces information onto the user that he doesn’t want, the user experience with the system sinks drastically and it becomes an annoyance, it is important to support inner locus of control [4]. A parallel to the previous example; if the user is walking in town looking for an electronics store, the system displays clothing stores. The possibilities of a perfect context sensitive system are today rather slim as it requires a rather intelligent AI and a set of sensors to detect all sorts of properties with the user and his surroundings.

## 4.2 Results of Social Acceptance Survey

The total amount participants in the survey landed at 35, of which 24 of these were female and 11 were male. Looking at this and the respondents’ answers there were no discrepancies between the male and female answers that can be supported by the rather small amount of respondents that participated. All graphs can be seen in Appendix B.

### 4.2.1 Social Acceptance

The result of the social acceptance part of the survey was a lot more positive than expected. On a scale between 1-5, all gestures received a mean of above 3,5 when both performing the gesture self and when seeing others performing it.

The gesture that received the best social acceptance results were “resize b”. This particular gesture was included knowing that it’s a common gesture used in popular interfaces already. Performing this gesture turned out to be very socially acceptable, with a mean of 4,51 when performing self and 4,54 when seeing others performing the gesture. The reason for the high scores is interpreted as being because it’s a very common gesture that most people have seen or performed today.

The result between performing the gesture self and seeing others perform turned out to be a clear advantage to performing the gesture on your own. This means that seeing others performing a gesture without seeing the result of it is less socially accepted than doing it yourself. Seeing others performing received lower scores across the board, if only slight in a few cases.

The second most socially accepted way of performing a gesture was in front of friends or seeing a friend performing. However, the score was to such a slight advantage compared to performing or seeing family performing that no real conclusions can be drawn with respect to the number of respondents in the survey.

The second least socially accepted to perform in front of or to see perform was colleagues and least accepted was strangers. In short, the result can only be described as the better you know a person, the more accepted it is to perform a gesture in front of or to see perform. The results compared between the different gestures were pretty much as expected, with the real surprise being the high scores. The standard deviation was also rather similar across all gestures, with nothing notable to point out.

#### 4.2.2 Intuitiveness

Making hand gestures intuitive is a rather hard task and knowing this; lower results were expected in this part of the survey. Intuitiveness is hard to convey, the survey participants easily believe something he knows well is intuitive because he knows the purpose of it. This was also reflected in the survey, where the gesture that already exists (pinch - *resize B*) received the highest intuitiveness score (3,91). Worth noting, however, is that the least socially accepted gesture (two handed - *resize C*) received second highest intuitiveness score by quite a margin (3,74). This gesture, while using two hands and being less socially accepted, is clearly more intuitive when looking at it and the result was quite expected. The practicality of this gesture is another issue. Another stand out turned out to be *close A* which received 4,43. The rest of the five gestures got somewhere in between 3,03 and 3,2.

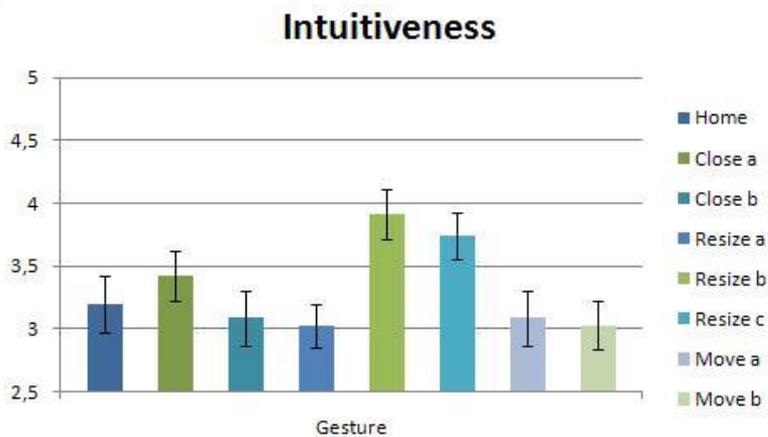


Figure 3 Intuitiveness of the different gestures

#### 4.2.3 Surfaces

When asked about which surface the participants would like to browse their data on, most people chose in the palm of the hand instead of on a wall. This ties in with the social acceptance factor and shows that people are more comfortable keeping things to themselves rather than doing gestures and working with data on something more public, like a wall. The palm of your hand is more private than a wall in that the user could have some kind of device in the palm without the spectators knowing. The fact that it is closer to the user also plays part in making it more private.

Another important aspect in this question is privacy. Users probably feel more secure browsing in their palm than on a wall. When having content on a wall it would probably feel as if you're showing it to everyone, even if you're not. Users are used to browsing private information on a cell phone or a laptop, which normally is close to the user.

#### 4.2.4 Survey Conclusion

The results from the survey, while important, don't provide a truly accurate key when designing gestures. To get the best result out of a survey like this, a lecture on the particular project is needed for the participants to really understand all the aspects of each gesture. While a description was provided with the survey of how the gestures was going to be used, all aspects weren't presented since this would have resulted in a survey too long for anyone to bother to participate in.

Knowing all the aspects, the survey can serve as a rough guideline and an insight into what works and what really doesn't. Gestures can be determined using the survey results with confidence knowing that the social acceptance factor won't be a problem. Doing intuitiveness tests also provides knowledge whether users will have trouble getting familiar with the interaction of a device or not.

Our idea to browse content in the palm is supported further by the survey, which shows that users rather browse in their palm than on a public surface. Allowing users to switch between the two is most likely the best option since it supports use in different situations and when hands are busy, but giving the user the option to keep it close and private seems to be a good starting point in order to make an AR system that users can trust and want to use.

## 4.3 Concept Video

In order to make a concept video that people like to look at it has to be visually appealing. Many projects have done this before with more experience and greater knowledge of how to make appealing concept videos than what would be possible to present in this project. In order to combat this, most focus was put on how to convey the usability and user experience to the viewer, but also on possibilities of AR that hasn't been explored before.

The main idea with the concept video was to show how a user might use the system a typical weekday doing typical activities. The vision was to show how our main character and his friends' weekdays could be enhanced using an AR visor system.

Since not all ideas could be presented in a rather short concept video, only the larger, more important and more general parts of the system were selected. The core interaction techniques were demonstrated along with a couple of typical use cases. This concluded into three different scenes.

The first scene features the main character waiting for the bus. He's bored, so he decides to play some ball. While bouncing the ball, a friend shows up and joins the game. After a while the buss arrives, they throw the ball into a trash bin and leave with the bus. The scenario includes basic system interaction, entertainment, profile look-up and system integration.

The second scene, called "Study Centre", features the main character and two of his friends, all of which are university students. They are planning to study together but one of his friends arrives a bit late. The scene includes basic system interaction, distanced communication and collaborative work.

The third and final scene features basic system interaction while hands are busy, geo-tagging and way finding. A friend of the main character has left a geo-tag on a bench. The main character walks past and activates the geo-tag. He continues by looking for a place to eat. He brings up the way finding system and selects what and where to eat. He then follows the systems recommended path to the location. The interaction here is performed on the ground in front of the user with the users feet. Feet interaction is shown to convey the possibilities of AR, rather than being the best way to

navigate in all cases. It also allows for hands busy interaction with the system, which might be important in many cases. However, the navigation in this particular system might as well have been performed with the hands. Feet interaction is a way to allow for simple system navigation while hands are busy.

## 4.4 Services

The core of the system is the services the system provides. Services are the programs or the applications that provide content to the user. To begin using the services and features of the system, the user needs to initiate this by performing an action. The alternative to this would be an automated sensor system that recognizes the user, his needs and his surroundings. Three different ways of how to initiate user content are given below:

- Browse the surroundings
  - The surroundings sense the user and triggers actions.
- Context based
  - Depending on the current context the system proposes actions accordingly.
- User initiated
  - The user initiates the content.

Automated systems are great when they function like the user wants them to, though they seldom do. All too often the result of automated systems is frustration to the end user and ultimately leads to bad user experience. Even though it might be time to eat and the user is hungry, he might want to finish the task he's currently busy with. It's a question of push vs pull (push - force events upon user, pull - user requests events), and the approach chosen for this project is pull for the reasons discussed. Using pull, the user is in full control of what he's doing and what's available to him. He initiates the starting of a service or feature on his own.

The "hardcore AR-way" of starting services and applications would be to use your surroundings. This approach isn't very practical in that you either need:

1. a gesture for each application, or
2. a set of objects close at hand at all times (tangible interaction).

The first option puts a huge strain on the working memory, which isn't very desirable. The second option requires the user to interact with the physical environment, which isn't very mobile.

The approach chosen for this project is based on two of the above given approaches; more around a traditional desktop system with integrated AR which is context based. Display a menu with all services/applications on a surface of choice. If the user is on the move he can choose to view content in his palm, and if he's sitting down by a table the table can instead be his display.



Figure 4 Menu with services

Using the browser, the user selects a service of choice by clicking the appropriate icon. The content of the chosen service is displayed on the current display, which is the surface the user has performed the gesture on. This means that the user can use any surface of choice to work on and this provides great mobility and customization to the users' preferences.

## 4.5 General System Interaction

The basic idea of the general interaction was hand gestures on surfaces. The purpose of doing gestures on a surface is because it's a lot easier ergonomically and because mid-air interaction introduces depth issues. Using mid-air interaction the user no longer touches any physical object and thus won't receive the feedback that normally would be had when touching a surface. The user will have trouble knowing at what depth to interact with the content he sees in front of himself. At all times there should be different kinds of feedback in terms of facilitating the different senses, utilizing multimodality as much as possible.

The main idea was also that all gestures preferably had to be able to be performed with one hand. One-hand interaction isn't a necessary requirement for a system of this kind, but since it facilitates mobility (hands-busy mode) and ergonomical aspects it's a great feature that users appreciate when trying one versus the other.

#### **4.5.1 Basic actions**

Figuring out the main actions that the system needs was the first step in designing the interaction.

1. Starting a display (start browsing applications)
2. Closing a display
3. Resizing a display
4. Moving a display
5. Selecting/activating content
6. Sharing content

The gestures representing these actions were chosen using a combination of methods. The brain storming session with the focus group resulted in a couple of gestures, both good and bad. These were later filtered and a couple were kept and a couple discarded. After careful consideration of the features of the system along with our novel ideas of how we wanted to do interaction the system ended up with a set of gestures. Each gesture corresponded to one of the above listed actions that were required to navigate the system. The gestures were chosen also based on our background knowledge of the system and of how to develop gestures as well as with the help of the survey.

#### **4.5.2 Starting a display**

The idea was born from the tagline "it starts in your hand". To open up a display (start browsing applications) the user performs a gesture specific to opening up the "browser". Opening the hand on a surface was chosen as the gesture since it symbolizes the opening of something or even enlarging something. The way it works is that the location of the display will have its center where the user places his hand. When the user opens his hand the starting screen of the display is opened. Depending on the size of the gesture the display gets progressively larger, meaning that the user can decide which size of the display he wants to work with when starting the display.



Figure 5 Starting a display

#### 4.5.3 Closing a display

Utilizing the fact that starting and closing a display is the opposite of each other, the decision to use the opposite gesture to close the display was quite natural and it provides good intuitiveness. By placing the fingers on top of the display and closing the hand the display closes gradually. This could also be seen as a minimization of the display, which when opened is maximized and resumed to its previous state.



Figure 6 Closing a display

#### 4.5.4 Resizing a display

By continuing the pattern with the opening and closing of the hand, the resizing gesture also comes natural. By putting the fingers on the display and either closing or opening the hand, the content is resized and zoomed.



Figure 7 Resizing display

#### 4.5.5 Moving a display

Knowing how to open, close and to resize, moving the display also comes natural. Placing the fingers on the display locks the fingers in place and allows the user to move the display around. The moving of the display stops when the user lifts his fingers from the surface.

To move a display between surfaces the user simply minimizes the display and opens it at the new location.

#### 4.5.6 Selecting/activating content

Activating applications, links, buttons etc. are easily performed with a simple click of the finger.

#### 4.5.7 Sharing

As the physical world is the medium of use when presenting information there is an obvious concern of what is seen by whom in different contexts. In some cases users might want to be certain that information shown to them is not seen by anyone else, but it might also be vice versa. When working with private content the user has the opportunity to share the content with whoever he wants to. By simply moving the desired content towards another user, whom receives an indication by seeing the display, although shaded (hiding the actual content to prevent strangers from sharing spam or similar), he knows that something is about to be “sent to his augmented world” [see figure 7]. The receiver might choose to accept it simply by placing his hand on the content [see figure 8] or he might choose to deny it by not accepting it or by brushing the shaded display away. This approach would preferably be augmented by gaze tracking which may ensure that users sharing are aware of the sharing as opposed to just moving a display on the surface. By looking at the user whom he wants to share with another layer of error preventing is added. Other users are recognized by their visors that may communicate with each other. The decision was taken to display this as an icon and feedback about the amount of people that can see the display.

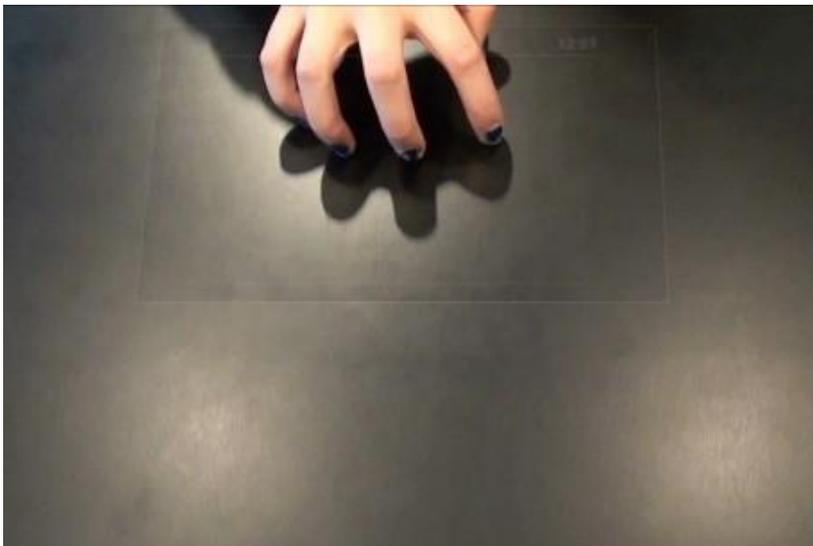


Figure 8 Spectators view of someone sharing a display



Figure 9 Accepting a display

## 4.6 Shortcuts

Experienced users want shortcuts [26]. It allows faster browsing of the system for common tasks in a way that can be customized to personal liking. By allowing the user to arrange shortcuts as he wants, he gets a place he can call his own, which is important in that it creates a bond to the system which in turn raises user experience. The user can choose to have any application of choice as a shortcut. Shortcuts are placed from the application browser. By holding the finger on top of an application icon for two seconds, a shortcut to the application sticks to the finger. The shortcut is only a copy of the icon representing the application and the actual application can't be removed from the application browser this way. The user can now place the shortcut anywhere he wants.

The user can choose to put his shortcuts anywhere. This means that he can make them mobile by putting them somewhere on his body, like for example on the back of his hand. This approach is based on the metaphor of "bag of favorite things". Compare this to a bag a person might carry when on the town: usually it contains keys, wallet, cell phone and analogous items. They are all stored in a well known place and always accessible. This allows for easy access at any time. The user can also choose make them stationary in places he often uses them. By placing a shortcut to the news paper by the dinner table he can read his news paper each morning when he's eating his serial and drinking his orange juice.



Figure 10 User's personal shortcuts

## 4.7 Notifications

Notifications (messages, calendar event etc.) are today received using either a cell phone or a laptop. These traditional communication types can be improved with the use of AR. As mentioned earlier, the physical world is the medium of data transfer. AR visor provides much better mobility since the user doesn't have to reach into his pocket and bring up his phone. The same applies for telephone calls, instead of reaching into the pocket to answer a call the user might for example instead tap the answer button on his forearm. As notifications can be received at all times it is of most importance that they do not interfere with the user's current context. The user might not want to be disturbed when participating in important business meetings or while driving a car during heavy traffic, which is why notifications should be implemented in a subtle and non-disturbing way.

When receiving a notification the receiver should be notified of this by vibrations from the arm band that tracks hand movement and by sound, similar to how it is today. In the suggested system the notifications peek up from under the arm band, indicating that an event has been triggered. By glancing at the arm strap the user can see which kind of notification it is and from whom it was sent. The user also receives feedback that there's more text in the notification, but it's covered by the arm strap. This was done to give feedback to the user that the notification can be pulled up to see it in its entirety, see figure 9.



Figure 11 Notification received and read

This method of using notifications provides great mobility and increases safety. An example of this is when the user is on the road on his bike. Normally when receiving notifications the user has to pick up his cell phone to check which can be troublesome when on a bike. If something drastic were to happen which requires immediate attention to the road the user would have to get rid of his cell phone before he's fully capable of dealing with the situation. By reading the notification on the back of the hand the user can, while still on his bike, read it and still have his hands on the handlebar. He only needs to release the handlebar when pulling up the notification, and is still very able to deal with upcoming situations. The system also mitigates driver focus when in a car. If the user has to check notifications while driving, he might do so by reading on the back of his hand while still having both hands on the steering wheel and the focus straight ahead still.

## 4.8 Wayfinding

Bowman et. al define wayfinding as a cognitive process defining a path using and acquiring spatial knowledge which is aided by natural and artificial cues [11]. Using AR is an excellent way of providing efficient and sufficient information as guidance. The mapping between map directions and real world comes a lot more natural when the information is overlaid compared to looking at a map and trying to discern which way the map is oriented or how far to go down a particular street. It's more of a synchronous and direct approach compared to the traditional indirect and asynchronous way. There is good reason to believe that the load on the short term memory is greatly reduced when using AR wayfinding since the user doesn't have to keep parameters such as distance and orientation in mind.

The suggested system utilizes a map of the surroundings projected in front of the user on the ground directly in front of his feet. Keeping the map directly in front of the user's feet makes it available at all times and doesn't obstruct the view when navigating. The map, obviously, needs to be transparent enough so that the user doesn't trip over obstacles hidden under the map. The positioning of the map is also quite discrete which likely means high social

acceptance. Alternatively the map could be positioned in the users field of view as in a HUD, but since we want to limit the amount of information shown to the user simultaneously as well as keeping the user safe when navigating in traffic the first approach is more appropriate for this situation.

Keeping a map on the ground doesn't automatically disconnect interaction. Since the map is located on the ground, interacting with the feet comes natural. Just as we're pressing buttons and icons with our finger, we could press with our feet. This is another great example of how versatile interaction in AR systems can be, if needed. Selecting a location on the map activates the location and the system displays the direction to the location as a line on the ground. The purpose of using a line is that it's easier to follow than the traditional GPS-based arrows, albeit still also utilizing lines to mark where to go. The reason for using arrows on GPS maps today is because it makes the mapping between real world and the map on the GPS unit easier, a problem that doesn't exist when using an AR based navigation.

## 4.9 Mirroring

The problem with objects out of reach has been solved by letting far away-objects be mirrored in the user's hand. By letting the user hold his hand in front of an interactive object the content is then mirrored into the hand. It can be argued that this does not provide very good social acceptance but in comparison with alternative solutions the gesture can be performed very discretely and fast, providing adequate social acceptance. When having the information at hand the user can interact with it just as with anything else (interaction surface is the palm).

Alternatives to mirroring consisted of selecting objects by pointing at or looking at objects. Pointing at icons or objects at a distance introduces technical accuracy issues as well as the common perception that it's not a very nice thing to point. Using eye-tracking the user could also activate far away objects by looking at them. The problem with Midas touch is once again introduced. An alternative to this could be to have an activation mechanism (something simple like tapping your leg with the hand) so that everything that is glanced upon isn't activated.



Figure 12 Example of mirroring

Using either of these methods or something similar the content would still end up inside the user's palm which means that the connection and the feedback between activating the content and receiving it inside the palm would be hard to make intuitive. A new idea was worked up, to perform distance selection by covering the object to be selected with the hand. This gesture has a rather high social acceptance factor as well as being able to provide ample feedback in most situations.

# 5 Conclusions

This master thesis project aimed at providing:

- Functional, intuitive and cognitively sound interaction methods.
- Information presentation that facilitates everyday duties.
- Socially acceptable interaction.

Overall we found that social acceptance is a factor of great importance. When moving the interaction space from a mobile phone's screen into our world, careful work has to be put into the interaction carried out while being on the move. By using metaphors and exploiting the human familiarity with the physical world an intuitive way of interacting can be achieved.

## 5.1 Design choices

### 5.1.1 Gestures

When designing gestures social acceptance, familiarity with the physical world as well as cultural aspects has to be taken into consideration. As proven by the social acceptance survey previous user experiences matters when being introduced to new gestures. Users are more comfortable performing certain gestures in front of specific social groups as well as performing gestures on a surface close at hand.

The survey also shows that intuitiveness plays part in remembering gestures which suggests that each gesture has to be carefully developed with respect to different cultures and habits.

### 5.1.2 Notifications

When discussing the weight of truly utilizing the potential and strength of AR and making notifications subtle there have been discussions about placing notifications in the peripheral view of the users. In this way there is a clear and yet very subtle indication that "something is happening". We felt that to use this approach there had to be more research and most important: more prototyping. The outcome would most likely result in users becoming distracted by items popping up in front of them and it also breaks Shneidermans rule of allowing the user to have full control in all situations [4]. Cognitive blindness could also occur depending on the time of occurrence of the notification.

### 5.1.3 Wayfinding

Davis et. al concludes that audio can provide useful directional and distance information [31]. By integrating 3D sound the user could get an extra informational cue of where he is situated at the moment regarding the distance to his final destination.

### 5.1.4 Prototyping

To further evaluate the gestures' functionality and feeling, a prototype session was planned. To get a feeling of the user experience it is important to prototype UIs.

When prototyping started a platform had to be chosen. A couple of different setups with SDKs exists however a lot of them are very limited with only simple tag tracking included. Some time was spent looking at and trying out different platforms, and in the end most of the work was done with ARToolkit<sup>5</sup> and Ogre 3D<sup>6</sup> engine, using a computer and a web camera. There were ideas of using the game development tool Unity<sup>7</sup> with a cell phone plug-in and an SDK that Qualcomm was developing that would be able to track real objects and register button presses. This would also have made the prototyping more functional and more advanced as well as providing better mobility, but due to scheduling issues with the release of the SDK this was never realized. Due to the restriction in the prototype that was developed using ARToolkit and lack of time, the prototyping was scrapped all together. It wouldn't have resulted in what we were aiming for because of the limitations with AR tag systems.

ARToolkit is used to identify tags and to track them. The system recognizes the tags through simple camera based vision and the user can then program what happens when the tag is recognized. With this technique we can decide which information to show at which time by systematically placing these tags in the environment and on the user. By holding or placing tags on the user's hands simple interaction can be performed using visibility and proximity algorithms.

The results of our ideas were instead presented only through the concept video.

---

<sup>5</sup> <http://www.hitl.washington.edu/artoolkit/>

<sup>6</sup> <http://www.ogre3d.org/>

<sup>7</sup> <http://unity3d.com/>

## 5.2 Obstacles Along the Way

During the course of the project there has naturally been a set of obstacles to overcome. Starting with the planning, we originally set out to do iterations of our work to refine each step in the process of developing the system, and while we have revisited some phases and reworked a lot of ideas, the iteration process have suffered, mostly due to lack of time.

Setting up requirements early on has also been troublesome. Since the work done in this project is a concept rather than a final product some aspects that normally would be taken into consideration hasn't been as important. Requirements have been added, and to some extent removed, all through the project and the requirements document has been rather messy.

An issue encountered that ties in with the requirements is that when talking to different people about the project, most would like to see it go in different directions. Some are more out-going while others are more traditional. It is important to know in what direction the system should go early on in order to not spend too much time going back and forth between different approaches during the course of the project.

## 5.3 Future work

Using the ideas and guidelines presented in this paper, more factual examples and implementations can be produced. There are several use cases and areas of use that some people might see as vital which have been left out on purpose. Many of these are application based and providing specific guidelines on application functionality is out of this thesis' scope.

Barakonyi et. al discusses the advantages of using virtual agents in AR systems [32]. They emphasize that a lot of research has been done to use agents as an intermediary, bridging the gap between the real and virtual world. Virtual agents and AR share the same goal. Since much has been done in the area of agents, the next step could be to investigate the use of agents in a system like the one described in this report. Could they help bridging the gap between real and virtual satisfactory enough while still retaining a clear distinction of what is real and not.

Another interesting topic concerning AR visor usage is the social media aspect. Some ideas have been presented in this report but there is much work still to be done in this area. Using the social acceptance survey as a starting point, further investigations of how peckish users would be using the system depending on social context can be done.

Context awareness could potentially break new grounds in AR systems.

As Johnson states it is nowadays common knowledge to test the ideas and the system on potential users, carrying out user tests with different kind of people with different backgrounds and earlier knowledge and experience [33]. Comments and information gathered in these kind of approaches usually proves to be invaluable. This becomes especially important when evaluating such a new and young area of technique with such enormous potential. It is easy to sprung away into developing interaction techniques and user interfaces. This was the case with virtual reality (VR) when that technology was new [11]. If such an evaluation was to be performed it should be performed in an iterative approach, thoroughly investigating new ideas as they arise.

# References

- [1] R.K. Azuma (1997), *A Survey of Augmented Reality*, In *Presence: Teleoperators and Virtual Environments* 6(4), Malibu, CA: Hughes Research Laboratories, pp. 355-385.
- [2] Slater, M. (1998) *Measuring Presence: A Response to the Witmer and Singer Questionnaire*, *Presence: Teleoperators and Virtual Environments*, 8(5), 560-566.
- [3] Cognetics Corporation, *LUCID - Logical user centered Interaction Design*, 1998.
- [4] B. Shneiderman & C. Plaisant, *Designing the User Interface*, 4th edition, Pearson Education, pp 74-75
- [5] Mattias Wallergård - lecture on interaction design, IKDC, 28 April 2010
- [6] L. Klein (2007), *Issues and Challenges of 3D User Interfaces: Effects of Distraction*, 5th Joint Advanced Student School (JASS 2007, St Petersburg, Russia March 25th - April 4th, 2007)
- [7] N. Petersen & D. Stricker. (2009), *Continuous Natural User Interface: Reducing the Gap Between Real and Digital World*, Proc. of the 2009 8th IEEE International Symp. on Mixed and Augmented Reality, pp-23-26
- [8] D., Holman, R. Vertegaal, M. Altosaar, N. Troje & D Johns. (2005), *PaperWindows: Interaction techniques for Digital Paper*, In Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (Portland, Oregon, USA, April 02 - 07, 2005). CHI '05. ACM, New York, NY, pp 591-599.
- [9] M. Nielsen, T. Moeslund, M. Storrøng & Erik Granum (2003), *A procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction*,. In The 5th Int. Workshop on Gesture and Sign Language based Human-Computer Interaction, Genova, Italy
- [10] H. Benko, E.W. Ishak & S Feiner (2005), *Cross-Dimensional Gestural Interaction Techniques for Hybrid Immersive environments*, Proc. of the IEEE Virtual Reality 2005, pp 209-216.
- [11] D. A. Bowman, E. Kruijff, J.J. LaViola & I.Poupyrev (2004), *3D User Interfaces, Theory and Practice book*, 1st edition, Addison-Wesley Professional, p.147, 350

[12] J. Rico & S. Brewster (2010), *Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability*, 28th ACM Conf. on Human Factors in Computing Systems, CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA

[13] J. Looser, M. Billinghamurst & A Cockburn (2004), *Through the Looking Glass: The use of Lenses as an interface tool for Augmented Reality interfaces*, GRAPHITE '04 Proc. of the 2nd international conf. on Computer graphics and interactive techniques in Australasia and South East Asia

[14] A.K. Dey, G.D. Abowd (1999), *Towards a better understanding of context and context-awareness*, HUC'99, pp. 304-307

[15] P. Dorish (2004), *Where the Action Is: The Foundations of Embodied Interaction*, 1st edition, MIT Pres, p 38, 102

[16] I.E. Sutherland (1968), *A head-mounted three dimensional display*, Proceeding AFIPS '68 (Fall, part I) Proc. of the December 9-11, 1968, fall joint computer conference, part I

[17] I.E. Sutherland (1965), *The ultimate display*, Proc. of the IFIP Congress, pp 506-508

[18] F. Zhou, H. B-L. Duh, M. Billinghamurst (2008), *Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR*, Proc.ISMAR '08 Proc. of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality

[19] M. Weiser (1991), *The Computer for the 21st Century*, Scientific American, September 1991, pp. 94-104

[20] C. de Lange (2010, August 2nd), *Future on display: the flavour-changing cookie*,  
[Online].<http://www.newscientist.com/article/dn19248-future-on-display-one-cookie-seven-flavours.html>

[21] K. Hermodsson (2010), *Beyond the keyhole*, Positional paper for the W3C Workshop: Augmented Reality on the web, June 15-16, 2010 Barcelona

[22] B. A. Parviz (2009), *For your eyes only*, IEEE Spectrum, September 2009, pp. 36-41

[23] A. Huckauf, M.H Arbina, J. Grubert et. al, *Perceptual issues in optical-see-through displays*, APGV '10 Proc. of

the 7th Symposium on Applied Perception in Graphics and Visualization

[24] S. Ronkainen, J. Häkkinen, S. Kaleva et. al (2007), *Tap input as an embedded interaction method for mobile devices*, TEI '07 Proc. of the 1st international conference on Tangible and embedded interaction

[25] H. Ishii, B. Ullmer (1997), *Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms*, Proc. of ACM Conference on Human Factors in Computing Systems -97

[26] D.A. Norman (1998), *The Design of Everyday Things*, MIT Pres pp 118-199

[27] I. Barakonyi, H. Prendinger, D. Schmalstieg et. al (2007), *Cascading Hand and Eye Movement for Augmented Reality Videoconferencing*, 3D User Interfaces, 2007 IEEE Symposium on 3D User Interfaces

[28] J. Nielsen (1992), *The Usability Engineering Life Cycle*, IEEE Computer 25, pp 12-22

[29] C. Magnusson, Topic: *Touch and gesture controlled interfaces*, Eye Openers: User experience workshop, SonyEricsson, Lund, June 1, 2010

[30] C.S Montero, J. Alexander, M.T Marshall, S. Subramanian (2010), *Would you do that? - Understanding Social Acceptance of Gestural Interfaces*, MobileHCI '10 Proc. of the 12th international conf. on Human computer interaction with mobile devices and services

[31] E.T Davis, K. Scott, J. Pair, L.F Hodges, J. Oliverio (1999), *Can Audio Enhance Visual Perception and Performance in a Virtual Environment?*, Georgia Institute of Technology

[32] I. Barakonyi, T. Psik, D. Schmalstieg (2004), *Agents that talk and hit back: Animated agents in augmented reality*, In IEEE and ACM International Symp. on Mixed and Augmented Reality (ISMAR 2004), pp 141-150

[33] J. Johnson (2010), *Simple Guide to Understanding User Interface Design Rules*, Elsevier

[34] N. Bevan (2008), *UX, Usability and ISO Standards*, CHI 2008, April 5 – April 10, 2008

[35] ISO 9241-210:2010, *Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems*

[36] B. Thomas, B. Close, J. Donoghue, J. Squires, P.D Bondi, W. Piekarski (2002), *First Person Indoor/Outdoor Augmented Reality Application: ARQuake*, *Personal and Ubiquitous Computing*, 6(1), 2002, pp. 75–86.

[37] Drgoldie (23 July 2009), available:  
[http://en.wikipedia.org/wiki/File:Milgram\\_Continuum.png](http://en.wikipedia.org/wiki/File:Milgram_Continuum.png)

[38] P. DiZio, J.R. Lackner (1997), *Motion sickness side effects and aftereffects of immersive virtual environment created with helmet-mounted visual displays*, RTO HFM Workshop on “The Capability of Virtual Reality to Meet Military Requirements”, Orlando USA, December 5-9.

# Appendix A, Personas

## **Marie Baxter - Mother**

Marie is 42 years old and works as a recreation instructor and has done so all her life. She likes to be around children and is a zealous collector of antiques. If there's a flea market around she's the first one at the stop.

In her spare time she likes to disconnect her noisy and hectic daytime by reading books. She likes detective novels because of her childhood dream of becoming a detective and because of her mundane weekdays. Whenever there's a gap in her schedule she likes to grab a coffee with the girls. They usually talk about family topics and the latest gossip.

## **Steve Baxter – Father**

Steve, 47 is an educated construction engineer. Most of his working days are spent planning construction of bridges. He's not very tech-savvy and uses computers more as a tool for getting the administrative tasks done at work. When he comes home from work he likes to do something different to completely get his mind off work. Since he's home from work before his wife he started cooking and has since gotten a taste for it. He also likes spending time in the garden, mowing the lawn and taking care of the flower beds.

Steve is a very systematic person who wants to keep work planned down to the last detail. He likes spending time with the family and during weekends he often takes them out on fishing trips and other activities.

## **Eric Baxter – Son**

Eric is 20 years old and has been practicing track and field since he was six years old. To keep up with studies, he now only jogs a couple of days a week. He's currently studying to become a sports therapist and has been doing so for the past year. His dream is to become a successful therapist and to work with professional athletes.

Eric is a quite spontaneous person who likes to spend time with his friends when he's not studying.

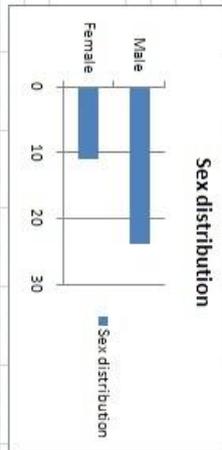
## **Ann Baxter - Daughter**

Ann is 10 years old and goes to third grade at the compulsory school. She's very fond of school and likes to do her homework. She likes to be around animals and to spend time outdoors with her friend. They're usually biking or rollerskating around the neighborhood.

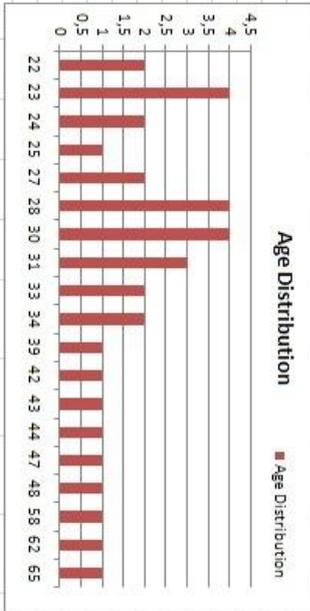
# Appendix B, User survey

Total Answers:

Female	11
Male	24



Age	22	23	24	25	27	28	30	31	33	34	39	42	43	44	47	48	58	62	65	Total
	2	4	2	1	2	4	4	4	3	2	2	1	1	1	1	1	1	1	1	35

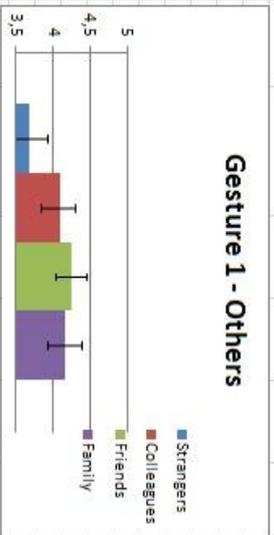
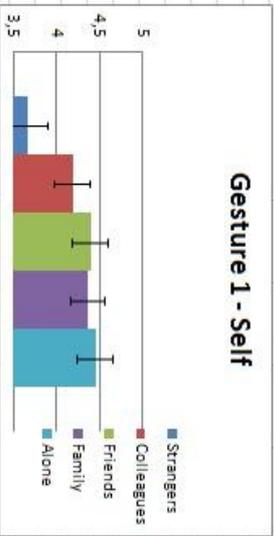


Average age: 33,43  
Median age: 30

Gesture one - Explore	1	2	3	4	5	Points
Performing self						
Strangers	4	4	8	3	16	128
Colleagues	2	2	5	4	22	147
Friends	2	2	2	3	26	154
Family	2	3	0	5	25	153
Alone	3	1	1	2	28	156
Seeing others						
Strangers	5	4	4	6	16	129
Colleagues	3	2	5	4	21	143
Friends	3	1	3	5	23	149
Family	3	2	4	3	23	146

Intuitiveness	%
1 (Bad)	14,00%
2	17,00%
3	23,00%
4	26,00%
5 (Good)	20,00%
	100,00%

Highest self: 156  
Highest others: 149



**Comments:**

I would suggest something that resembles how you would open something in real life... maybe not exactly the same, but more similar? To knock on the table seems more intuitive (knocking on a door means asking sb to open it), more comfortable, less awkward (although it certainly is for other persons). Otherw/ise, I like: [http://www.youtube.com/watch?v=yHGODp0b8k&feature=player\\_embedded](http://www.youtube.com/watch?v=yHGODp0b8k&feature=player_embedded) When you know what the gesture means it is quite obvious. If I only had seen the gesture without the explanation I would have to consider it for a while. When opening something, like a book/paper I "split" it, or if a paperbag I just pull/push the sides apart or press on an opening device... Not by putting my hand on its surface and stretching my fingers as my hand/fingers then would end up in the content of whatever I open. Thinking about a bottle/jar - why not a "twist" to unscrew something. Certainly not the first thing I would try out. Mainly because I have never seen that action before but if some (preferable many) application would use it I would probably find it "intuitive" after a while. The action of the hand, out-reaching, is un-intuitive. It's better to close the hand instead. Does gesture open shorcuts around your hand only or will it open behind and therefore be blocked by the hand?

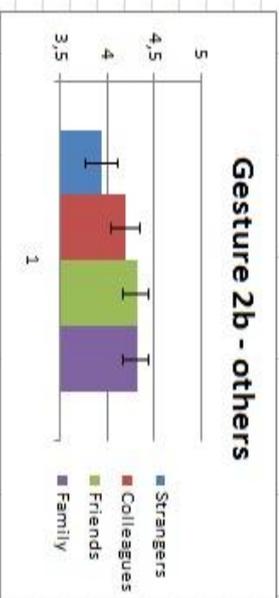
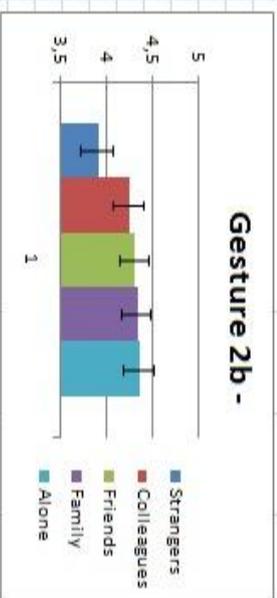
TOTAL GESTURE ONE  
Highest Gesture One SELF: Explore  
Highest Gesture One OTHERS: Explore



Gesture two - Closing2	1	2	3	4	5	Points
Performing self						
Strangers	1	1	11	9	13	137
Colleagues	1	0	6	10	18	149
Friends	1	0	5	10	19	151
Family	1	0	5	9	20	152
Alone	1	1	4	7	22	153
Seeing others						
Strangers	0	2	13	5	15	138
Colleagues	0	1	9	7	18	147
Friends	0	0	7	10	18	151
Family	0	0	8	8	19	151

Intuitiveness	%
1 (Bad)	11,00%
2	23,00%
3	31,00%
4	14,00%
5 (Good)	20,00%
	99,00%

Highest self: 153  
Highest oth: 151



Comments:

the second could 'accidentally' be performed if you just want to wipe your screen from all the touching

I prefer this one, one just gets rid of the programme. Maybe can be confusing with throwing a file in the recycle bin.

Clearly shows that something is ending

It's more intuitive than the previous one, but not perfect. It is more a feeling of someone discarding something

Could be a gesture that simply "reacts" to something, like "no, that was stupid of me to do". This makes it more comfortable to do.

Inferior to the closed hand

Put aside for the moment, perhaps not closing

That's not closing that is an minimise gesture.

Gesture 2 might intuitively mean "minimize window" or something similar, instead of "close current application".

Can cause conflicts with other gestures like moving objects

The right-to-left action isn't good. Some read right-to-left, and swiping rtl means "forward" in their view. Perhaps if you swiped only top-to-bottom?

It would, perhaps, be more intuitive if one was able to shake or wipe away the window?

More suitable for minimizing a program perhaps?

Best of the two

TOTAL GESTURE TWO

Highest Gesture Two SELF: Closing1

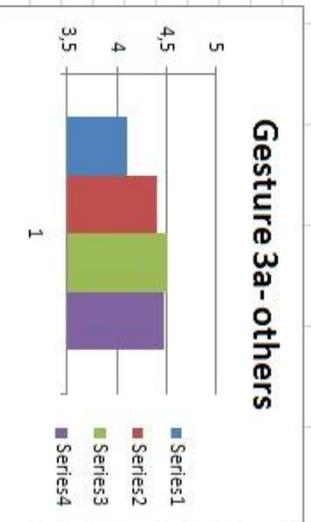
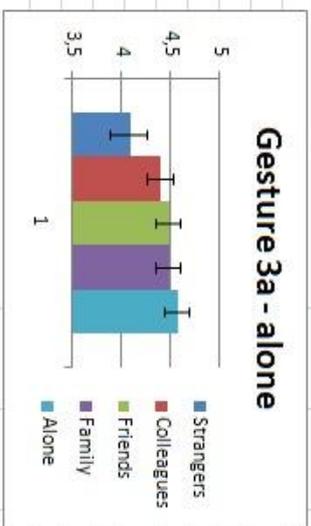
Highest Gesture Two OTHER: Closing1

Gesture three - Resizing1	1	2	3	4	5	Points
Performing self						
Strangers	1	2	7	8	17	143
Colleagues	0	2	4	10	20	156
Friends	0	1	3	9	22	157
Family	0	1	3	9	22	157
Alone	0	1	2	8	24	160
Seeing others						
Strangers	1	2	6	9	17	144
Colleagues	0	1	4	10	20	154
Friends	0	0	4	9	22	158
Family	0	0	5	8	22	157

Intuitiveness	%
1 (Bad)	3,00%
2	29,00%
3	43,00%
4	14,00%
5 (Good)	11,00%
	100,00%

Highest self: 160

Highest others: 158



Comments:

Pro: that it follows the same format you began with con: not as intuitive maybe for the goal

Can be confused with programme opening. Difficult to see how this can be accurate.

Follow how earlier gestures is shown and is therefore understandable. But it is not as clear cut.

Its applicability is related to which opening gesture to use, if the opening gesture presented in this survey is used this gesture is too similar.... I think.

Too similar to the movement when you close

Confusing with previous gestures that look similar. Otherwise it works fine.

Obvious conflict with my preferred close gesture

Firstly, you have a usage clash between this and other gestures. Nevertheless, it's good to do a small-to-big gesture (or the other way around).

I'm not that keen on a full-hand gesture, though, for a general gesture. It's not always true that you have one full hand to perform the gesture with.

I'm wondering if one might accidentally close the window? Other than that, I like the interaction.

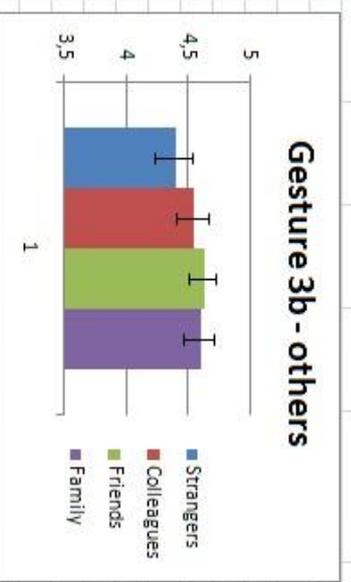
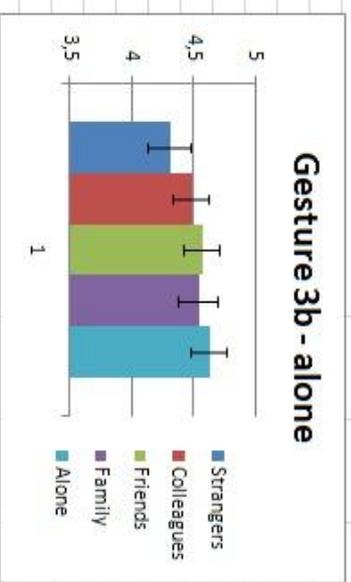
Too much of an effort

Gesture three - Resizing2	1	2	3	4	5	Points
Performing self						
Strangers	1	1	4	9	20	151
Colleagues	1	0	3	8	23	157
Friends	1	0	2	7	25	160
Family	1	1	1	7	25	159
Alone	1	0	2	5	27	162
Seeing others						
Strangers	0	2	4	7	22	154
Colleagues	0	1	3	7	24	159
Friends	0	0	3	7	25	162
Family	0	1	2	7	25	161

Intuitiveness	%
1 (Bad)	3,00%
2	11,00%
3	17,00%
4	29,00%
5 (Good)	40,00%
	100,00%

Highest self: 162

Highest others: 162



Comments:

this one is very good

Standard gesture now.

basically follow nr 1 but is a smaller gesture

Better discrimination from previous gestures

For smaller screens/devices

Apple might have influenced my opinion

My favourite

Being an iphone user, this gesture is already in my muscle memory. It's good that only two digits are needed.

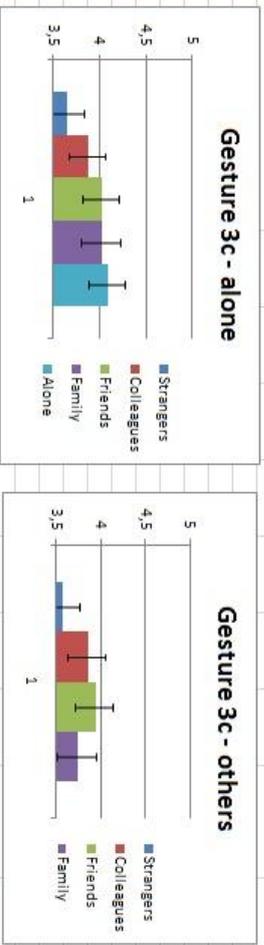
It looks a bit weak, and i'm thinking my fingers might tire if i worked a lot in an interface using this type of gesture...

Familiar.

Great gesture

Pinch, i map to zooming in and out elements in the application such as photos

Gesture three - Resizing3	1	2	3	4	5	Points		
Performing self								
Strangers	1	4	13	5	12	128	Intuitiveness	%
Colleagues	2	2	7	11	13	136	1 (Bad)	3,00%
Friends	2	2	4	12	15	141	2	11,00%
Family	3	1	4	11	16	141	3	20,00%
Alone	2	1	6	9	17	143	4	40,00%
Seeing others							5 (Good)	26,00%
Strangers	2	4	13	4	12	125		100,00%
Colleagues	2	3	10	6	14	132	Highest self:	143
Friends	3	1	9	7	15	135		
Family	3	1	9	6	16	136	Highest other:	136



Comments:

It could possibly work for large screens, it is also quite intuitive.. but maybe not as comfortable because it does require both hands.. so depends on purpose.

Can be annoying to require the use of two hands.

Two problems: requires a lot of place and the use of the two hands: to vivid and broad

It is intuitive, but it seems unnecessary to have to use both hands too messy.

Complicated since it needs two hands

A bit too much, feels more like shaping something of "clay", not a constrained window

For bigger screens/devices

Intuitive, but a bit strange with two hands.

Most intuitive but to large, physical and not practical to use two hands: It looks like a very large and invasive gesture. I would wonder whether or not the user has mental issues. On the other hand, it may be fun and intuitive to use this type of interaction in a controlled environment.

way too much effort for performing that function.

Grab the corners och expand the window. Excellent.

TOTAL GESTURE THREE

Highest Gesture Three SELF: Resizing1/Resizing2

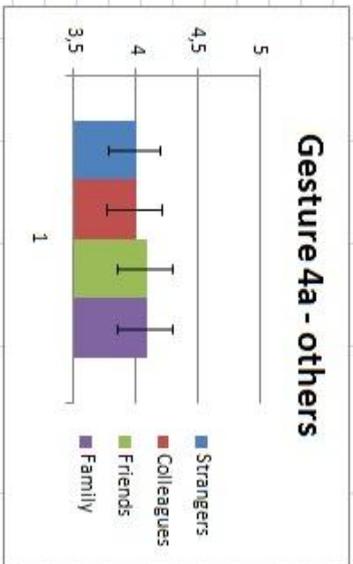
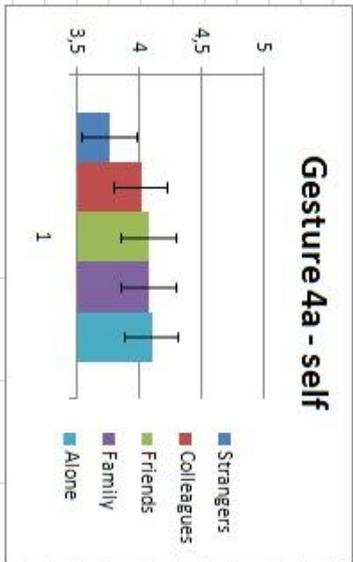
Highest Gesture Three OTHER Resizing2

Gesture four - Moving1	1	2	3	4	5	Points
Performing self						
Strangers	2	5	7	6	15	132
Colleagues	2	4	3	8	18	141
Friends	2	4	3	6	20	143
Family	2	4	3	6	20	143
Alone	2	4	2	7	20	144
Seeing others						
Strangers	4	3	7	6	15	130
Colleagues	3	3	3	8	18	140
Friends	2	4	4	4	21	143
Family	2	4	4	4	21	143

Intuitiveness	%
1 (Bad)	17,00%
2	14,00%
3	26,00%
4	29,00%
5 (Good)	14,00%
	100,00%

Highest self: 144

Highest others: 143



Comments:

The hand position when moving an application around seems a little bit awkward, why not use the same gesture as one does when moving a piece of paper on a table?

Difficult to get intuitively!

Complicated with a mirror gesture for gripping and placing

I wouldn't try a infinity gesture for that.

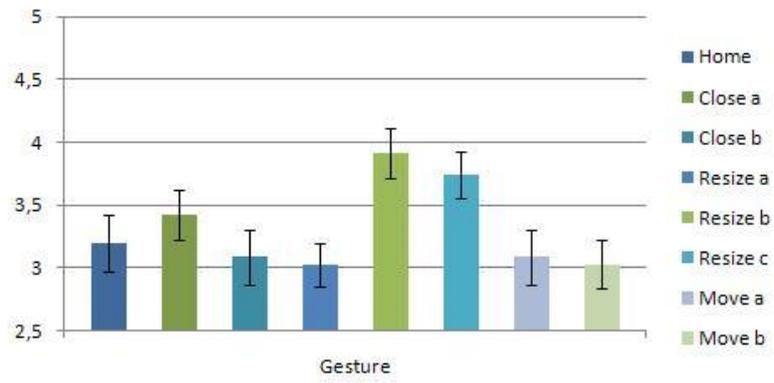
Not very intuitive, I would probably try open hand first

Too much effort in making sure that the hand touches all five points correctly along with the movement.

Seems like a full hand is the best one.



## Intuitiveness



	1	2	3	4	5	Points
Comfortable with palm						
In public	2	4	7	3	13	108
With friends	0	5	4	5	15	117
At work	2	2	4	4	17	119
At home	0	3	1	4	21	130

Comments:

The questions is not clear!

I would feel comfortable doing that but the hand palm seems to small to me.

it is of course more private that way

Early adapters would look insane. But if is common use you can get away with as talking on the phone with blue tooth devices.

Cool!

I think it will feel strange if you are the only one doing it and the technology is unknown to common public :)

Cool!

Today, there's not much difference as compared to interacting with a mobile phone when you're only working in the palm, no?

	1	2	3	4	5	Points
Comfortable with wall						
In public	10	10	6	0	3	63
With friends	6	1	13	3	6	89
At work	3	5	8	4	9	98
At home	3	1	5	5	15	115

Comments:

I dont necessarily want to share the contents with everyone.. so I wouldnt want it projected on a wall

The questions is not clear!

I just do not want anybody looking at my mails. Having them on the wall seems an invitation for others to read them.

generally speaking there is no problem to have others see what I do but it can be difficult if you attract a lot of attention in an environment where you are not totally comfortable

Why would one need to display ones mail on a wall? just seems...unnecessary.

I am not the uncomfortable type

I'd feel slightly weird, standing there and prodding at the wall. Also, sitting down comfortably, how would I reach a wall?



Figure 13 Start gesture (1)



Figure 14 Close gesture (2a)



Figure 15 Close gesture (2b)



Figure 16 Resize gesture (3a)



Figure 17 Resize gesture (3b)



Figure 18 Resize gesture (3c)



Figure 19 Move gesture (4a)



Figure 20 Move gesture (4b)

# Appendix C, Storyboards

## At the breakfast table



Steve wakes up and sits down for breakfast. As always, time is of utmost importance and Steve wants to read work stuff while eating.



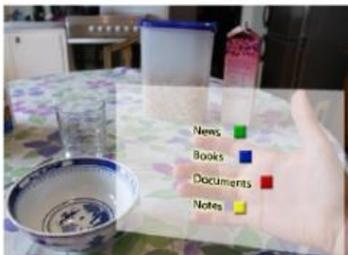
He opens up a working pane by dragging the desired size in his palm. The working pane ...



...contains everyday stuff like news...



...notes and so on.



When created Steve's holding the window in his hand. He's free to choose whatever he would like to read by..



...pinching the icon...



...and putting it...



..in the real world.

## Assembly



Eric has bought a wall mount for his TV that he's about to assemble.



He scans the bar code to signal to the system that he's about to start assemble something. The system looks up the assembly instructions.



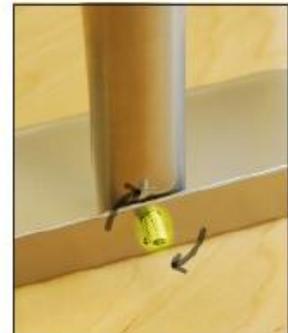
It then calibrates each physical part to correspond to the software instructions.



Eric grabs the green part and starts assembling. A green part indicates a piece that is correctly positioned.



He grabs the next part and the system shows him the step he's about to perform in yellow.



It then tells him to screw the two pieces together.



And here's Eric's wall mount in place.

## Waiting for the bus



Eric is 10 minutes early for the bus and is bored by the wait.



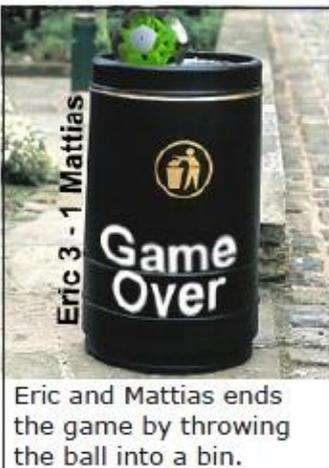
He decides to play around some with his virtual ball, he places his hands to form a ball. He's holding still for a moment and a ball starts growing inside his palms, until the ball fits in between his hands.



His friend Mattias arrives and sees Eric bouncing the ball against a wall. Mattias wants to join the game so Eric throws the ball to him.



Mattias catches the ball and by doing so he joins Eric's game automatically.

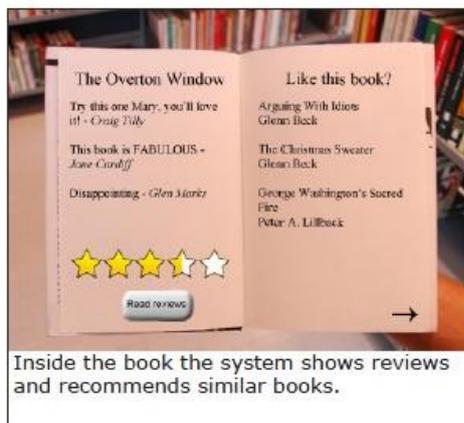
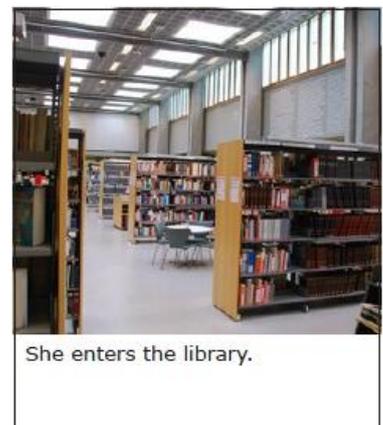


Eric and Mattias ends the game by throwing the ball into a bin.



Just in time for the bus' arrival.

## In the library



## At the supermarket



Steve arrives at the supermarket.



He's often confused by all the chaos in the store but the visor helps him to find the way to the products he and his family prefers.



After finding his way to the dairy products he checks his shopping list which is automatically updated depending on what groceries are missing in his refrigerator.



Steve's confused about which brands to choose.



The system selects the best product for Steve based on his and his family's likings and preferences.



Steve feels a relief after finding the right products



After finding the dairys needed the system guides him to the freezers since the salmon is the next grocery on the shopping list.



He picks up the salmon and receives a picture of the dish he's about to cook.

## On the town



Eric and his friends are on a break. They've decided to go out to town and have a meal.



They do not know where to go.



Eric pats himself on the stomach...



..and the application starts.



Eric wants to know more about this specific fast food restaurant. By using his finger...



..he gets a nice view of the restaurant and any good deals they might have.