Prototyping Methods for Augmented Reality Interaction

Alce, Günter

2015

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Prototyping Methods for Augmented Reality Interaction

Günter Alce

LICENTIATE THESIS
Ergonomics and Aerosol Technology, Department of Design Sciences,
Faculty of Engineering, Lund University, Sweden

Department of Design Sciences, Faculty of Engineering
ISBN 978-91-7623-365-8 (Printed)
ISBN 978-91-7623-366-5 (Pdf)

Printed in Sweden by Media-Tryck, Lund University
Lund 2015

Contents

Abstract

The age of wearable technology devices is upon us. These devices are available in many different form factors including head-mounted displays (HMDs), smartwatches and wristbands. They enable access to information at a glance. They are intended to always be “on”, to always be acting and to always be sensing the surrounding environment in order to offer a better interface to the real world. A technology suitable for these kinds of user interfaces (UIs) is augmented reality (AR) due to its ability to merge real and virtual objects.

It can be difficult and time consuming to prototype and evaluate this new design space due to components that are undeveloped or not sufficiently advanced. To overcome this dilemma and focus on the design and evaluation of new user interfaces instead, it is essential to be able to quickly simulate undeveloped components of a system to enable the collection of valuable feedback from potential users. The aim of the research presented in this thesis was to develop and evaluate two methods that can be used for prototyping AR interaction. The thesis is based on the four attached papers.

Paper 1 presents a Wizard of Oz tool called WozARd and the set of tools it offers. The WozARd device allows the test leader to control the visual, tactile and auditive output that is presented to the test participant. WozARd is also suitable for use in an AR environment where images are overlaid on the smartphone’s camera view or on glasses. The main features that were identified as necessary for simulating AR functionality were: presentation of media such as images, video and sound; navigation and location based triggering; automatically taking photos; capability to log test results; notifications; and the integration of the Sony SmartWatch for interaction possibilities.

The study described in Paper 2 is an initial investigation of the capability of the WozARd method to simulate a believable illusion of a real working AR city tour. A user study was carried out by collecting and analyzing qualitative and quantitative data from 21 participants who performed the AR city tour using the WozARd with a HMD and smartwatch. The data

analysis focused on seven categories that can have a potential impact on how the WozARd method is perceived by participants: precision, relevance, responsiveness, technical stability, visual fidelity, general user experience, and human operator performance. Overall, the results seem to indicate that the participants perceived the simulated AR city tour as a relatively realistic experience despite a certain degree of technical instability and human operator mistakes.

Paper 3 presents a proposed method, called IVAR (Immersive Virtual AR), for prototyping wearable AR interaction in a virtual environment (VE). IVAR was developed in an iterative design process that resulted in a testable setup in terms of hardware and software. Additionally, a basic pilot experiment was conducted to explore what it means to collect quantitative and qualitative data with the proposed prototyping method. The main contribution is that IVAR shows potential to become a useful wearable AR prototyping method, but that several challenges remain before meaningful data can be produced in controlled experiments. In particular, tracking technology needs to improve, both with regards to intrusiveness and precision.

The goal of Paper 4 was to apply IVAR to evaluate the four interaction concepts from Paper 3: two for device discovery and two for device interaction implemented in a virtual environment. The four interaction concepts were compared in a controlled experiment. Overall, the results indicate that the proposed interaction concepts were found natural and easy to use.

Overall, the research presented in this thesis found the two prototyping methods, the WozARd and the IVAR method, to be useful for prototyping AR interaction but several challenges remain before meaningful data can be produced in controlled experiments. WozARd is flexible in terms of being easy to add new UI, and is sufficiently stable for prototyping an eco-system of wearable technology devices in outdoor environments, but it relies on a well-trained wizard operator. IVAR is suitable for simulations of more complex scenarios, e.g. since registration and tracking easily can be simulated. However, it has the disadvantage of being static, since users need to sit down and their movements are somewhat limited because they are connected to a computer with cables.

Sammanfattning

Bärbara enheter har på senare tid fått stor uppmärksamhet. Dessa enheter finns i många olika formfaktorer så som huvudmonterade skärmar eller på engelska headmounted displays (HMDs), smarta klockor och armband. Den här typen av enheter gör det möjligt att lätt få tillgång till information. De är avsedda att alltid vara aktiva och alltid känna av den omgivande miljön för att kunna erbjuda ett bättre användargränssnitt. Förstärkt verklighet eller på engelska Augmented Reality (AR) är en teknik som lämpar sig för dessa typer av användargränssnitt tack vare dess förmåga att slå samman verkliga och virtuella objekt.

Outvecklade komponenter, eller komponenter som inte är tillräckligt avancerade, gör det svårt att bygga och utvärdera prototyper av nya interaktionskoncept som de bärbara enheterna möjliggör. För att kringgå detta dilemma och istället fokusera på design och utvärdering av nya användargränssnitt är det viktigt att snabbt kunna simulera outvecklade komponenter i ett system för att göra det möjligt att samla värdefull återkoppling från potentiella användare. Syftet med forskningen som presenteras i denna licentiatuppsats var att utveckla och utvärdera två metoder som kan användas för att bygga och utvärdera prototyper av AR-interaktion. Licentiatuppsatsen baseras på fyra artiklar.

Artikel 1 introducerar ett Wizard of Oz-verktyg som heter WozARd. WozARd möjliggör för testledaren att styra den visuella, taktila och auditiva stimuli som presenteras för testdeltagaren. WozARd är även lämplig för användning i en AR-miljö där bilderna överlagras på mobiltelefonens kameravy eller på HMD. De centrala tjänster som identifierades som nödvändiga för att kunna simulera AR-funktionalitet var: presentation av media så som bilder, video och ljud; navigation och platsbaserad aktivering; automatisk bildtagning; samla testdata; presentera notifieringar; samt integration av Sony's SmartWatch som interaktionsenhet.

Studien som beskrivs i artikel 2 är en första undersökning av WozARd metodens förmåga att simulera en trovärdig illusion av en verklig AR-stadstur. En användarstudie genomfördes genom att samla in och

analysera kvalitativ och kvantitativ data från 21 deltagare som utförde AR-stadsturen med hjälp av WozARd kopplad till en HMD och Sony's SmartWatch. Data- analysen fokuserade på sju kategorier som kan ha en potential inverkan på hur WozARd-metoden uppfattades av deltagarna: precision, relevans, responsivitet, teknisk stabilitet, visuell trovärdighet, allmän användarupplevelse, och testledarens prestationsförmåga. Sammantaget visar resultaten från användarstudien på att deltagarna upplevde den simulerade AR-stadsturen som en relativ realistisk upplevelse trots viss teknisk instabilitet och misstag av testledaren.

Artikel 3 presenterar en metod kallad IVAR (Immersive Virtual AR). Tanken med metoden är att kunna bygga och utvärdera prototyper av bärbar AR-interaktion i en virtuell miljö. IVAR utvecklades i en iterativ designprocess som resulterade i en testbar uppställning i form av hård-och mjukvara. Dessutom genomfördes ett pilot-experiment för att undersöka vad det innebär att samla kvalitativ och kvantitativ data med den föreslagna metoden. Det största bidraget från studien är att IVAR visar potential att bli en användbar metod för att bygga och utvärdera prototyper av bärbar AR-interaktion. Dock kvarstår flera utmaningar innan meningsfull data kan samlas in i kontrollerade experiment. Framför allt spårningstekniken (tracking) måste förbättras med avseende på precision och påträngdhet.

Målet med artikel 4 var att tillämpa IVAR-metoden och utvärdera fyra interaktionskoncept från artikel 3. Två koncept för att hitta enheter och två för att interagera med enheter implementerades i en virtuell miljö. De fyra interaktionskoncepten jämfördes i ett kontrollerat experiment. Sammantaget tyder resultaten på att deltagarna tyckte att de föreslagna interaktionskoncepten var lätta och naturliga att använda.

Sammantaget verkar de två metoderna WozARd och IVAR vara användbara för att bygga och utvärdera prototyper av bärbar AR-interaktion men flera utmaningar kvarstår innan meningsfull data kan samlas in i kontrollerade experiment. WozARd är flexibel när det gäller att lätt kunna ändra på användargränssnitt och är tillräckligt stabil för att kunna bygga och utvärdera prototyper av bärbara enheter i en utomhusmiljö men kräver en erfaren testledare. IVAR är lämplig för simulering av mer komplexa scenarier bland annat eftersom registrering och spårning av virtuella objekt enkelt kan simuleras. Emellertid har metoden nackdelen av att vara statisk, eftersom användarna behöver sitta ner och deras rörelser är något begränsade på grund av att de är anslutna till en dator med kablar.

Acknowledgements

Although, there is only one author, a licentiate thesis could never be written without the support from colleagues, friends and family. There are a lot of people I would like to thank for contributing to this research and for supporting me in different ways:

My two lovely sons Berk and Beran and my lovely wife Beyhan for supporting and listening to all the challenges that I have faced during this exciting adventure.

Klas Hermodsson and Mattias Wallergård for introducing me to the fascinating worlds of human-computer interaction, augmented reality and the internet of things.

Gerd Johansson for showing me all those important things about research.

Lars Thern and Tarik Hadzovic for all their hard work on developing IVAR. Special thanks go to Lars for helping me with the papers regarding IVAR.

Troed Sångberg for helping me to formulate the research questions.

Magnus Svensson for support in determining the areas I should look into and for finding ways to allow me to continue my research.

Sarandis Kalogeropoulos for pushing the management at Sony Mobile to sign the papers.

My Ph.D. candidate colleagues for sharing their challenges with me; in that way I knew that I was not alone.

Joakim Eriksson for always sharing his great knowledge, support and assisting during my studies at the usability lab.

Eileen Deaner for proofreading and for other valuable input to this thesis.

All the people who participated in my research for taking the time and effort to share their thoughts.

My parents Kadri and Mürvet for support and help in leaving and picking up the kids from daycare.

My two sisters Gülseren and Gülsen for always taking care of their “little” brother.

The European 7th Framework Program under grant VENTURI (FP7-288238) for funding the research.

The VINNOVA sponsored Industrial Excellence Center EASE (Mobile Heights) for funding the research.

List of included papers

Paper 1: WozARd: A Wizard of Oz Tool for Mobile AR

Alce, G., Hermodsson, K. and Wallergård, M. (2013). Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services – MobileHCI '13. (pp. 600-605). ISBN: 9781450322737.

This conference paper describes the Wizard of Oz tool called WozARd. The tool was developed by master thesis students and the respondent stabilized and redesigned it. The respondent was responsible for writing the conference paper. Klas Hermodsson and Mattias Wallergård critically reviewed the text.

Paper 2: WozARd: A Wizard of Oz Method for Wearable Augmented Reality Interaction – A Pilot Study

Alce, G., Wallergård, M. and Hermodsson, K. (2014). Resubmitted after minor revision to the journal of Advances in Human-Computer Interaction, Hindawi Publishing Corporation.

The goal of the presented pilot study was to perform an initial investigation of the capability of the WozARd method to simulate a believable illusion of a real working AR city tour. The respondent was responsible for the execution of the experiment, analysis of the data, and writing the article. Mattias Wallergård helped to plan the experiment. He also critically reviewed the text along with Klas Hermodsson.

Paper 3: A Prototyping Method to Simulate Wearable Augmented Reality Interaction in a Virtual Environment – A Pilot Study

Alce, G., Wallergård, M., Thern, L., Hermodsson, K., and Hadzovic, T. (2015). To be submitted, to the International Journal of Virtual Worlds and Human Computer Interaction, Avestia publishing.

This paper describes a prototyping method that was used for simulating wearable AR interaction in a virtual environment with relatively inexpensive, off-the-shelf devices. The work was done and presented in a master thesis by Lars Thern and Tarik Hadzovic. The respondent was responsible for the execution of the experiment, analysis of the data, and writing the paper. All five authors jointly analyzed the data, wrote the paper and critically reviewed it.

Paper 4: Feasibility Study of Ubiquitous Interaction Concepts

Alce, G., Thern, L., Hermodsson, K. and Wallergård, M. (2014). Proceedings of the 6th International Conference on Intelligent Human-Computer Interaction – iHCI '14. (pp. 35-42). DOI: 10.1016/j.procs.2014.11.007.

This conference paper discusses four interaction concepts that were developed using the IVAR method described in Paper 3. The respondent was responsible for writing the conference paper and wrote it with Lars Thern. Klas Hermodsson and Mattias Wallergård participated in the initial planning of the paper and critically reviewed it.

Other publications by the respondent

Chippendale, P., Prestele, P., Buhrig, D., Eisert, P., BenHimane, S., Tomaselli, V., Jonsson, H., Alce, G., Lasorsa, Y., de Ponti, M. and Pothier, O. (2012). VENTURI – immersiVe ENhancemenT of User-woRld Interactions. White paper. https://venturi.fbk.eu/documents/2012/09/venturi-white-paper-year-1.pdf.

Chippendale, P., Tomaselli, V., D'Alto, V., Urlini, G., Modena, C.M., Messelodi, S., Strano, M., Alce, G., Hermodsson, K., Razafimahazo, M., Michel, T. and Farinella, G. (2014). Personal Shopping Assistance and Navigator System for Visually Impaired People. 2nd Workshop on Assistive Computer Vision and Robotics – ACVR at ECCV. Zurich, Switzerland, 12 September 2014.

Introduction

Imagine the following scenario:

Adam is very interested in wearable technology devices and just recently bought Sony SmartEyeglasses and a Sony smartband. The glasses can augment the world around him with overlaid information and the wristband can track biometric values that are useful for identifying the user and detecting his mood. Adam visits a town where he has never been before. He enjoys exploring a new city and drinking coffee at old local cafés. He decides to go for a walk. He can see in the corner of his glasses that there is a historical café in the neighborhood. He strolls in that direction, but to get to the café he has to cross a wonderful bridge. The glasses detect that Adam is spending a lot of time looking at the bridge and the wristband senses from his biometric numbers that he is interested in the bridge. Because of this, the glasses show Adam a picture of the Ottoman architect who designed and built the bridge, and told him that it was constructed in 1566. He locates the historical café after crossing the bridge. Upon arrival, the glasses recommend a coffee on the menu based on his preferences. He can also see that a friend has been here who recommends a pastry called baklava to have with his coffee. After enjoying his visit, he heads back to the hotel. He goes to his room, lies down on the bed and wants to watch a documentary about the bridge. He takes control of the TV by making a grabbing gesture in the air towards the screen. He opens his hand on the bed in front of him on which the glasses project a list of documentaries about the wonderful bridge. He chooses to watch the one called "Mostar – A City with Soul in 1 Day."

The scenario above is an example of how people, wearable technology devices and communication are seamlessly integrated. The age of wearable devices is upon us. These devices are available in many different form factors including head-mounted displays (HMDs), smartwatches and wristbands (Genaro Motti & Caine, 2014). Wearable devices enable information at a glance (Baker, Hong, & Billinghurst, 2014). They are intended to always be "on", to always be acting and to always be sensing the surrounding environment in order to offer a better interface to the real world (Rekimoto, Ayatsuka, & Hayashi, 1998). Ideally, in a world where the digital and physical are bridged, users would not think of how to interact with systems. Everything would just seamlessly work perfectly as in the scenario.

A technology suitable for these kinds of user interfaces (UIs) is augmented reality (AR) due to its ability to merge real and virtual objects. AR technology has reached consumers through smartphones because they come with inexpensive, powerful embedded processors and sensors (Barba, MacIntyre, & Mynatt, 2012). However, according to Barbra et al. (2012), the ubiquity of the smartphone is owed, in part, to its emergence as the "Swiss army knife" of handheld computing. It is capable of many things, but ideal for none of them. Hermodsson (2010) lists some known limitations of using a smartphone for experiencing AR:

- **Limited view.** Instead of augmenting the user's world, the user looks at the augmented reality through a keyhole.
- **Awkward interaction.** AR users should not need to hold a device in front of them (this feeling of awkwardness is similar to that of most people standing in a public spot and holding up a camera in front of them for extended periods). The smartphone is both socially awkward and physically tiring.
- **Relying on a camera sensor.** When the display shows an augmented camera view, the world is degraded to the quality and speed of the camera sensor. A camera sensor drains battery power and is inferior to the human eye for sensing the world around us.
- **Limited use.** The user must actively initiate the use of the AR application and point the device in the desired direction for there to be any augmented information. This usage method results in use for a short time and only when the user has decided that he or she would like to know more about something.

The next natural step towards a more usable, immersive and comfortable AR experience would be to have *full peripheral view* using, for example, a head-mounted display (HMD).

Although HMDs have been developed and used in research since the 1960s (Sutherland, 1968), it has not been until recently that they have become available outside of the research lab. Example are Google Glass (2013), Meta Pro (2014), Recon Jet (2014), Vuzix M100 (2014), Epson Moverio BT200 Smart Glasses (2014), and Sony SmartEyeGlass (2015). Recently, Microsoft HoloLens (2015) was presented, which is able to create high quality holograms and enables the user to interact using gestures, touch and voice.

However, it is difficult and time consuming to prototype and evaluate this new design space due to components that are undeveloped or not sufficiently advanced (Davies, Landay, Hudson, & Schmidt, 2005). To overcome this dilemma and focus on the design and evaluation of new user interfaces instead, it is essential to be able to quickly simulate undeveloped components of the system to enable the collection of valuable feedback from potential users. The aim of the research presented in this thesis was to develop and evaluate two methods that can be used for prototyping AR interaction. By using these methods the scenario described in the beginning of the introduction can be experienced, at least on an elementary level.

Theoretical Overview

This section provides the reader with a basic description of the areas that the thesis covers.

Wearable technology

Wearable technology is based on computational power which can be worn. According to Mann, wearable computing is defined as "the study or practice of inventing, designing, building, or using miniature body-borne computational and sensory devices" (2014). This means the device will be worn and will always be on and running (Mann, 1998). Examples of wearable devices include smartwatches, glasses, jewelry and clothing (Figure 1).

According to Billinghurst & Starner (1999), the elements of a wearable device work to satisfy three goals. The first and most obvious is that it must be mobile. By definition, a wearable must go where its wearer goes.

The second goal is to augment reality, for example, by overlaying computer-generated images or audio on the real world. Unlike virtual reality (VR), augmented reality (AR) seeks to *enhance* the real environment, not replace it.

The third goal is to provide context sensitivity. When a computer device is worn it can be made aware of the user's surroundings and state. Context-sensitive applications can be developed to exploit the intimacy between the human, the computer, and the environment. An example is the Touring Machine (Feiner, MacIntyre, & Höllerer, 1997), developed by Steve Feiner of Columbia University, which uses a global positioning system (GPS) receiver and a head-orientation sensor to track the wearer as he walks around looking at various buildings on campus.

a)

b) c)

Figure 1. Example of wearable technology devices, a) Sony SmartEyeGlass (2015)**, b) Sony SmartWatch 3** (2015)**, c) Misfit Shine Bloom Necklace** (2015)**.**

Augmented reality

Augmented reality (AR) is a variation of virtual environments (VE), or virtual reality (VR) as it is more commonly called. VR technologies aim to completely immerse a user inside a synthetic environment. While immersed, the user cannot see the real world around him. In contrast, AR allows the user to see the real world with virtual objects superimposed on it (Figure 2) or composited with the real world. Thus, AR supplements reality, rather than completely replacing it (Azuma, 1997). In his survey, Azuma (1997) defines AR as a system that has the following three characteristics:

- Combines real and virtual
- Interactive in real time

- Registered in 3D

This definition allows other senses then vision to be augmented. Examples are hearing, smell, touch, temperature and taste.

Figure 2. Example of an AR application superimposing virtual objects on the real world (Byrne, 2010).

Milgram's Continuum defines the differences between real and virtual environments (Figure 3). Virtual environments (VE) immerse a user inside a virtual world. In opposition to VE, AR still resides in the real world but provides overlaid virtual information. To summarize, you could say that users of a VE are a part of the computer world while AR aims to make computers become a part of the real world.

Figure 3. Mixed reality continuum (Milgram & Kishino, 1994).

The design process

Designing an interactive system typically involves an iterative process of brainstorming, prototyping, development, user testing, and evaluation (Dow, MacIntyre, & Lee, 2005). This is not a clear-cut process; it often iterates through many cycles before reaching a final system.

According to Buxton (2010) sketches dominate the early ideation stages, whereas prototypes are more concentrated at the later stages. Much of this has to do with the related attributes of cost, timeliness, quantity, and disposability. This is illustrated by the design funnel in Figure 4. At the front end of the funnel, when there are lots of different concepts to explore and things are still quite uncertain, sketching dominates the process. The change in color reflects a transition from a concentration on sketching at the front to one on prototyping at the back (Buxton, 2010). The role of prototyping is to facilitate the exploration of a design space and uncover relevant information about users and their work practices by giving more details than a sketch and being testable.

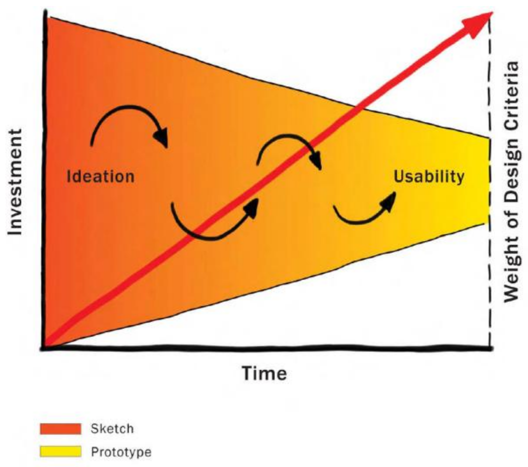


Figure 4. The dynamics of the design funnel (Buxton, 2010).

Prototyping methods

Prototyping is an important component in developing interactive systems (Rogers, Sharp, & Preece, 2011). Prototypes serve different purposes in interaction design. They are used, for example, to communicate between designers as well as with users, developers and managers. Prototypes are also used to expand the design space, to generate ideas and for feasibility studies.

Beaudouin-Lafon and Mackay (2003) define a prototype as a concrete representation of part or all of an interactive system. Designers, managers, developers, customers and end-users can use these artifacts to envision and reflect upon the final system.

Methods that are commonly used when prototyping interactive systems include low fidelity prototyping (e.g., paper prototyping and sketches), bodystorming, pretotyping, and Wizard of Oz. Each method has its advantages and disadvantages, which will be explained in the following sections.

Low fidelity prototyping

Lo-fi prototyping includes paper prototypes and sketches. Buxton (2010) lists a set of characteristics for lo-fi prototyping: quick to make, inexpensive, disposable and easy to share (Figure 5). However, they serve best for standard graphical UI interaction. Lo-fi prototyping dominates at the beginning of new projects, when ideas are considered to be “cheap”, “easy come, easy go” and “the more the merrier.” Low fidelity prototyping can be very effective in testing issues of aesthetics and standard graphical UI interaction. However, higher fidelity is preferable when designing for an eco-system of wearable devices and/or for AR interaction, (Carter, Mankoff, Klemmer, & Matthews, 2008).

Figure 5. Low fidelity prototyping of a smartwatch (Mattsson & Alvtegen, 2014).

Bodystorming

The idea of bodystorming is that the participants and designers go to a representative environment; if studying shopping malls, they will go to a representative shopping mall (Figure 6). Oulasvirta, Kurvinen, & Kankainen (2003) state that in this way, the descriptions of a problem domain (i.e., design questions) given to the bodystorming participants can concentrate more on different aspects of the problem that are not observable: the psychological (e.g. user needs), the social (e.g. interpersonal relationships) and the interactional (e.g. turn-taking in conversations). Bodystorming allows the participants to actively experience different, potential use cases in real time. Additionally, bodystorming sessions have proven to be memorable and inspiring.

Bodystroming is inexpensive, quick and helps to detect contextual problems. However, it is not easy to share the outcome of the session. In addition, a representative environment is sometimes hard to find.

Figure 6. Bodystorming at a shopping mall.

Pretotyping

The idea behind pretotyping is to start building the design idea with a low fidelity prototype using cardboard or even a piece of wood as did Jeff Hawkins, the founder and one of the inventors of the Palm Pilot (Figure 7). He used the wood and pretended as if the "thing" was working, which helped him figure out what did work and what did not (Savoi, 2011).

Alberto Savoi (2011), originator of the word "pretotyping" defines it as: "Testing of the initial attractiveness and actual use of a potential new product with minimal investment of time and money by simulating the experience of its core."

According to Savoi, prototyping is important and should be used to answer questions including: Is it possible to build? Will it work? What size should it be? How much should it cost? How much power should it use?. Pretotyping, on the other hand, focuses on answering the question: Is this the right "thing" to build?

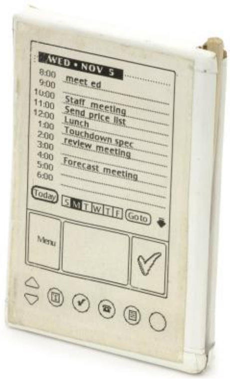


Figure 7. Jeff Hawkin's wooden PalmPilot (PalmPilot wooden model, 1995).

Wizard of Oz

The Wizard of Oz (WOZ) technique lets users experience interactive systems before they are real, even before their implementation (Buxton, 2010).

The idea is to create the illusion of a working system. The person using it is unaware that some or all of the system's functions are actually being performed by a human operator, hidden somewhere "behind the screen." The method was initially developed by J.F. Kelley in 1983 to simulate a natural language application (Kelley, 1983). The WOZ method has been used in a wide variety of situations, particularly those in which rapid responses from users are not critical. WOZ simulations may consist of paper prototypes, fully-implemented systems and everything in between (Beaudouin-Lafon & Mackay, 2003).

The WOZ method is a good way to quickly test new design ideas; it is easy and inexpensive. However, it relies highly on the human operator, which can compromise the validity and reliability of user test data.

Virtual reality

Virtual reality (VR) uses computer-generated graphical simulations to create "the illusion of participation in a synthetic environment rather than external observation of such an environment" (Gigante, 1993). The term VR is used more specifically to describe the technology that consists of the devices used to generate the virtual environment (Stanney, 2002). However, both terms are used as synonyms to each other.

Two important concepts in the field of VR are “presence” and “immersion.” According to Slater (1998), “Immersion is a description of a technology, and describes the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant.” Factors that contribute to immersion include field of view, resolution, stereoscopy, type of input and latency.

Slater defines presence as “the subjective experience of being in one place or environment, even when one is physically situated in another”. According to Slater, presence includes three aspects:

- The sense of “being there” in the environment depicted by the VE.
- The extent to which the VE becomes the dominant one, that is, that the participant will tend to respond to events in the VE rather than in the “real world.”
- The extent to which participants, after the VE experience, remember it as having visited a “place” rather than just having seen images generated by a computer.

In the last couple of years, a lot of technology that enables VR has become more inexpensive and easier to work with. For instance, when Oculus VR started shipping their Oculus Rift Developer Edition in 2013, the subject of immersive VR exploded on the scene. This inspired others to develop similar devices such as OpenVR (Yildirim, 2014), OpenDive (Welker, 2013), Google Cardboard (2014), Samsung Gear VR (2014) and Sony Morpheus (2014).

Overview of the prototyping methods

This section presents an overview of the methods that were used for prototyping AR interaction in the research. The first method called WozARd is based on the Wizard of Oz (WOZ) method and the second method called IVAR (Immersive Virtual AR) is based on VR technology.

Although, WOZ has been used for a long time and in various application areas, there is still no WOZ tool that the author is aware of that can be used to prototype AR interaction that works in both indoor and outdoor environments, and that can be used with HMDs and other wearable devices integrated with a smartphone (e.g., based on Android) for mobility. In an attempt to meet these requirements we developed a tool that consists of two Android devices communicating with each other wirelessly (Figure 8). The tool is called WozARd and is suitable for AR interaction since it allows an eco-system of wearable devices; it is usable both indoors and outdoors and flexible in terms of being easy to add new UI (Figure 9). See Papers 1 and 2 for more details.

Figure 8. System overview of WozARd.

Figure 9. WozARd in use.

Although WozARd is easy and flexible to use, it also has some undeveloped parts or parts that do not function very well. These include the registration and tracking of virtual objects and the reliance on a human operator.

The prototyping method, IVAR, uses off-the-shelf input/output devices to prototype wearable AR interaction with in a Virtual Environment (VE). The devices (Figure 10) that were used include:

1. The Oculus Rift Development Kit (Oculus Rift-Virtual Reality Headset for Immersive 3D Gaming, 2014), a head mounted display showing the VE.

2 a, 2b. Razer Hydra|Sixense (2014), a game controller system that tracks the position and orientation of the two wired controllers.

3. 5DT Data Glove Ultra (2014), that tracks finger joint flexion in real time.

4. Sony Xperia Tablet Z (2013), the tablet allows the system to capture and react to touch input from the user. Additionally, it offers tactile feedback, resulting in higher immersion.

5. Android powered smartphone. This device is attached to the wrist of the user's dominant arm and is used to give haptic feedback through vibrations.

6. Desktop computer with a powerful graphics card. This computer executes and powers the VE through the use of the Unity game engine (2014).

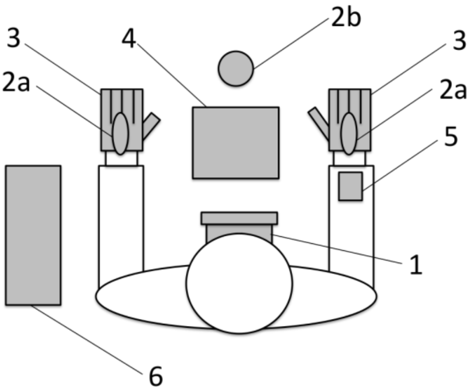


Figure 10. System overview of IVAR.

Most of the IVAR system components are wired, making this setup unsuitable for interaction where the user needs to stand up and walk around. However, the setup works for use cases that involve a seated user. For this reason, it was decided to implement a VE based on a smart living room scenario in which a user sitting in a sofa can interact with a set of consumer electronic devices. Four well-known interaction concepts with relevance for wearable AR were implemented in the VE (Figure 11). The concepts support two tasks that can be considered fundamental for a smart living room scenario: device discovery and device interaction. IVAR is capable of simulating technologies that are not yet developed, and to simulate the registration and tracking of virtual objects such as text description popping up in front of the TV. It is also easy and inexpensive to add more virtual devices such as a TV, tablets and wristband. It is different from the WozARd in that it does not rely on a human operator; the user interacts as he or she wishes. However, the method has the disadvantage of being static, since users need to sit down and their movements are somewhat limited because they are connected to a computer with cables (Figure 11). See Papers 3 and 4 for more details.

Figure 11. IVAR in use.

Methodology

This section, describes the methods used when conducting user studies, followed by a presentation of the participants.

Methods

Different research methods were used for the different experiments. The methods included observations, interviews, questionnaires, and think aloud.

Observations. Observation is a useful data gathering technique at any stage during product development. Observation conducted later in development, e.g. in evaluation, may be used to investigate how well the developing prototype supports the tasks and goals. Users may be observed directly by the investigator as they perform their activities, or indirectly through records of the activity (Rogers et al., 2011). In the experiment described in Paper 2, indirect observation of the recorded videos was performed and in the experiment described in Paper 3 and 4 direct observation was performed.

Interviews. There are four types of interviews: open-ended or unstructured, structured, semi-structured, and group interviews (Frey & Fontana, 1994). If the goal is to gain first impressions about how users react to a new design idea, then an informal, open-ended interview is often the best approach. But if the goal is to get feedback about a particular design feature, such as the layout of a new web browser, then a structured interview or questionnaire is often better (Rogers et al., 2011). In the experiment described in Paper 2, open-ended interviews were conducted together with a questionnaire. An open-ended interview was used to gather qualitative data, and a questionnaire designed particularly for the experiment to collect quantitative data. For the experiments described in Papers 3 and 4, though, semi-structured interviews were conducted together with the NASA-TLX Workload Questionnaire (Hart, 2006).

Questionnaires. Questionnaires are a well-established technique for collecting demographic data and users' opinions. They are similar to interviews in that they can have closed or open questions (Rogers et al., 2011). Efforts are needed to ensure that questions are clearly worded and the data collected can be analyzed efficiently. As mentioned, a questionnaire was designed particularly for the experiment in Paper 2 to collect demographic and quantitative data regarding six categories that can have a potential impact on how the WozARd tool is perceived by participants: responsiveness, precision, relevance, visual fidelity, general user experience, and technical stability. The questionnaire was inspired by the System Usability Scale (SUS) (2013). In Papers 3 and 4, the NASA-TLX Questionnaire was used to measure the perceived workload for the specific tasks.

Think aloud. Think aloud is one of the most direct and widely used methods to gain information about participants' internal states (Ericsson & Simon, 1980). The think-aloud method was used only in the experiment described in Paper 2. The method had two purposes: to gain information on the participants' experience when attending to the information and to aid the human operator in understanding if the participants were experiencing any problems. However, very few participants actually said anything during the city tour since they probably were focused on the task of following the instructions given from the "system."

Video analyses were used in all experiments. Data logging included time, distance, performed errors and recovery time. For the experiment described in Paper 2, all test sessions were recorded and transcribed. Each participant's video recordings were analyzed, with individual quotes categorized and labeled. From the experiment described in Papers 3 and 4, the participant's comments from the test session were transcribed and analyzed. The total perceived workload was calculated for each participant based on the NASA-TLX data. A Wilcoxon signed rank test for two paired samples ($p < 0.05$) was used to analyze the quantitative data and find out whether there were any significant differences.

Participants

The participants for the experiment described in Paper 2 consisted mainly of students with no engineering background except for one. 21 participants (6 women and 15 men, mean age = 26.2, SD = 14.17) were recruited.

Participants for the experiments described in Papers 3 and 4 were mainly recruited from university students. 24 participants (9 women and 15 men, mean age = 24.5, SD = 5.43) participated in the device discovery part. 20 participants (9 women and 11 men, mean age = 23.8, SD = 5.06) participated in the device interaction part. The device interaction participants were a subset of the device discovery group (due to technical problems, four participants' data could not be used). The participants were mainly students with an engineering background.

Paper Summaries

The papers are briefly described in this section.

Paper 1: WozARd: A Wizard of Oz Tool for Mobile AR

This paper describes the Wizard of Oz tool called WozARd and presents the set of tools it offers. The WozARd device lets the test leader control the visual, tactile and auditive output that is presented to the test participant. Additionally, WozARd is suitable for using in an augmented reality environment where images are overlaid on the smartphone's camera view or on glasses.

The main features that were identified as necessary for simulating augmented reality functionality were: presentation of media such as images, video and sound; navigation and location based triggering; automatically taking photos; capability to log test results; notifications; and the integration of the Sony SmartWatch for interaction possibilities.

Paper 2: WozARd: A Wizard of Oz Method for Wearable Augmented Reality Interaction – A Pilot Study

This paper presents an initial investigation of the capability of the WozARd method to simulate a believable illusion of a real working AR city tour. Mainly aspects concerning the method itself were studied but also the limitations of current hardware were considered since they contribute to the participants' experience. A pilot study was carried out by collecting and analyzing qualitative and quantitative data from 21 participants who performed a predefined city tour using the WozARd on wearable technology. The data analysis focused on seven categories which potentially can have an impact on how the WozARd method is perceived by participants: precision, relevance, responsiveness, technical stability, visual fidelity, general user experience, and human operator performance. Overall, the results seem to indicate that the participants perceived the simulated AR city tour as a relatively realistic experience despite a certain degree of technical instability and human operator mistakes. Their subjective experience of the simulated AR city tour, as measured by the questionnaire, was overall positive and in general the city tour seemed to induce a feeling of a real, autonomous system rather than a system being controlled by someone else. The observation data seemed to confirm this. All participants managed to accomplish the AR city tour and in general they seemed to enjoy walking the simulated AR experience. Based on the experiences of this study, the authors believe that two of the most important factors contributing to these results are the design of the wizard device of the WozARd tool and the skill of the human operator. In conclusion, the WozARd method seemed to work reasonably well at least for this specific use case. In the present study only one specific use case for wearable AR was simulated. No real claims about the general usefulness of the WozARd method in a design process can therefore be made based on the presented data.

Paper 3: A Prototyping Method to Simulate Wearable Augmented Reality Interaction in a Virtual Environment – A Pilot Study

Building prototypes of such wearable AR systems can be difficult and costly, since it involves a number of different devices and systems with varying technological readiness level. The ideal prototyping method for this should offer high fidelity at a relatively low cost and the ability to simulate a wide range of wearable AR use cases.

This paper presents a proposed method, called IVAR (Immersive Virtual AR), for prototyping wearable AR interaction in a virtual environment (VE). IVAR was developed in an iterative design process that resulted in a testable setup in terms of hardware and software. Additionally, a basic pilot experiment was conducted to explore what it means to collect quantitative and qualitative data with the proposed prototyping method. The main contribution is that IVAR shows potential to become a useful wearable AR prototyping method, but that several challenges remain before meaningful data can be produced in controlled experiments. In particular, tracking technology needs to improve, both with regards to intrusiveness and precision.

Paper 4: Feasibility Study of Ubiquitous Interaction Concepts

This paper applies the IVAR method from paper 3 to evaluate the two concepts for device discovery and the two concepts for device interaction implemented in a virtual environment. The interaction concepts were compared in a controlled experiment.

Although statistically there were notable differences regarding how fast participants could finish their tasks, only small and moderate correlations were found between the task completion time and the perceived workload. This is probably due to task completion time being affected by aspects not covered by the six categories of the NASA-TLX. For the device discovery concepts, significant differences were found in perceived physical demand.

System limitations that may have affected the participants were the cables and equipment that the users had to wear as well as not being able to lean forward or backward.

Overall, the results indicate that the proposed interaction concepts were found natural and easy to use.

Discussion

In this section the strength and weaknesses of the prototyping methods and methodological issues are discussed.

The prototyping tool and method

This thesis has presented two prototyping methods, WozARd and IVAR, which can be used for exploring AR interaction. According to Liddle (1996), when designing and exploring UI one should distinguish between three different aspects: 1) graphical design, 2) interaction, and 3) conceptual model.

Graphical design, deals with what appears on the user's screen. Both WozARd and IVAR are suitable for prototyping and evaluating graphical design. The advantage of using WozARd for graphical design is that there is no need for recompiling the code when trying out new graphical user interfaces. However, since WozARd does not support tracking, IVAR is more suitable for graphical user interfaces that need to be correctly registered in a 3D space.

The second aspect, interaction, is about the control mechanism or the input method to control the commands. Interaction can be prototyped and evaluated with both WozARd and IVAR. WozARd offers more detailed interaction. It lets the user make small gestures on small areas such as the smartwatch display; it can simulate speech and gesture interaction but this requires a trained wizard who can interpret and react to user behavior and actions in a fast and correct manner. Since IVAR uses VR technology to simulate the environment in which participants test the interaction, the test cases can be run in a controlled manner without relying on a human operator. However, the devices used for input in IVAR were relatively cumbersome with several tracking and mobile devices attached to the user, resulting in a tangle of cables and straps. This probably had a negative effect on the perception of immersion and precision. An alternative setup could consist of Leap Motion's Dragonfly (Sixense, 2014) mounted at the front of the Oculus Rift DK2 (Oculus VR, 2014), which would reduce arm restrictions.

The third aspect, which is the most important component to design properly according to Liddle (1996), is the system's conceptual model. Everything else should be subordinated to making that model clear, obvious and substantial. IVAR is more suitable for prototyping and evaluating advanced conceptual models, as in Paper 3 where it is used to simulate the registration and tracking of virtual cards such as text descriptions popping up in front of the TV in a smart living room. If the same scenario were prototyped with WozARd, there would be a problem with latency of the virtual cards since the wizard would need to carefully observe that the user was pointing at the TV and quickly try to press the correct button to show the correct virtual card; by then there is a risk that the user would have already moved to the next device.

Another important aspect is the role of prototyping in the design process. Its role is to facilitate the exploration of a design space and uncover relevant information about users and their work practices by giving more details than a sketch and being testable. Additionally, prototypes are used to communicate an idea between designers, engineers, managers and users. They also permit early evaluation since they can be tested in various ways, including traditional usability studies and informal user feedback throughout the design process. In the early stage of the design process, low fidelity tools are preferable such as paper sketches, pretotyping and bodystorming. Software prototypes are usually more effective in the later stages when the basic design strategy has been decided (Beaudouin-Lafon & Mackay, 2003).

Based on the research results, I believe that WozARd can be used closer to the front end of the design funnel, since as a designer you can sketch an idea, take a photo of the sketch and use it. In addition, it has the strengths of being flexible, mobile and able to add other form factors but it relies on the wizard and does not facilitate high fidelity AR prototyping due to the lack of tracking functionality. Furthermore, Carter et al. (2008) state that WOZ prototypes are excellent for early lab studies but do not scale to longitudinal deployment because of the labor commitment for human-in-the-loop systems. IVAR is suitable to use closer to the narrow part of the design funnel, since it requires more hands on to simulate an idea. On the other hand, IVAR can provide three dimensional illustrations of more complex devices and can simulate more complex scenarios and registration and tracking of virtual objects. However, I believe that you can get a higher sense of presence, closer feeling to reality if WozARd is used with a well-trained wizard, at least for less complicated systems.

Methodological issues

Two evaluations were conducted using WozARd and IVAR. WozARd was used for an outdoor AR city tour study and IVAR to simulate an indoor home environment.

The goal of the AR city tour pilot study was to perform an initial investigation of the capability of the WozARd method to simulate a believable illusion of a real working AR city tour. Mainly aspects concerning the method itself were studied but also the limitations of current hardware were considered since they contribute to the participants' experience. Based on the experiences of this study, the two most important factors contributing to the findings are the design of the wizard device of the WozARd tool and the skill of the wizard. The wizard device was designed to aid the wizard in controlling the events of the WOZ experience during the pilot study and to reduce the risk of wizard mistakes. However, one aspect that was not implemented prior to the study due to time constrains was the audio feedback to the wizard, that is, feedback indicating that the audio information was played on the test person's device and when it was finished. Because of this, the test scenario heavily relied on a skilled wizard who could not be replaced by another wizard at short notice.

The goal of the "home environment" study was to explore the possibility of using IVAR to prototype AR interaction concepts before any physical prototypes were built. The validity of a method based on participants' perceptions and actions inside a VE must be carefully considered. One could argue that the proposed method constitutes a sort of Russian nested doll effect with "a UI inside a UI." This raises the question: Are observed usability problems caused by the UI or by the VR technology, or by both? To validate the results of the interaction concepts developed with IVAR, we need to compare the results of Paper 4 with those from a real system. This has not yet been done, but plans have been made to build a similar setup in a real room with real devices to compare the results.

In both evaluation studies, methodological triangulation was used to increase the quality of the data. Triangulation refers to the investigation of a phenomenon from (at least) two different perspectives (Rogers et al., 2011). According to Rogers et al. (2011), there are four types of triangulation:

1) Triangulation of data

2) Investigator triangulation

3) Triangulation of theories

4) Methodological triangulation

As mentioned in our studies, we used methodological triangulation which means applying different data gathering techniques. Examples of methods which we used in conducting evaluations included observations, interviews, questionnaires, and think aloud.

Another aspect of the design of the conducted evaluations is the fact of having relatively young people in the studies. The participants were primarily students and male. Having a better mixture of gender and age are preferable to gain a wider range of users' thoughts on a potential future of using other form factors than smartphones. The results from the studies show that the systems seem to work for relatively young people but do not say anything on how they would work for older people or people who are less accustomed to new technologies. Furthermore, we are unable to say anything about how the systems would work for people with cognitive and motor limitations.

Further research

This thesis has focused on developing and evaluating prototyping methods: WozARd and IVAR. More experiments should be performed to further explore the methods. We would like to conduct a similar study such as the one described in Papers 3 and 4, using real devices, such as Google Glass, SmartWatch 3 and Xperia Tablet Z, instead of a VE to be able to compare the findings. Additionally, we would like to continue adding features to the WozARd tool such as tracking to be able to register AR objects correctly in a 3D space. We would also like to investigate the importance of the WozARd operator by letting other users run the test instead of having one dedicated wizard.

A natural next step is to apply the methods described in this thesis to develop and evaluate user interaction combining several modalities such as gaze tracking, gestures and speech to explore the areas of affective user experience and intrusiveness. Example of research questions that I would like to investigate include:

What subjective experiences do different interaction techniques give rise to?

How can context-aware functionality ensure that the users' attention resources are not overloaded?

References

5DT Data Glove 5 Ultra. (2014). Retrieved from http://www.5dt.com/products/pdataglove5u.html

Azuma, R. (1997). A survey of augmented reality. *Presence, 4*(August), 355–385. Retrieved from http://nzdis.otago.ac.nz/projects/projects/berlin/repository/revisions/22/raw/trunk/Master's Docs/Papers/A Survey of Augmented Reality.pdf

Baker, M., Hong, J., & Billinghurst, M. (2014). Wearable Computing from Jewels to Joules. *IEEE Pervasive Computing*, *4*, 20–22.

Barba, E., MacIntyre, B., & Mynatt, E. D. (2012). Here We Are! Where Are We? Locating Mixed Reality in The Age of the Smartphone. In *Proceedings of the IEEE* (Vol. 100, pp. 929–936). doi:10.1109/JPROC.2011.2182070

Beaudouin-Lafon, M., & Mackay, W. E. (2003). Prototyping Tools and Techniques. *Human Computer Interaction—Development Process*, 122–142.

Billinghurst, M., & Starner, T. (1999). Wearable devices: New Ways to Manage Information. *Computer*, *32*(January), 57–64.

Buxton, B. (2010). *Sketching User Experiences: Getting the Design Right and the Right Design: Getting the Design Right and the Right Design*. Morgan Kaufmann.

Byrne, C. (2010). Most augmented reality companies not doing augmented reality? Retrieved March 4, 2015, from http://venturebeat.com/2010/12/22/forrester-most-augmented-reality-companies-not-doing-augmented-reality/

Carter, S., Mankoff, J., Klemmer, S., & Matthews, T. (2008). Exiting the Cleanroom: On Ecological Validity and Ubiquitous Computing. *Human-Computer Interaction*, *23*(1), 47–99. doi:10.1080/07370020701851086

Davies, N., Landay, J., Hudson, S., & Schmidt, A. (2005). Guest Editors' Introduction: Rapid Prototyping for Ubiquitous Computing. *IEEE Pervasive Computing*, *4*(4), 15–17. doi:10.1109/MPRV.2005.78

Dow, S., MacIntyre, B., & Lee, J. (2005). Wizard of Oz support throughout an iterative design process. *Pervasive Computing, IEEE CS and IEEE ComSoc, 4*(4), 18–26. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1541964

Epson Moverio BT-200 Smart Glasses. (2014). Retrieved from http://www.epson.com/cgi-bin/Store/jsp/Product.do?sku=V11H560020

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215 – 251.

Feiner, S., MacIntyre, B., & Höllerer, T. (1997). A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. In *Proceedings of the International Symposium on Wearable Computers.* (pp. 74–81). Boston, MA.

Frey, J. H., & Fontana, A. (1994). Interviewing: the art of science. *Handbook of Qualitative Research*, 361 – 376.

Genaro Motti, V., & Caine, K. (2014). Understanding the wearability of head-mounted devices from a human-centered perspective. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers - ISWC '14* (pp. 83–86). New York, New York, USA: ACM Press. doi:10.1145/2634317.2634340

Gigante, M. A. (1993). Virtual reality: Enabling technologies. *Virtual Reality Systems*, 15–22.

Google Cardboard – Google. (2014). Retrieved March 4, 2015, from http://www.google.com/get/cardboard/

Google Glass. (2013). Retrieved from https://www.google.com/glass/start/

Hart, S. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Retrieved from http://pro.sagepub.com/content/50/9/904.short

Hermodsson, K. (2010). *Augmented Reality on the Web*. Retrieved from http://www.w3.org/2010/06/w3car/beyond_the_keyhole.pdf

Kelley, J. F. (1983). An empirical methodology for writing user-friendly natural language computer applications. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems - CHI '83* (pp. 193–196). New York, New York, USA: ACM Press. doi:10.1145/800045.801609

Liddle, D. (1996). Bringing Design to Software Ch. 2 - Liddle. Retrieved January 10, 2015, from http://hci.stanford.edu/publications/bds/2-liddle.html

Mann, S. (1998). Definition of “wearable computer” (Taken from Prof. Mann’s Keynote speech of 1998 International Conference on Wearable Computing). Retrieved January 10, 2015, from http://wearcam.org/wearcompdef.html

Mann, S. (2014). Wearable Computing. *The Encyclopedia of Human-Computer Interaction, 2nd Ed.* Retrieved from https://www.interaction-design.org/encyclopedia/wearable_computing.html

Mattsson, S., & Alvtegen, C. (2014). Communicating beyond the word - designing a wearable computing device for Generation Z. Retrieved from https://lup.lub.lu.se/student-papers/search/publication/4451017

Meta Pro. (2014). Retrieved from https://www.spaceglasses.com/

Microsoft HoloLens. (2015). Retrieved March 4, 2015, from http://www.microsoft.com/microsoft-hololens/en-us

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE TRANSACTIONS on Information and Systems*, *77*(12), 1321–1329.

Misfit shine bloom necklace. (2015). Retrieved from http://bionicly.wpengine.netdna-cdn.com/wp-content/uploads/2014/11/misfit-shine-bloom-necklace.jpeg

Oculus Rift - Virtual Reality Headset for Immersive 3D Gaming. (2014). Retrieved from http://www.oculusvr.com/rift/

Oculus VR, I. (2014). The All New Oculus Rift Development Kit 2 (DK2) Virtual Reality Headset. Retrieved January 1, 2015, from http://www.oculusvr.com/dk2/

Oulasvirta, A., Kurvinen, E., & Kankainen, T. (2003). Understanding contexts by being there: case studies in bodystorming. *Personal and Ubiquitous Computing*, *7*(2), 125–134. doi:10.1007/s00779-003-0238-7

PalmPilot wooden model. (1995). Retrieved December 5, 2014, from http://www.computerhistory.org/revolution/mobile-computing/18/321/1648

Razer Hydra | Sixense. (2014). Retrieved from http://sixense.com/hardware/razerhydra

Recon Jet. (2014). Retrieved from http://www.reconinstruments.com/products/jet/

Rekimoto, J., Ayatsuka, Y., & Hayashi, K. (1998). Augment-able reality: situated communication through physical and digital spaces. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)* (pp. 68–75). IEEE Comput. Soc. doi:10.1109/ISWC.1998.729531

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction Design - beyond human-computer interaction* (Third Edit). A John Wiley and Sons, Ltd, Publication.

Samsung Gear VR. (2014). Retrieved March 4, 2015, from http://www.samsung.com/global/microsite/gearvr/gearvr_features.html

Savoi, A. (2011). *Pretotype It-Make sure you are building the right it before you build it right*.

Sixense. (2014). Leap Motion Sets a Course for VR. Retrieved January 1, 2015, from http://blog.leapmotion.com/leap-motion-sets-a-course-for-vr/

Slater, M. (1998). Measuring Presence: A Response to the Witmer and Singer Questionnaire. *Presence: Teleoperators and Virtual Environments*, *8*(5), 560–566.

Sony Morpheus. (2014). Retrieved May 14, 2014, from http://www.sony.com/SCA/company-news/press-releases/sony-computer-entertainment-america-inc/2014/sony-computer-entertainment-announces-project-morp.shtml

Sony SmartEyeGlass. (2015). Retrieved March 4, 2015, from https://developer.sony.com/devices/mobile-accessories/smarteyeglass/

Sony SmartWatch 3. (2015). Retrieved from http://www.sonymobile.com/global-en/products/smartwear/smartwatch-3-swr50/

Sony Xperia Tablet Z. (2013). Retrieved from http://www.sonymobile.com/se/products/tablets/xperia-tablet-z/

Stanney, M. K. (Ed.). (2002). *Handbook of Virtual Environments - Design, Implementation, and Applications*. London: Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS '68 (Fall, part I)* (pp. 757 – 764). New York, New York, USA: ACM Press. doi:10.1145/1476589.1476686

System Usability Scale (SUS). (2013, September 6). Retrieved October 22, 2014, from http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

Unity - Game Engine. (2014). Retrieved January 1, 2015, from http://unity3d.com/

Welker, S. (2013). OpenDive. Retrieved from http://www.durovis.com/opendive.html

Vuzix M100. (2014). Retrieved from http://www.vuzix.com/consumer/products_m100/

Yildirim, A. (2014). OpenVR. Retrieved October 13, 2014, from http://mclightning.com/

WozARd: A Wizard of Oz Tool for Mobile AR

Alce, G., Hermodsson, K. and Wallergård, M. (2013).

I

Proceedings of the 15th International Conference on Human-Computer Interaction
with Mobile Devices and Services – MobileHCI '13.
(pp. 600-605). ISBN: 9781450322737.

WozARd: A Wizard of Oz Tool for Mobile AR

Günter Alce
Sony Mobile Communications
Nya Vattentornet
221 88 Lund, Sweden
Gunter.alce@sonymobile.com

Klas Hermodsson
Sony Mobile Communications
Nya Vattentornet
221 88 Lund, Sweden
Klas.Hermodsson@sonymobile.com

Mattias Wallergård
Lund University
P.O. Box 118
221 00 Lund, Sweden
Mattias.Wallergard@design.lth.se

MobileHCI 2013, Aug 27 – 30, 2013, Munich, Germany.
ACM 978-1-4503-2273-7/13/08.

Abstract

Wizard of Oz methodology is useful when conducting user studies of a system that is in early development. It is essential to be able to simulate part of the system and to collect feedback from potential users. Using a human to act as the system is one way to do this.

The Wizard of Oz tool presented here is called WozARd and it aims at offering a set of tools that help the test leader control the visual, tactile and auditive output that is presented to the test participant. Additionally, it is suitable for using in an augmented reality environment where images are overlaid on the phone's camera view or on glasses.

The main features that were identified as necessary include presentation of media such as images, video and sound, navigation and location based triggering, automatically taking photos, capability to log test results and visual feedback, and the integration of Sony SmartWatch for interaction possibilities.

Author Keywords

Augmented Reality, Wizard of Oz, multimodal interaction, multimodal user interface

ACM Classification Keywords

H.5.2 User Interfaces: Evaluation/methodology, Graphical user interfaces (GUI), Prototyping, User-centered design, Voice I/O; D.2.2 Design Tools and Techniques: User interfaces;

General Terms

Human Factors, Experimentation

Introduction

When conducting user studies of a system that is in early development, it is essential to be able to quickly simulate missing “parts” of the system and to collect valuable feedback from potential users. One way of doing this is with Wizard of Oz (WOZ) testing, which is a well-known method in human computer interaction used for testing a non-complete system on users initially developed by J.F. Kelley in 1983 to simulate a natural language application [3]. Additionally, WOZ testing is a powerful tool for uncovering design ideas in limited evolved technologies, especially for systems performing in physical environments, since the designers are less constrained by technical specifications [6]. S. Dow et al. state that WOZ testing is "leading to more frequent testing of design ideas and, hopefully, to better end-user experiences" [2].

WOZ methodology has been used in a variety of studies to explore design concepts. Most notably, simulating speech recognition systems [1, 3, 4, 9] was an early application area. There are also WOZ tools that focus on Augmented Realty (AR) user interfaces [5, 8]. However, they tend to focus on creating 3D content. Furthermore, they might not provide a mobile set-up. One mobile WOZ tool that works on Android devices is presented in [7] but is not suitable for exploring AR user interfaces. Li et al. [6] have developed *Topiary* which let users interact with the user interface mockup while a “wizard” follows them and updates the locations. However, *Topiary* focuses specifically on location-based services and is not suitable for AR environments. Furthermore, it is based on rather outdated devices.

A WOZ tool that works on mobile phones (based on e.g. Android) which is easy to control and can simulate visual, tactile and auditive AR user interfaces, is currently missing. Another important feature of such a tool is that it helps researchers and designers to plan and prepare user studies, assist during execution of the user study and to analyze the data afterwards.

This paper presents our attempts to develop such a WOZ tool called WozARd.

WozARd

WozARd consists of two Android devices communicating with each other wirelessly (Figure 1).

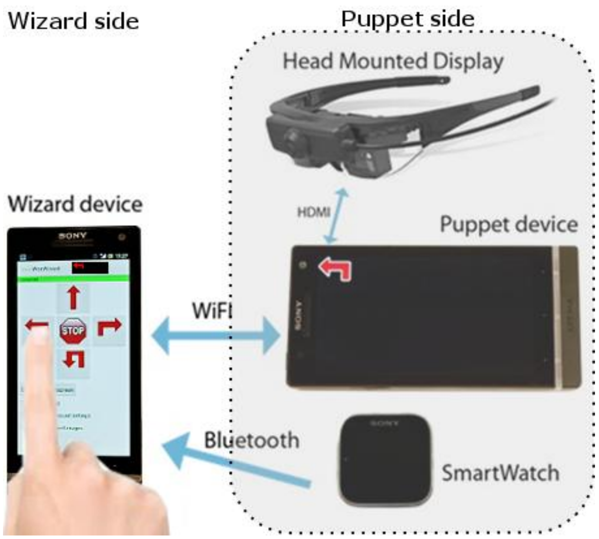


Figure 1: Shows an example of WozARd setup. In this example the wizard use a Sony Xperia S phone while the test person use Vuzix Star 1200, Sony Smartwatch and Sony Xperia S phone. The devices are not in correct scale.

One device acts as the wizard controlled by the test leader and the other one is the puppet used by test person. The wizard application can control the user interface on the puppet device. Main features which were identified as necessary include:

- present media such as images, video and sound
- navigation and location based triggering
- offer features to plan and prepare for user studies
- capability to log test and visual feedback
- integrate Sony SmartWatch [10] for interaction possibilities

The wizard

The wizard contains most of the functionalities and can be used to control what is shown on the puppet device. There are several views customized for different scenarios including notification, navigation and camera view. In an initial pilot test, the importance of feedback to the wizard emerged. The wizard must always know what is shown on the puppet device. The connection is indicated with a green bar and a small thumbnail image indicating what is shown on the puppet side. It is easy to add content and create lists of notifications without recompiling the application. The features that the wizard offers are filebrowser, notifications, navigation, camera, tour and logging.

FILEBROWSER

In the filebrowser view, it is possible to list all files that can be of interest to show or play to the test person on the puppet device (Figure 2).

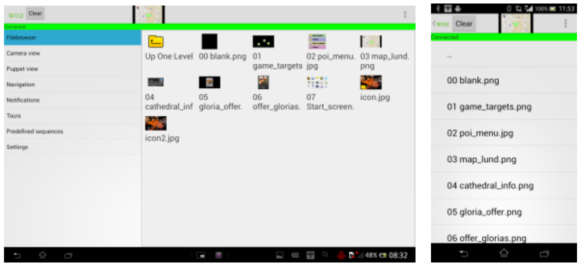


Figure 2: Example of how filebrowser view appears on the wizard side of a Sony Xperia tablet Z and Sony Xperia TL phone.

Camera

From camera view (Figure 3) the wizard can start the camera on the puppet device and also request the puppet device to send the camera feed from the front- or back camera of the phone or tablet. It is also possible to set the puppet to take pictures automatically within a certain time interval. This data could facilitate analysis and could also be used when debriefing the subject. It should be noted that it is not necessary to display the camera view on the puppet device. It can be started in the background without the test person knowing it.

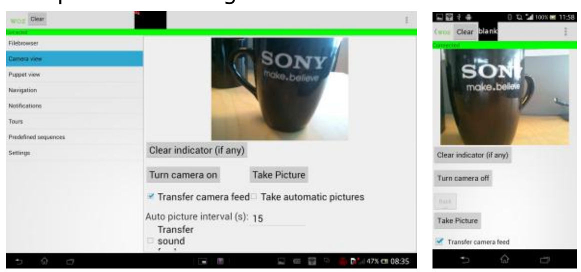


Figure 3: Example of how camera view appears on the wizard side of a Sony Xperia tablet Z and Sony Xperia TL phone.

Navigation

The purpose of this feature is to simulate navigation. It is also possible for the wizard to enable/disable audio navigation together with the visual navigation (Figure 4). Additionally, it is possible to customize the audio navigation since the Android Text-to-Speech engine is used.

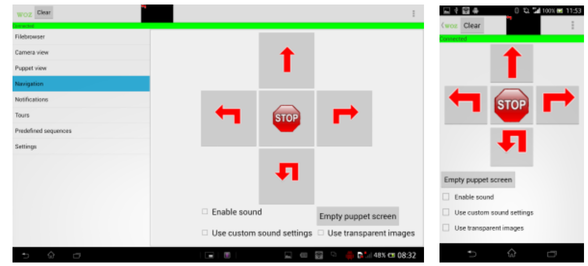


Figure 4: Example of how navigation view appears on the wizard side of a Sony Xperia tablet Z and Sony Xperia TL phone.

Notifications

The wizard can choose to send default notifications but can also create and send different simulated notifications on the fly such as Gmail, Alarm, SMS, etc. Type of notification and type of icon can be dynamically changed through an xml file. It is also possible to have the notification messages read aloud on the puppet device.

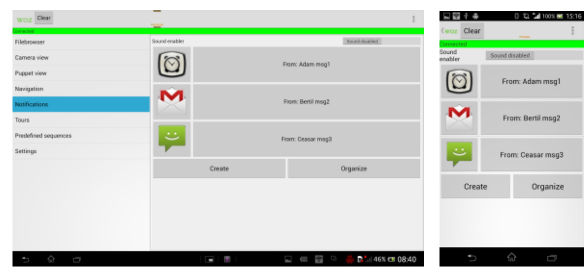


Figure 5: Example of how notification view appears on the wizard side of a Sony Xperia tablet Z and Sony Xperia TL phone.

Tours

The wizard can also trigger different actions at different locations in order to offer the opportunity to create a tour with WozARd. The wizard can walk around and choose to set different actions, for example showing an image in different locations presented on a map (Figure 6). The wizard can also start the tour, which will be running in the background, and still be able to use the other features.

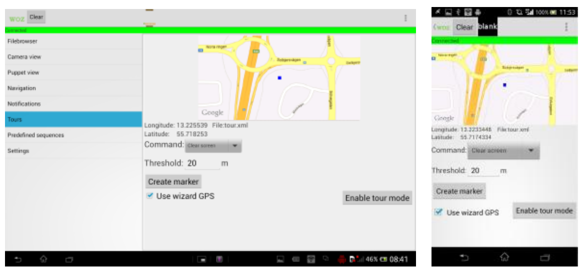


Figure 6: Example of how tour view appears on the wizard side of a Sony Xperia tablet Z and Sony Xperia TL phone.

PREDEFINED SEQUENCE

This feature helps to predefine a user study. The idea is to list all commands and simply click through without needing to switch views.

LOG

Both the wizard and puppet applications have support for logging the activities. The logs are saved on the SD-card. All entries have a timestamp and if the position is known, latitude and longitude are also saved. This is essential for analyzing the data collected.

The puppet

The user interface of the puppet is designed not to have too much point and click interaction, other than the SmartWatch, which is used as an input device for selecting, dismissing and scrolling notifications.

The puppet has a black background since black is transparent when using HMDs with optical see-through. There is also support to show navigation arrows on a video see-through device (Figure 7).

Figure 7: Example of how an overlaying navigation arrow on a mobile video-see through device is presented on the puppet side.

Conclusions and Future work

The purpose of WozARd is to make it possible to prototype mid-fi user interfaces for AR on mobile phones, tablets and glasses. WozARd already supports many useful features but can of course not simulate all sorts of use cases. For example, one obvious drawback of WozARd is its inability to simulate scenarios that involve real-time tracking. In its current form WozARd is probably best suited for testing scenarios limited to visualization of 2D data, tactile feedback and audititive feedback. However, it is our intention to release WozARd as open-source to make it possible for others in the AR community to expand the features and tweak it to fit their own requirements. For our part, we will continue to develop WozARd and research its usefulness as an AR prototyping tool. Much effort will be put on making the user interface of the wizard part of WozARd as intuitive and fast as possible in order to minimize the risk for wizard errors, which in worst case might render a whole user test session useless. Another important research question concerns whether the ecological validity of the proposed tool is good enough to produce reliable user test data.

Acknowledgements

The authors would like to extend their gratitude to Per Sörbris and Mikael Möller who developed the first version of WozARd in their master thesis. This research is partially being funded by the European 7th Framework Program, under grant VENTURI (FP7-288238) and the VINNOVA funded Industrial Excellence Center EASE (Mobile Heights). Special thanks go to Eileen Deaner for linguistic support.

References

[1] Dahlbäck, N., Jönsson, A., Ahrenberg, L., Wizard of Oz Studies – Why and How. In *Proc. of Intelligent User Interfaces '93*. 1993, 193-200.

[2] Dow, S., Lee, J., Oezbek, C., MacIntyre, B., Bolter, J. D., Gandy, M., Wizard of Oz Interfaces for Mixed Reality Applications, *CHI 2005,* ACM April 2005.

[3] Kelley, J.F., An empirical methodology for writing user-friendly natural language computer applications, In *Proc. of the SIGCHI Conference of Human Factors in Computing Systems,* 1983, 193-196.

[4] Klemmer, S., Sinha A., Chen J., Landay J., Aboobaker N., Wag A, "SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces", In *ACM Symp. On User Interface Software and Technology (UIST'00)*, 2000, 1-10.

[5] Lee, M., Billinghurst M., A Wizard of Oz Study for an AR Multimodal Interface, In *Proc.* of *the 10th International Conference on Multimodal Interfaces, ICMI'08*, 2008, 249-256.

[6] Li, Y., Hong, J. I., Landay, J. A., Design Challenges and Principles for Wizard of Oz Testing of Location-Enhanced Applications, *IEEE Pervasive computing,* 2007, 70-75.

[7] Linnell, N., Bareiss, R., Pantic, K., A Wizard of Oz Tool for Android, MobileHCI'12, 2012.

[8] MacIntyre, B., Gandy, M., Dow, S., Bolter, J.D., DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences, In *ACM Symp. On User Interface Software and Technology (UIST'04),* 2004, 197-206.

[9] Maulsby, D, Greenberg, S., Mander, R., Prototyping an intelligent agent through Wizard of Oz, In *Proc. CHI,* 1993, 277-284.

[10] Sony Mobile Communications. Sony SmartWatch. http://www.sonymobile.com/gb/products/accessories/smartwatch/. [Online] [Accessed 24 April 2013].

Paper 2: WozARd: A Wizard of Oz Method for Wearable Augmented Reality Interaction – A Pilot Study

Alce, G., Wallergård, M. and Hermodsson, K. (2014).

Resubmitted after minor revision to the journal of
Advances in Human-Computer Interaction,
Hindawi Publishing Corporation.

II

WozARd: A Wizard of Oz Method for wearable Augmented Reality Interaction – A Pilot Study

Abstract

Head-mounted displays (HMDs) and other wearable devices open up for innovative types of interaction for wearable augmented reality (AR). However, to design and evaluate these new types of AR user interfaces, it is essential to be able to quickly simulate undeveloped components of the system and collect valuable feedback from potential users early on in the design process. One way of doing this is the so called wizard of oz (WOZ) method. The basic idea behind WOZ is to create the illusion of a working system by having a human operator, hidden somewhere "behind the screen", performing some or all of the system's functions. WozARd is a WOZ method developed specifically for wearable AR interaction. The goal of the presented pilot study was to perform an initial investigation of the capability of the WozARd method to simulate a believable illusion of a real working AR city tour. Mainly aspects concerning the method itself were studied but also the limitations of current hardware were considered since they contribute to the participants' experience. Qualitative and quantitative data was collected when 21 participants performed an AR city tour simulated with the WozARd method. The data analysis focused on seven categories that can have a potential impact on how the WozARd method is perceived by participants: precision, relevance, responsiveness, technical stability, visual fidelity, general user experience, and human operator performance. Overall, the results seem to indicate that the participants perceived the simulated AR city tour as a relatively realistic experience despite a certain degree of technical instability and human operator mistakes.

Keywords:

Wizard of Oz, Augmented Reality, Wearable Technology, Evaluation Methods, Prototyping, Interaction Design

Introduction

The age of wearable devices is upon us and they are available in many different form factors including head-mounted displays (HMDs), smartwatches and smartbands [1]. Wearable devices are intended to always be ‘‘on’’, always acting, and always sensing the surrounding environment to offer a better interface to the real world [2]. Taking into account recent advances in wearable devices, we can expect that people will be able to carry their wearables at all time. One example of a wearable form factor that follows this trend are HMDs. HMDs have been developed and used in research since the 1960s [3], but it is not until recently that they have become available outside of the research lab. Examples of HMDs or glasses that are available are *Google Glass* [4], *Meta-pro* [5], *Recon Jet* [6], *Vuzix M100* [7] and *Epson Moverio BT-200* [8].

The HMD form factor facilitates Augmented Reality (AR), a technology that mixes virtual content with the users' view of the world around them [9]. Azuma [10] defines AR as having three characteristics: 1) Combines real and virtual, 2) interactive in real time and 3) registered in 3-D. According to Narzt et al. [11] the AR paradigm opens innovative interaction facilities to users: human natural familiarity with the physical environment and physical objects defines the basic principles for exchanging data between the virtual and the real world, thus allowing gestures, body language, movement, gaze and physical awareness to trigger events in the AR space. However, it is difficult and time consuming to prototype and evaluate this new design space due to components that are undeveloped or not sufficiently advanced [12]. To overcome this dilemma and focus on the design and evaluation of new user interfaces (UIs) instead, it is essential to be able to quickly simulate undeveloped components of the system in order to enable the collection of valuable feedback from potential users early on in the design process. One way of doing this is with the Wizard of Oz (WOZ) method. The basic idea behind WOZ is to create the illusion of a working system. The person using it is unaware that some or all of the system's functions are actually being performed by a human operator, hidden somewhere "behind the screen". This allows testing interaction concepts before a system is fully working. The method was initially developed by J.F. Kelley in 1983 to simulate a natural language application [13].

The WOZ method has been used in a variety of studies to explore design concepts for interactive systems. An early application area was simulating speech recognition systems [14].

Another application area in which it is suitable to use the WOZ method includes AR [15]. A WOZ tool called *DART* [16] enables designers to design AR UIs and to integrate live video, tracking technology and other sensor data. Lee and Billinghurst [17] used WOZ to study multimodal AR interfaces but only in an indoor static setup.

Although WOZ has been used for a long time and in various application areas, there is still no WOZ tool known by the authors that can be used to prototype AR UIs that work in both indoor and outdoor environments, and that can be used with HMDs and other wearable devices integrated with a mobile phone (e.g. based on Android) for mobility.

The authors have developed a WOZ tool called *WozARd* in an attempt to meet these requirements. The set of features that *WozARd* offers is described in more details in [18]. With WozARd it is possible to control what is shown and played on the user's HMD and/or smartphone or tablet. WozARd lets the user interact with the system through a smartwatch. The human operator can easily change the UI without reprogramming the application, which makes WozARd flexible and easy to use for non-programmers.

One important aspect when using the WOZ method is to ascertain that the participants' behavior in the simulated system is reasonably similar to that in the corresponding real system [14]. The extent to which a study comprises "real-world" use of a system is called "ecological validity" [19]. Another term which is closely related to ecological validity is "external validity", which means the extent to which the results of a study can be generalized to other situations [20].

For example, low fidelity prototyping such as paper prototyping has a low ecological validity but it can be very effective in testing issues of aesthetics and standard graphical UI. In other words, by using low fidelity prototyping with low ecological validity, it is still possible to achieve high external validity. However, to do so when designing for an eco-system of wearable devices, a richer ecological validity is often required [19].

The context in which WozARd was developed was the three year European project VENTURI [21]. The project's first year focused on AR gaming, the second year was about supporting visually impaired people and the third year had AR city tours as theme. The goal of the third year was to deliver an AR application that let people experience a city's cultural heritage, through their own smartphones and/or tablets. Part of the objective was also to allow participants to experience parts of the city tour with HMD. For this reason WozARd was developed and used within VENTURI to explore fundamental design issues connected to AR city tours early on in the project.

Furthermore, AR navigation systems such as "The Touring machine" developed by Steven Feiner et al. [22], Narzt et al. [11] and Bolter et al. [23] research was used as inspiration for this study.

The goal of the presented pilot study was to perform an initial investigation of the capability of the WozARd method to simulate a believable illusion of a real working AR city tour. Mainly aspects concerning the method itself were studied but also the limitations of current hardware were considered since they contribute to the participants' experience.

The study presented was carried out by collecting and analyzing qualitative and quantitative data from 21 participants who performed a predefined city tour using WozARd on wearable devices. The data analysis focused on six categories that are believed to have a potential impact on how the WozARd method is perceived by participants: precision, relevance, responsiveness, technical stability, visual fidelity, and general user experience.

The next section presents relevant related work. Then the WozARd tool is described followed by a presentation of the method, results, discussion, conclusions and future work.

Related work

As mentioned, WOZ is a well-known method where a human operates undeveloped components of a technical system. Above all, the WOZ method has been widely used in the field of human-computer interaction to explore design concepts. WOZ testing is a powerful method for uncovering design ideas in limited evolved components, especially for systems performing in physical environments, since the designers are less constrained by technical specifications [24]. Dow et al. [25] state that WOZ testing is "leading to more frequent testing of design ideas and, hopefully, to better end-user experiences."

An early application area of WOZ was in speech recognition systems [14]. To simulate both input and output language technology components, Schlögl et al. developed an open source tool called *WebWOZ* [26] that uses an internet based WOZ framework.

The WOZ method has also been used to combine speech and gestures to control a robot by speech and gestural interaction [27]. Two human wizards were used in the evaluation, one responsible for the dialogue and the other for the robot navigation. Other gesture based WOZ studies include [28].

Other examples of research tools that used the WOZ method include *ConWIZ* [29], which is a WOZ tool with a mobile application that is capable of controlling the simulation of a WOZ prototype as well as contextual objects such as fans and lights. Fleury, Pedersen, and Bo Larsen [30] used a WOZ setup to evaluate four different methods for transferring video content from a smartphone to a TV screen. Li et al. [24] developed *Topiary* which lets users interact with the UI mockup while a human wizard follows them and updates the locations.

As already shown, there are several WOZ tools available for different use cases. However, none of them fulfill the requirements of being flexible, mobile, able to add other form factors, and able to explore AR interaction. Some of the listed WOZ tools are flexible but not mobile [26]. Example of mobile WOZ tools include *ConWIZ* [29], Linnel et al.'s tool [31] and *Topiary* [24] but they do not support integration of other form factors nor can they be used for exploring AR. In most of the studies, the human operator's role is stationed in a control room hidden from the participants. However, with mobile tools such as *ConWIZ* [29], Linnel et al.'s tool [31] and *Topiary* [24] the wizard was able to follow the participants and update the UI accordingly, but it is not clear whether the participants knew that the system was controlled by the human operator or not.

None of the mentioned WOZ tools, however, fulfilled the requirements which were needed to perform studies for the VENTURI project. Examples of requirements for the VENTURI project included:

- Not focus on one form factor.
- Be useable both indoors and outdoors.
- Aid the human operator when adding scenarios on the fly.
- Support the easy adding of other form factors.
- Be suitable for prototyping AR.

In past research by the authors, the WOZ tool WozARd [18] was developed in an attempt to meet these requirements.

The WozARd tool

This section introduces the WozARd WOZ tool, a more detailed description of the tool can be found in [18]. First, an overview is presented of how the tool works, followed by examples of features that the tool supports.

WozARd consists of two Android devices that communicate with each other wirelessly (Figure 1).

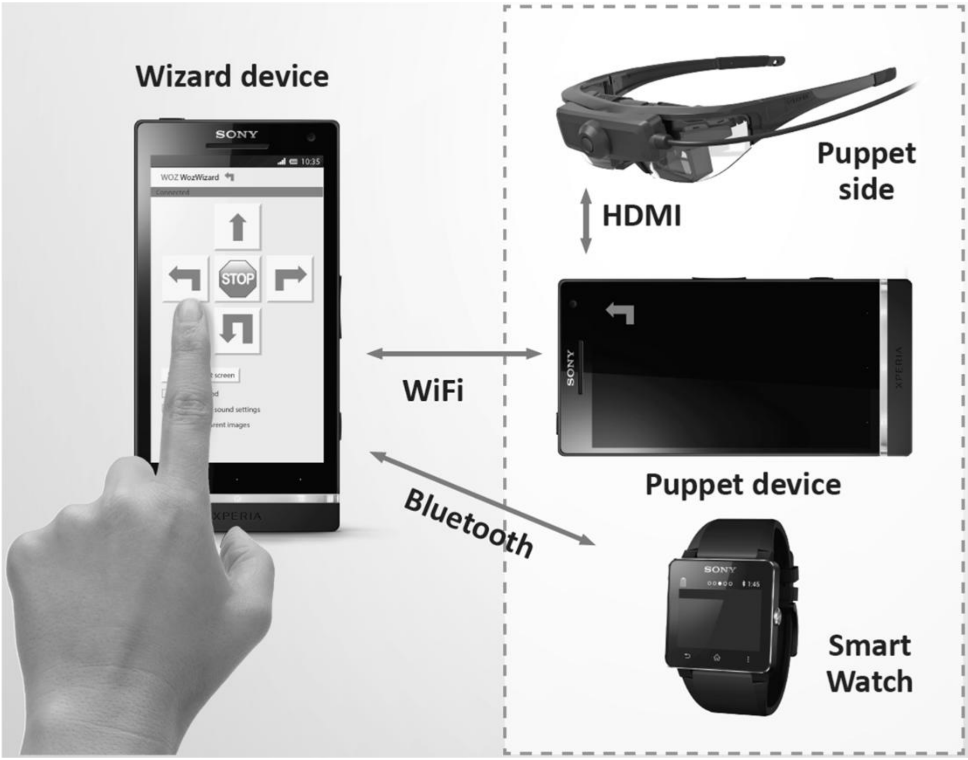


Figure 1. The WozARd architectural setup.

On the left is the wizard device which is controlled by the human operator and on the right within the dashed lines is the devices used by the participant (Figure 1). Through the WozARd wizard application, the human operator can control the participant's UI by pressing the buttons in the application. Examples of features that WozARd is suitable for are:

- Presentation of media such as images, video and sound (Figure 2a).
- Navigation and location based triggering (Figure 2b).
- Showing notifications (Figure 2c).
- Features to plan and prepare for user studies.
- Capability to log test and visual feedback.
- Being able to work with both tablet and phone form factors.
- Integrating the Sony *SmartWatch* [32] and the Sony *SmartWatch 2* [33] for interaction possibilities.
- Adding, HMDs, which can be connected through HDMI, e.g. Vuzix *Star 1200* [34].
- Adding, HMDs, which runs on Android, for example *Epson Moverio BT-200* [8], *Vuzix M100* [7] and *Google Glass* [4].

Figure 2. The wizard device UI: a) list of playable multimedia, b) navigation, and c) notifications.

The only type of interaction that the participant can perform is touch gestures on a Sony *SmartWatch*, which catch the gesture performed by the participant and sends it through the Bluetooth connection to the wizard device. Of course, other interaction types based on e.g. voice and mid-air hand gestures could be simulated as long as the human operator can hear and see the participant properly and interpret his/her intentions correctly. Figure 3 show what a participant sees through a video see-through display when the human operator pushes the turn right button.

Figure 3. The participant's view through a video see-through display, when the human operator pushes the turn right button.

Method

This section describes the setup of the pilot study.

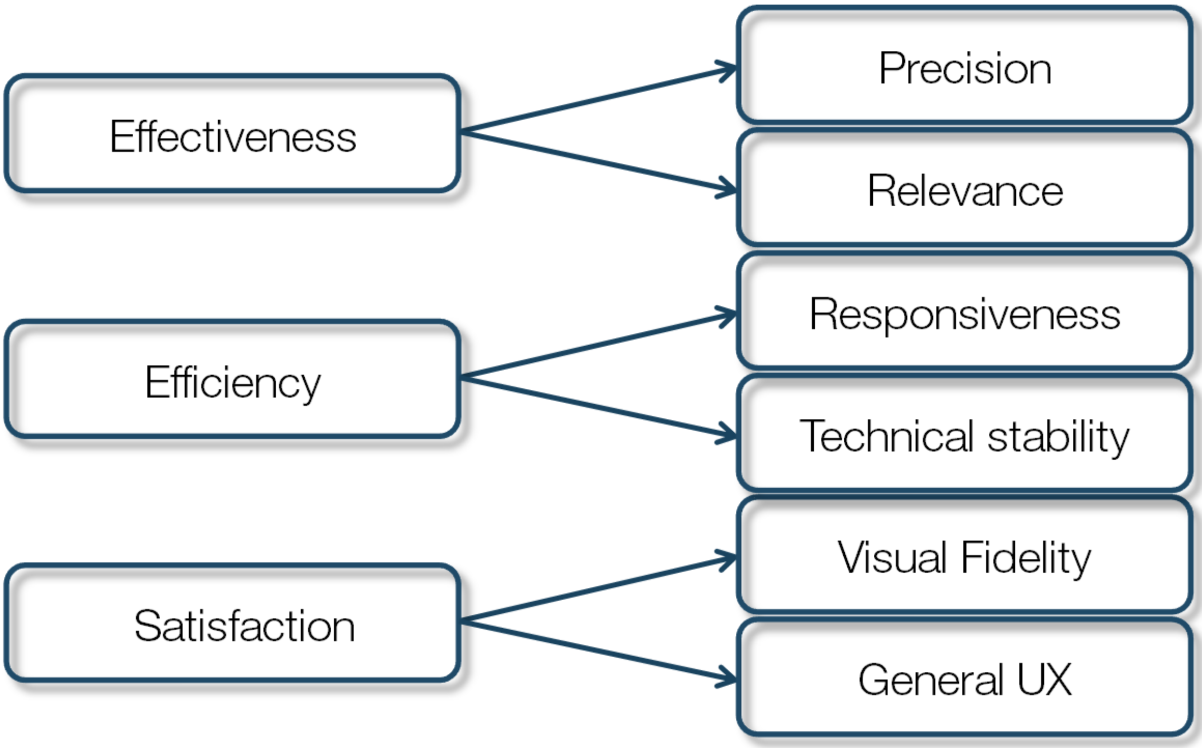


Figure 4. Six categories emerged from the ISO 9241-210.

The approach to this pilot study was to first define categories that can have a potential impact on how the WozARd method was perceived by participants. The ISO definition of usability [35], which includes effectiveness, efficiency and satisfaction was used as starting point. Each of the three usability categories was subcategorized resulting in a total of six categories (Figure 4):

- Precision: Is the augmented information shown at the right time and place?
- Relevance: Is relevant information shown at the right time and place?
- Responsiveness: How quickly does the system respond to user input?
- Technical stability: Did the user notice any technical difficulties?
- Visual fidelity: What fidelity does the visual input have? Since the WozARd does not currently support tracking, it is not possible to impose virtual content correctly registered in the 3D space. Instead, the image is "hanging" in front of the user (i.e. when the user turns his/her head, the image follows the head movement).
- General user experience: What is the general user experience of the WozARd method including the ability to hear and read the augmented information?

Nine pilot experiments were conducted iteratively, which resulted in continuous improvements of the tool and the experimental setup.

The AR city tour took place in a small city in southern Sweden called Trelleborg. The tour was based on a predefined route. All information and images were collected prior to the study and included different types of urban environments and target objects. The information that the participants experienced contained an image and audio, mainly text to speech. Examples of participant experiences included historical information, informative notifications (Figure 5a), lunch specials at restaurants, tourist attractions (Figure 5b) and sculptures (Figure 5c). Participants had to interact with a Sony *SmartWatch* [32] to start the city tour, to continue the city tour and to remove notifications.

a) "Tab to start the city tour."

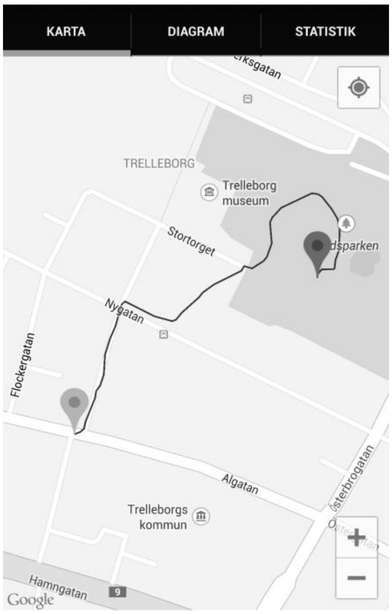
b) "Böst is a fountain made in bronze …"

c) "These creatures are made by Ralf Borselius who is a famous sculptor."

Figure 5. Three samples of what was shown during the city tour.

The tour was designed to let the participants walk approximately 500 m (Figure 6a). The average time to walk the city tour was eight minutes.

a) b)



c)

Figure 6. a) a map of the route, b) the human operator on the left guiding one of the participants, c) a participant interacting with the system.

Materials

Equipment used during the pilot study included:

- HMD, Vuzix *Star 1200* [34] connected to Sony *Xperia S* [36].
- Sony *Xperia S* [36] used as puppet device.
- Sony *Xperia Z* [37] used as wizard device.
- Sony *SmartWatch* [32] used by the participants to interact with the system.
- Sony *Handycam HDR-CX190* [38] to record during the user study.

Participants

21 participants (6 women and 15 men), mainly students, were recruited for the study. The average age was 26.2 years (*SD* = 14.2). The participants reported that they used computers or tablets 3.67 hours per day (*SD* = 2.92), and smartphone on average 4.98 hours per day (*SD* = 3.61).

Procedure

The sessions involved a participant; the human operator who simulated the AR city tour with the WozARd wizard device and managed the experiment; and a test assistant who walked along the participant and video recorded each session for data capture and to measure elapsed time (Figure 7). The session started with the participant signing an informed consent form and filling out a background questionnaire. The questionnaire included participant age, gender and occupation. Next, a short introduction of AR was given by describing Azuma's definition [10], followed by instructions on how to interact with the system. The participants were also asked to follow the instructions from the system and to think aloud while walking the city tour. Think aloud is one of the most direct and widely used methods to gain information about participants' internal states [39]. Using the think aloud method had two purposes: to gain information on the participants' experience when attending to the information and to aid the human operator during the city tour in understanding if the participants were experiencing any problems. In addition, participants were informed that the human operator would walk behind them taking notes.

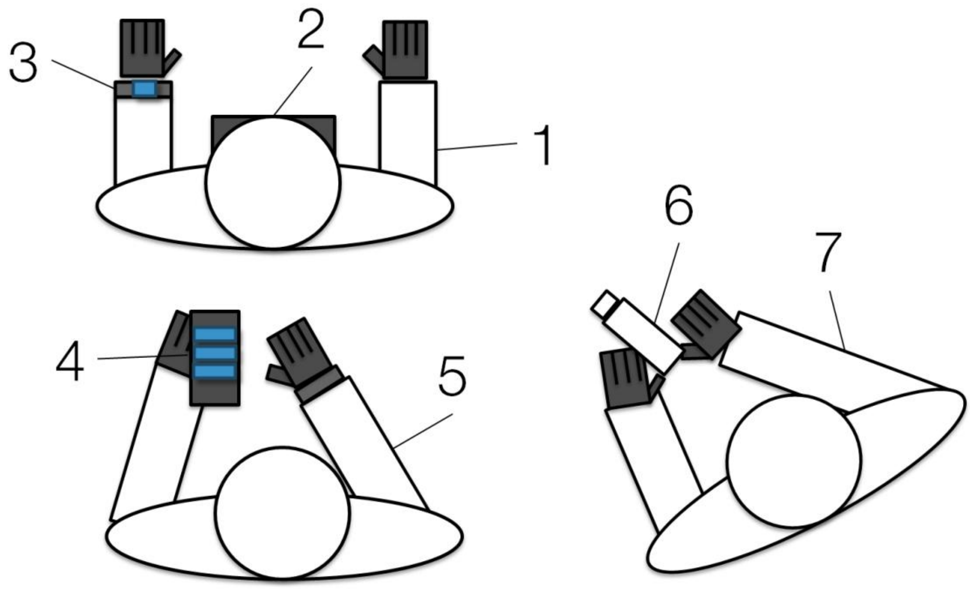


Figure 7, A conceptual setup of the study. 1 Participant, 2 Vuzix Star 1200 [34], *3 Sony Smartwatch* [32], *4 Sony Xperia Z* [37] *WozARd wizard side, 5 Human operator, 6 Sony Handycam HDR-CX190*[38], *7 Test assistant.*

All participants filled in a questionnaire after the tour. It contained fifteen statements inspired by the System Usability Scale (SUS) [40] to which the participant agreed or disagreed on a five-point Likert scale. The questionnaire was designed to target the six categories: precision, relevance, responsiveness, technical stability, visual fidelity, and general user experience (Table 1). Each session lasted about 30 min.

Table 1. Questionnaire statements.

Number	Statement
S1:	The system was responsive.
S2:	The system was precise considering showing right picture on the right place.
S3:	The city tour was interrupted due to technical difficulties.
S4:	The system showed me relevant information at the right time and place.
S5:	The “hanging” images were annoying.
S6:	The system showed too much information too often.
S7:	I enjoyed walking the city tour.
S8:	The pace of the tour was appropriate.
S9:	The interaction with the watch was intuitive.
S10:	I would recommend the city tour to a friend.
S11:	I would like to use this system when I visit a new city.
S12:	It was hard to read the notifications.
S13:	It was hard to hear what the system said.
S14:	It felt like the system was working completely automatically.
S15:	It felt like someone was controlling the system.

The session was concluded with an informal, open interview to collect qualitative data. Each session was video recorded. Each participant's video recording was transcribed with individual quotes categorized and labeled. Furthermore, events of special interest were noted e.g. human operator induced errors. The answers from the five-point Likert scale questionnaire responses were given as numerical value from 1 to 5 (disagree = 1; agree = 5) for the statistical calculations of median (Mdn) and inter-quartile range (IQR), which is the distance between the 75th and 25th percentile.

Results

This section presents quantitative and qualitative data from the user pilot study. Overall, all of the 21 participants managed to accomplish the AR city tour and the majority of them showed signs of enjoying the AR experience. The data in the following is divided into the seven categories: precision, relevance, responsiveness, technical stability, visual fidelity, general user experience and human operator performance. The last category was not part of the original six categories that were hypothesized to have a potential impact on how the WozARd method is perceived by participants but emerged as a new category during the data analysis.

Since the distribution was not symmetric and an ordinal scale was used, the median was calculated for the questionnaire responses [41].The whiskers show the range of the data set, i.e. max/min value. The values of the statements are presented in Figure 8 and Table 2.

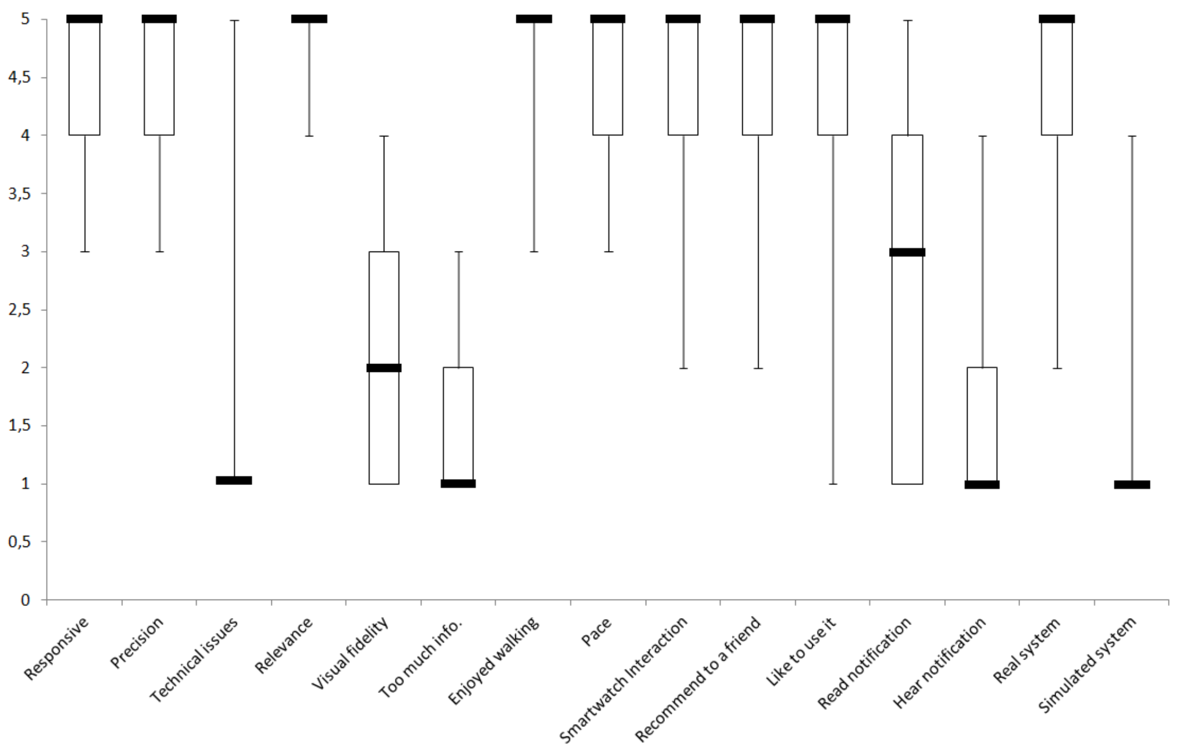


Figure 8. Median and the whiskers show the range of the data set.

Table 2. Median and min. and max. values.

Questionnaire statement	Median	Min.	Max.
S1: The system was responsive.	5	3	5
S2: The system was precise considering showing right picture on the right place.	5	3	5
S3: The city tour was interrupted due to technical difficulties.	1	1	5
S4: The system showed me relevant information on the right time and place.	5	4	5
S5: The "hanging" images were annoying.	2	1	4
S6: The system showed too much information too often.	1	1	3
S7: I enjoyed walking the city tour.	5	3	5
S8: The pace of the tour is appropriate.	5	3	5
S9: The interaction with the watch was intuitive.	5	2	5
S10: I would recommend the city tour to a friend.	5	2	5
S11: I would like to use this system when I visit a new city.	5	1	5
S12: It was hard to read the notifications.	3	1	5
S13: It was hard to hear what the system said.	1	1	4
S14: It felt like the system was working completely automatically.	5	2	5
S15: It felt like someone was controlling the system.	1	1	4

Precision

The majority of the participants thought that the system was precise; three participants neither agreed nor disagreed. The median value was 5 (IQR = 1).

Several participants made comments regarding precision. One participant, for example seemed impressed that "The system knew that I was close to the street and told me to stop and then continue when I had crossed the street." Another participant liked that the "Daily specials from the restaurants popped up about 15 m before so you had a chance to think if you wanted to eat there or not."

Relevance

All participants found the information they received to be relevant both considering the right time and place. That is showing today's offer when passing the restaurant and not what movie is shown in the cinema. This category had a median value of 5 (IQR = 0).

Positive feedback was given about the relevance of information during the interviews. One participant particularly liked that the daily menu was shown in such a way that you knew what was being offered without the need to take out the mobile phone to search for that information. Two participants, though, thought that all the information would be intrusive when they went down a street that only consisted of restaurants.

Responsiveness

The majority of the participants agreed that the system felt responsive. For example, one participant commented that "The system reacted immediately when I turned." Only one participant selected neutral (i.e. neither agreed nor disagreed). The median value was 5 (IQR = 1).

Technical stability

There were technical problems which the human operator needed to manage but from the participants' point of view there were two participants who commented on experiencing technical problems, one which resulted in aborting at the end of the city tour and for one who received low battery notification (Table 3).

Table 3, Listing problems that occurred during the city tour.

Type of error	Description	Number of occurrences
Lost Wi-Fi connection	Notification was not sent due to loss of Wi-Fi connection. For one participant this resulted in a missed turn. The human operator, however, managed to redirect the participant. In the case of the second participant, the city tour had to be ended prematurely. The only thing missing, however, was the showing of an image and information that the tour was over.	2
Low Battery	Low battery notification resulted in that the participant needed to hurry up to finish the city tour.	1
SmartWatch touch functionality	The touch functionality of the smartwatch did not register that the participant had pressed or made a swipe gesture. However, none of the eight participants noticed the lack of touch functionality since the human operator noticed the problem and sent the notification as if the touch functionality had worked.	15
Human operator error	The human operator missed to send a notification. However, the two participants showed no signs of noticing that something was missing.	2

Visual fidelity

The input from the participants on visual fidelity was diverse, but in general they did not express extreme opinions. The mean value was 2 (IQR = 2).

The feedback was quite varied in the interviews. One participant pointed out that when the image is "hanging" in front of you (i.e. when you turn your head, the image follows the head movement) it disfigures the view of what you actually want to see. Additionally, one participant disliked the current solution and suggested that if the image was correctly placed in the 3D space, it could be used as a means of interaction (i.e. if you did not look at the building the image would disappear). Another participant, though, liked it since it helped to find the "target" that one might have missed if the augmented information was correctly registered in the 3D space.

General user experience

Several statements were used to collect data about the general user experience such as S6, S7, S8, S9, S10, S11, S12 and S13 (Table 1).

The participants seemed to enjoy walking the tour; only one participant was neither positive nor negative towards the AR city tour simulated with WozARd. The amount of information was also considered to be well balanced.

The participants used the smartwatch to interact during the city tour and most of the participants found it discrete and intuitive. Four participants mentioned that they would like to be able to use speech as well.

Several participants commented on the industrial design of the glasses during the interviews. Example included: “It would be embarrassing, people would think that I had lost my mind”, “The design should be woman-friendly”, “I would like to use the system when I visit a new city, if it looked nicer.”

The answers were more diverse about both hearing and reading the notifications. The main problem with reading the notifications was glare due to the sun. Ten participants reported that they had problems with the sun. One stated, “I had to find a place in the shade to be able to read the notifications.”

Because some areas were crowded, some participants could not clearly hear the instructions. One user could not hear the information being presented when an ambulance passed by, and also stated that it was impossible to repeat the information.

One participant reported that it was disturbing that the arrows used for navigation had 90° angles, which resulted in several unnecessary turns.

Human operator performance

During the nine pilot trials, the importance of letting the human operator see what was displayed in the participant’s view was noticed. In the experimental set-up it was therefore arranged so that the human operator got visual feedback from the participant side. As for audio, however, there was no feedback indicating that the audio information was being played on the participant side or when it had finished. Due to this, the human operator could accidently interrupt the audio information by sending a new command to the participant. However, after the pilot trials the human operator knew when the sound was about to finish and could adjust the timing of the notifications. Consequently, none of the participants reported that they had any problems with the audio information being cut short.

Despite the attempt to make the wizard device UI as usable as possible, at two occasions the human operator missed to send a notification to the participant’s view (Table 3) when the two participants passed a point of interest of the AR city tour. Since the participants were unknowing about the information that should have been shown to them, they of course did not notice any problem.

Since the touch functionality of the smartwatch did not work properly and sometimes failed to catch the participant’s swipe gesture, the human operator had to be proactive and react when this occurred. The human operator managed to notice and address the problem every time it occurred (Table 3).

Another concern that made it difficult for the human operator was the Wi-Fi connection, which occasionally started to fluctuate in crowded areas. However, the connection managed to stabilize quickly enough so that all information and notifications were sent to the user with one exception (Table 3).

Also power usage constituted a problem. Since the screen, Wi-Fi and Bluetooth were always on, the phone needed to be recharged often. One misjudgment by the human operator resulted in a “Low battery” notification for one participant (Table 3).

Discussion

Overall, the results seem to indicate that the participants perceived the simulated AR city tour as a relatively realistic experience despite a certain degree of technical instability and human operator mistakes. Their subjective experience of the simulated AR city tour, as measured by the questionnaire, was overall positive and in general the city tour seemed to induce a feeling of a real, autonomous system rather than a system being controlled by someone else. The observation data seemed to confirm this. All participants managed to accomplish the AR city tour and in general they seemed to enjoy walking the simulated AR experience.

Based on the experiences of this study, the authors believe that two of the most important factors contributing to these results are the design of the wizard device of the WozARd tool and the skill of the human operator. The wizard device of WozARd was designed to aid the human operator in controlling the notifications during the user study and to reduce the risk of human operator mistakes. However, despite this notifications to be sent to the participant’s AR view were missed. This risk for mistakes could be decreased by letting WozARd’s wizard device provide the human operator with visual hints that aids him/her when activating commands in the GUI (Figure 2). For example, in the notification view, already shown notifications could be grayed out and the upcoming notification in the list could be highlighted in some way.

Since the participants were walking in an outdoor environment, unpredicted turns took place and therefore the human operator needed to react accordingly. Another example of a small detail that could have easily been missed by a novice human operator was that the smartwatch [32] did not work properly and sometimes failed to catch the participant’s swipe gesture. However, participants appeared not to pay attention to the problem since the human operator noticed and reacted when the touch functionality did not work. This indicates that it is important to have a skilled human operator who can control both WozARd and the test situation simultaneously in order to react to unpredictable and unexpected events. It has been suggested that the skill of the human operator is often a general problem with the WOZ method since it relies on the human operator not making any mistakes [42].

Another aspect that potentially can have a large impact on how a test participant perceives a WOZ test is the actual hardware used. The HMD [34] used in this study is one of the earliest HMDs available for early adapters. The insufficient display technology was reflected in the results concerning readability, one of the aspects that had the most diverse responses. The main reason for the participants’ troubles to read the notifications was glare due to the sun.

The participants who tested the tour in the afternoon reported the most problems with glare. This demonstrates a potential problem in using a WOZ tool like WozARd: the difficulties observed could be due to insufficient hardware rather than design issues connected to the AR user interface itself. The potential bias this can introduce in test results must be carefully considered when using a method like WozARd. The glare problem, nevertheless, can be expected to diminish with the development of newer display technology such as the one used in *Epson Moverio BT-200* [8]. Another aspect concerning hardware is how participants' behavior and performance may be affected by the actual industrial design. Several participants commented that they did not think that the HMD was very attractive. However, the present study did not target potential effects of the system's industrial design and we therefore only report that this is a possible source of bias to be aware of. Naturally, the use case as well as the chosen field setting of being in an outdoor environment has an impact on how a test participant perceives an AR city tour. For example, problems such as glare from the sun, rain, crowded areas, and Wi-Fi connection could have been avoided if the study was conducted in a controlled indoor environment. However, since the VENTURI third year theme was about AR city tours conducting the study outdoor meant a higher ecological validity and richer feedback for the project.

In the present study only one specific use case for wearable AR was simulated. No real claims about the general usefulness of the WozARd method in a design process can therefore be made based on the presented data. Even though only a few participants commented on the lack of realistic tracking of the augmented data in the present study, many use cases involve moving people and objects and would depend on a real AR tracking algorithm that correctly integrate the augmented data with the real world 3D space. For such use cases, the WozARd method might not be able to facilitate meaningful prototyping. In its current form, WozARd is probably best suited for testing scenarios limited to visualization of 2D data, tactile feedback and auditive feedback. However, WozARd has been released as an open-source project [43], which makes it possible for others in the AR community to expand the features and tweak it to fit their own requirements and use cases.

Conclusions

In conclusion, the WozARd method seemed to work reasonably well at least for this specific use case. Further, two of the most important factors that contributed to the simulated AR experience and induced a feeling of a real, autonomous system are the design of the wizard device of the WozARd tool and the skill of the human operator. Last, although WozARd does not support AR tracking in its current form this did not emerge as a critical problem for this particular use case. However, if other use case with e.g. moving people and objects is intended to be studied AR tracking might be needed.

Future work

We will continue to develop WozARd and investigate its usefulness as a prototyping tool for wearable AR. Much effort will be put into the design of the wizard device in order to help the human operator not making any mistakes. Examples of improvements include letting WozARd's wizard device provide the human operator with visual hints that aids him/her when activating commands in the GUI and adding visual feedback when the audio has stopped playing on the participant's side. Further, it is of great interest to study the importance of the human operator's level of expertise in using the WozARd tool. One way of doing this could be to investigate the difference in performance between a group of novice human operators, who will only get a short introduction to WozARd, and a group of expert human operators.

A feature that could add to WozARd's ability to facilitate meaningful prototyping is AR tracking. This could be especially useful for studies that involve moving people and objects. Since WozARd has been released as an open-source project [43], this and other features could be developed together with other developers in the AR community.

The software development kit [44] for *Google Glass* was released after this study and WozARd has been updated to be able to run with *Google Glass*. It would thus be interesting to conduct another study in which the user only needs to wear *Google Glass* and a *Sony SmartWatch* to investigate what results WozARd could produce with better HMD hardware.

Finally, it would be interesting to investigate how WozARd can be used to explore AR UIs in other environments, such as the home, by simulating e.g. a smart living room in which the user can control consumer electronics with different wearable devices such as *Google Glass*, *Sony SmartWatch* or *Smartband*.

Acknowledgments

This research was partially funded by the European 7th Framework Program, under grant VENTURI (FP7-288238), and the VINNOVA funded Industrial Excellence Center EASE (Mobile Heights). The authors would like to extend their gratitude to Ertan Muedin who helped conduct the user studies. Special thanks to Eileen Deaner for language support.

References

[1] V. Genaro Motti and K. Caine, "Understanding the wearability of head-mounted devices from a human-centered perspective," in *Proceedings of the 2014 ACM International Symposium on Wearable Computers - ISWC '14*, 2014, pp. 83–86.

[2] J. Rekimoto, Y. Ayatsuka, and K. Hayashi, "Augment-able reality: situated communication through physical and digital spaces," in *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, 1998, pp. 68–75.

[3] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS '68 (Fall, part I)*, 1968, pp. 757 – 764.

[4] "Google Glass," 2013. [Online]. Available: https://www.google.com/glass/start/.

[5] "Meta Pro," 2014. [Online]. Available: https://www.spaceglasses.com/.

[6] "Recon Jet," 2014. [Online]. Available: http://www.reconinstruments.com/products/jet/.

[7] "Vuzix M100," 2014. [Online]. Available: http://www.vuzix.com/consumer/products_m100/.

[8] "Epson Moverio BT-200 Smart Glasses," 2014. [Online]. Available: http://www.epson.com/cgi-bin/Store/jsp/Product.do?sku=V11H560020.

[9] E. Barba, B. MacIntyre, and E. D. Mynatt, "Here We Are! Where Are We? Locating Mixed Reality in The Age of the Smartphone," in *Proceedings of the IEEE*, 2012, vol. 100, no. 4, pp. 929–936.

[10] R. Azuma, "A survey of augmented reality," *Presence*, vol. 4, no. August, pp. 355–385, 1997.

[11] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, H. Hörtner, and C. Lindinger, "Augmented reality navigation systems," *Univers. Access Inf. Soc.*, vol. 4, no. 3, pp. 177–187, 2005.

[12] N. Davies, J. Landay, S. Hudson, and A. Schmidt, "Guest Editors' Introduction: Rapid Prototyping for Ubiquitous Computing," *IEEE Pervasive Comput.*, vol. 4, no. 4, pp. 15–17, Oct. 2005.

[13] J. F. Kelley, "An empirical methodology for writing user-friendly natural language computer applications," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems - CHI '83*, 1983, no. 12, pp. 193–196.

[14] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of Oz studies—why and how," *Knowledge-based Syst.*, pp. 193–200, 1993.

[15] S. Dow, J. Lee, C. Oezbek, B. MacIntyre, J. D. Bolter, and M. Gandy, "Wizard of Oz interfaces for Mixed Reality Applications," in *CHI '05 extended abstracts on Human factors in computing systems - CHI '05*, 2005, pp. 1339 – 1342.

[16] S. Dow, B. MacIntyre, and J. Lee, "Wizard of Oz support throughout an iterative design process," *Pervasive Comput. IEEE CS IEEE ComSoc*, vol. 4, no. 4, pp. 18–26, 2005.

[17] M. Lee and M. Billinghurst, "A Wizard of Oz study for an AR multimodal interface," in *Proceedings of the 10th international conference on Multimodal interfaces - IMCI '08*, 2008, pp. 249 – 256.

[18] G. Alce, K. Hermodsson, and M. Wallergård, "WozARd: A Wizard of Oz Tool for Mobile AR," in *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI '13*, 2013, pp. 600 – 605.

[19] S. Carter, J. Mankoff, S. Klemmer, and T. Matthews, "Exiting the Cleanroom: On Ecological Validity and Ubiquitous Computing," *Human-Computer Interact.*, vol. 23, no. 1, pp. 47–99, 2008.

[20] J. E. McGrath, "Methodology matters: Doing research in the behavioral and social sciences," *Readings Human-Computer Interact. Towar. year 2000*, pp. 152–169, 1995.

[21] P. Chippendale, P. Benjamin, D. Buhrig, P. Eisert, S. BenHimane, V. Tomaselli, H. Jonsson, G. Alce, Y. Lasorsa, M. de Ponti, and O. Pothier, "VENTURI – immersiVe ENhancemenT of User-woRld Interactions," 2012.

[22] S. Feiner, B. MacIntyre, and T. Höllerer, "A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment," in *Proceedings of the International Symposium on Wearable Computers.*, 1997, pp. 74–81.

[23] J. Bolter, M. Engberg, and B. MacIntyre, "Media studies, mobile augmented reality, and interaction design," *interactions*, vol. 20, no. 1, pp. 36–45, 2013.

[24] Y. Li, J. I. Hong, and J. a. Landay, "Design Challenges and Principles for Wizard of Oz Testing of Location-Enhanced Applications," *IEEE Pervasive Comput.*, vol. 6, no. 2, pp. 70–75, 2007.

[25] S. Dow, B. Macintyre, J. Lee, C. Oezbek, J. D. Bolter, and M. Gandy, "Wizard of Oz Support throughout an Iterative Design Process," *Publ. by IEEE CS IEEE ComSoc*, vol. 4, no. 4, pp. 18 –26, 2005.

[26] S. Schlögl, G. Doherty, N. Karamanis, and S. Luz, "WebWOZ: a wizard of oz prototyping framework," in *Proceedings of the 2nd ACM SIGCHI symposium on Engineering interactive computing systems - EICS '10*, 2010, pp. 109 – 114.

[27] A. Green, H. Huttenrauch, and K. S. Eklundh, "Applying the Wizard-of-Oz framework to cooperative service discovery and configuration," in *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, 2004, pp. 575–580.

[28] S. Connell, P.-Y. Kuo, L. Liu, and A. M. Piper, "A Wizard-of-Oz elicitation study examining child-defined gestures with a whole-body interface," *Proc. 12th Int. Conf. Interact. Des. Child. - IDC '13*, pp. 277–280, 2013.

[29] T. Grill, O. Polacek, and M. Tscheligi, "ConWIZ: a tool supporting contextual Wizard of Oz simulation," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia - MUM '12*, 2012.

[30] A. Fleury, J. S. Pedersen, and L. Bo Larsen, "Evaluating user preferences for video transfer methods from a mobile device to a TV screen," *Pervasive Mob. Comput.*, vol. 9, no. 2, pp. 228–241, Apr. 2013.

[31] N. Linnell, R. Bareiss, and K. Pantic, "A wizard of oz tool for android," in *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion - MobileHCI '12*, 2012, pp. 65 – 70.

[32] "Sony SmartWatch," 2012. [Online]. Available: http://www.sonymobile.com/global-en/products/accessories/smartwatch/.

[33] "Sony SmartWatch 2," 2013. [Online]. Available: http://www.sonymobile.com/global-en/products/accessories/smartwatch-2-wrist-strap-se20/.

[34] “Vuzix Star 1200,” 2010. [Online]. Available: http://www.vuzix.com/UKSITE/augmented-reality/products_star1200.html.

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

[36] “Sony Xperia S,” 2012. [Online]. Available: http://www.sonymobile.com/global-en/products/phones/xperia-s/.

[37] “Sony Xperia Z,” 2013. [Online]. Available: http://www.sonymobile.com/global-en/products/phones/xperia-z/.

[38] “Sony High Definition Camcorder HDR-CX190,” 2012. [Online]. Available: http://store.sony.com/gsi/webstore/WFS/SNYNA-SNYUS-Site/en_US/-/USD/ViewProduct-Start?SKU=27-HDRCX190/B&CategoryDomainName=&CategoryName=.

[39] K. A. Ericsson and H. A. Simon, “Verbal reports as data,” *Psychol. Rev.*, vol. 87, no. 3, pp. 215 – 251, 1980.

[40] “System Usability Scale (SUS),” 06-Sep-2013. [Online]. Available: http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html. [Accessed: 22-Oct-2014].

[41] P. K. Janert, *Data Analysis with Open Source Tools*. 1005 Gravenstein Highway North, Sebastopol, CA 95472: O’Reilly Media, Inc., 2011.

[42] D. J. Lazar, D. J. H. Feng, and D. H. Hochheiser, *Research Methods in Human-Computer Interaction*. 2010.

[43] “WozARd source code,” 2014. [Online]. Available: https://github.com/sonyxperiadev/WozARd.

[44] “Google Developers,” 2014. [Online]. Available: https://developers.google.com/glass/develop/gdk/. [Accessed: 21-Oct-2014].

Paper 3: IVAR: A Prototyping Method to Simulate Augmented Reality Interaction in a Virtual Environment – A Pilot Study

Alce, G., Wallergård, M., Thern, L., Hermodsson, K. and Hadzovic, T. (2015).

To be submitted, to the

International Journal of Virtual Worlds and Human Computer Interaction,

Avestia publishing.

III

Avestia Publishing
International Journal of Virtual Worlds and Human Computer Interaction
Volume 2, Year 2014
Journal ISSN: 2368-6103
DOI: 10.11159/vwhci.2014.xxx

A Prototyping Method to Simulate Wearable Augmented Reality Interaction in a Virtual Environment - A Pilot Study

Günter Alce[1], Mattias Wallergård[2], Lars Thern[2], Klas Hermodsson[1], Tarik Hadzovic[2]
[1]Sony Mobile Communications
Mobile vägen 4, Lund, Sweden 221 88
Gunter.Alce@sonymobile.com; Klas.Hermodsson@sonymobile.com
[2]Lund University, Department of Design Sciences
P.O. Box 118, Lund, Sweden 221 00
Mattias.Wallergard@design.lth.se; Lars.Thern@gmail.com; Hadzovic.Tarik@gmail.com

Abstract-
Recently, we have seen an intensified development of head mounted displays (HMD). Some observers believe that the HMD form factor facilitates Augmented Reality technology (AR), a technology that mixes virtual content with the users' view of the world around them. One of many interesting use cases that illustrate this is a smart home in which a user can interact with consumer electronic devices through a wearable AR system. Building prototypes of such wearable AR systems can be difficult and costly, since it involves a number of different devices and systems with varying technological readiness level. The ideal prototyping method for this should offer high fidelity at a relatively low cost and the ability to simulate a wide range of wearable AR use cases.
This paper presents a proposed method, called IVAR (Immersive Virtual AR), for prototyping wearable AR interaction in a virtual environment (VE). IVAR was developed in an iterative design process that resulted in a testable setup in terms of hardware and software. Additionally, a basic pilot experiment was conducted to explore what it means to collect quantitative and qualitative data with the proposed prototyping method.
The main contribution is that IVAR shows potential to become a useful wearable AR prototyping method, but that several challenges remain before meaningful data can be produced in controlled experiments. In particular, tracking technology needs to improve, both with regards to intrusiveness and precision.

Keywords: Augmented Reality, Virtual Reality, Prototyping, User Interaction.

1. Introduction

Recently, we have seen an intensified development of head mounted displays (HMD), i.e. display devices worn on the head or as part of a helmet. Two of the most well-known examples are Google Glass [1] and Oculus Rift [2]. The HMD form factor facilitates Augmented Reality (AR), a technology that mixes virtual content with the users' view of the world around them [3]. Azuma [4] defines AR as having three characteristics: 1) Combines real and virtual, 2) interactive in real time and 3) registered in 3-D. One of many interesting use cases that illustrate this is a smart home in which a user can interact with consumer electronic devices through a wearable AR system. For example, such an AR system could help the user discover devices, explore their capabilities and directly control them.

Building prototypes of such wearable AR systems can be difficult and costly, since it involves a number of different devices and systems with varying technological readiness level. The ideal prototyping method for this should offer high fidelity at a relatively low cost and the ability to simulate a wide range of wearable AR use cases. Creating a prototyping method which fulfils these requirements is problematic due to underdeveloped or partially developed technology components, such as display technology and object tracking. Also the lack of development tools and methodologies is a hindrance [5]. In particular, it is difficult to achieve prototypes that offer an integrated user experience and show the full potential of interaction concepts.

There are numerous examples of prototyping methodologies and tools used for prototyping AR interaction. Some offer low fidelity at low cost, e.g. low fidelity mock-ups [6] and bodystorming [7], whereas some offer high fidelity at high cost e.g. a ''military grade'' virtual reality (VR) system [8]. In between these two extremes there is a huge variety of prototyping methods. For example, a prototyping method widely used within human-computer interaction is Wizard of Oz (WOZ), in which a human operator simulates undeveloped components of a system in order to achieve a reasonable level of fidelity.

This paper presents a proposed method for prototyping wearable AR interaction in a virtual environment (VE). From here on we refer to the method as IVAR (Immersive Virtual AR). IVAR was developed in an iterative design process that resulted in an adequate setup in terms of hardware and software. Additionally, a small pilot study was conducted to explore the feasibility of collecting quantitative and qualitative data from the proposed method.

The main contribution of this paper is to present IVAR, a method for exploring the design space of AR interaction using a VE.

In the next section relevant related work is presented. The method is described in the section called the IVAR method, which is followed by pilot experiment, results, discussion and conclusions.

2. Related Work

Prototyping is a crucial activity when developing interactive systems. Examples of methods which can be used for prototyping AR systems include low-fidelity prototyping, bodystorming and WOZ.

Each method has its advantages and disadvantages. For example, low fidelity prototyping such as paper prototyping can be very effective in testing issues of aesthetics and standard graphical UIs. However, to do so when designing for wearable AR interaction, a higher fidelity is likely to be required [9].

As already mentioned, a well-known prototyping method widely used within human-computer interaction is WOZ. For example, the WOZ tool WozARd [10] was developed with wearable AR interaction in mind. Some advantages of using WozARd are flexibility, mobility and the ability to combine different devices.

Carter et al. [9] states that WOZ prototypes are excellent for early lab studies but they do not scale to longitudinal deployment because of the labor commitment for human-in-the-loop systems.

Furthermore, WOZ relies on a well trained person, the wizard, controlling the prototyping system during the experiment. The skill of the wizard is often a general problem with the WOZ method since it relies on the wizard not making any mistakes [11].

VE technology has been used as a design tool in many different domains, such as architecture, city planning and industrial design [12]. A general benefit of using a VE to build prototypes of interactive systems is that it allows researchers to test systems or hardware that do not actually exist in a controlled manner. Another advantage, compared to the WOZ method, is that functionality in the prototype can be handled by the VE, instead of relying on an experienced human operator to simulate the technical system's behaviour. One of the main drawbacks when it comes to simulating AR interaction in a VE is related to the fidelity of the real world component in the system, i.e. the simulated real environment and objects upon which augmented information is placed [13]. Examples of issues include the lack of tactile feedback in the VE and physical constraints due to limitations of VE navigation compared to real world movements. However, these issues depend on the goal of the study and are likely to be less of an issue for some use cases.

Ragan et al. [13] used VEs to simulate AR systems for the purposes of experimentation and usability evaluation. Their study focused on how task performance is affected by registration error and not on user interaction. They conducted the study using a four-sided CAVE™ with an Intersense IS-900 tracking system for head and hand tracking and their setup can be considered to be a high cost system.

Baricevic et al. [14] present a user study, evaluating the benefits of geometrically correct user-perspective rendering using an AR magic lens simulated in a VE. Similar to Whack-A-Mole, the participants were asked to repeatedly touch a virtual target using two types of magic lenses, phone-sized and tablet-sized. The focus of the study was on which rendering method the participants preferred and not on the user interaction

Lee et al. [15] used a high-fidelity VR display system to achieve both controlled and repeatable mixed reality simulations of other displays and environments. Their study showed that the completion times of the same task in simulated mode and real mode are not significantly different. The tasks consisted of finding virtual objects and reading information.

Despite these research efforts, there are no AR prototyping tools, to the knowledge of the authors of this paper, that focuses on using a VE to prototype interaction concepts for wearable AR.

3. The IVAR method

We reason that a method for prototyping wearable AR interaction in a VE should have the following characteristics in order to be effective:

- It should offer a degree of immersion, i.e. "*the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality*" [16], high enough to induce some degree of presence in the user.
- The interaction with the VE (navigation and manipulation of objects) should be intuitive and comfortable.

Compromising too much with these two issues would likely have a negative effect on prototyping outcomes. For example, a user who experiences a very low degree of presence in the VR is less likely to behave as he/she would have done in the corresponding real environment [17]. Furthermore, difficult and awkward VE interaction would bias the results of prototype evaluations [18], making it very difficult to say if user performance is due to the VE interaction or due to prototyped AR interaction.

With the fast development of off-the-shelf technology it has become increasingly easier to build VR systems that live up to the two characteristics described above. For example, low price HMDs for VR are now sold by a number of companies. The development of input devices for VR is even faster and has produced products such as Microsoft Kinect [19], Razer Hydra [20] and Leap Motion [21], which offer different opportunities for tracking user input depending on the underlying technology.

To build a setup for the IVAR method, a number of off-the-shelf input/output devices (available in late 2013 when this study was initiated) were tested in different configurations and their advantages and disadvantages were reflected upon. Through this exploratory development the final IVAR system components were chosen (Figure 1). An important design guideline during the development was to map the physical world to the VE in an attempt to create higher immersion and more intuitive VE interaction (Figure 3).

The total effort spent in designing, implementing and testing the set-up was 45 person weeks.

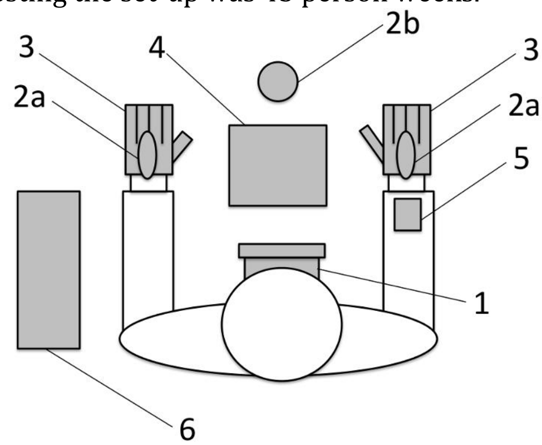


Figure 1. IVAR system components, 1) Oculus Rift [2], 2) Razer Hydra [20], 3) 5DT Data Glove 5 Ultra [22], 4) Sony Xperia Tablet Z [23], 5) Smartphone, 6) Desktop computer.

1. Oculus Rift Development Kit [2]. A VR head mounted display (HMD), having head tracking in three rotational degrees of freedom (DOF) and approximately 100^0 field of view.
2. Razer Hydra [20]. A game controller system that tracks the position and orientation of the two wired controllers (2a) in six DOF relative to the base station (2b) through the use of magnetic motion sensing. The two controllers are attached to the back of the user's hands. This enables the system to track location and orientation of the user's hands, and in extension via inverse kinematics the flexion and rotation of the user's arms.
3. 5DT Data Glove 5 Ultra [22]. A motion capture glove, which tracks finger joint flexion in real time.
4. Sony Xperia Tablet Z [23]. An Android powered 10'' tablet. The tablet is placed on a table in front of the user in alignment with the location of a tablet in the VE (Figure 2). The tablet allows the system to capture and react to touch input from the user. Additionally, it offers tactile feedback, which is likely to result in higher immersion and more intuitive VE interaction.

5. Android powered smartphone. This device is attached to the wrist of the user's dominant arm and is used to simulate a wristband that gives vibration feedback. The location of this feedback device is aligned with the virtual wristband in the VE (Figure 2).
6. Desktop computer with a powerful graphics card. This computer executes and powers the VE through the use of the Unity [24] game engine. The computer is powerful enough to render the VE at a high and steady frame rate.

Figure 2. The VE.

Most of the IVAR system components are wired, making this setup unsuitable for interaction where the user needs to stand up and walk around. However, the setup works for use cases that involve a seated user. For this reason, it was decided to implement a VE based on a smart living room scenario in which a user sitting in a sofa can interact with a set of consumer electronic devices (Figure 2). A tablet device is lying on the table in front of the user. The tablet is one of the input devices that can be used to control the living room. A TV is hanging on the wall in front of the user where media can be played back. Other consumer electronic devices including speakers, audio receiver, game console and printer, are located throughout the living room. The user is presented with overlaid information as spatially placed, visual feedback while interacting with the VE. In a real world scenario, this augmentation could originate from an optical see-through HMD, ceiling mounted projector or similar.

It has been suggested that physical awareness is important for effective interaction [25] in a VE. Therefore, some of the furniture and devices in the VE were mapped with corresponding objects in the real world (Figure 3). For instance, 1) The virtual table was mapped with a physical table, 2) The virtual tablet was mapped with a physical tablet, and 3) the virtual smartband was mapped with a smartphone (Figure 3). The physical table and tablet provided tactile feedback and the smartphone provided haptic feedback.

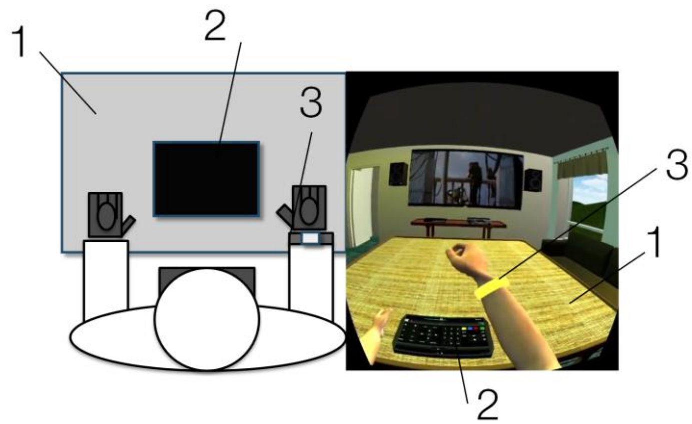


Figure 3. The physical devices mapped with virtual devices, 1) table, 2) tablet, 3) smartband.

In order to facilitate immersion, ease of interaction and physical awareness, the VE was equipped with a virtual representation of the user's own body. The virtual body was based on a 3D model of a young male whose two virtual arms could be moved in six DOF by the user. The ten virtual fingers could be moved with one rotational DOF. This was the most realistic input that could be achieved with the available tracking devices.

Four well-known interaction concepts with relevance for wearable AR were implemented in the VE. The concepts support two tasks that can be considered fundamental for a smart living room scenario: device discovery and device interaction, see video[1]. The purpose of the interaction concepts was to use them for starting exploring what it means to collect quantitative and qualitative data with a prototyping method like IVAR. The four interaction concepts and their usability in relation to wearable AR are not the focus of this paper but are instead described and reflected upon in more detail in [26].

[1] Concepts at youtube: http://goo.gl/zewYjm

Device discovery - gesture: This concept allows the user to discover the identity and location of consumer electronic devices by moving the dominant hand around, as if "scanning" (Figure 4). When a discoverable device is pointed at by the user's hand for more than one second, the user receives vibration feedback from his/her wristband and a window with information of the device is displayed. When the user is no longer aiming at the device with his hand, the window disappears.

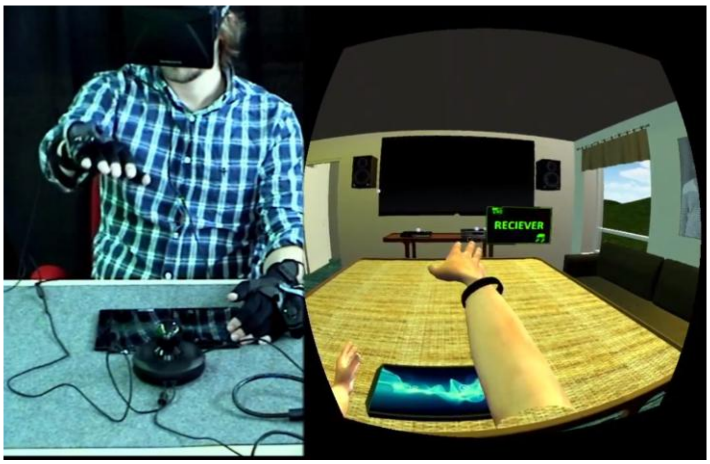


Figure 4. Device discovery – gesture.

Device discovery - gaze: This concept works like gesture except that the user is using his gaze instead of his hand movements to discover devices. Another difference is that the user does not receive vibration feedback when the information window appears. The hardware system does not include a gaze tracking component, instead the centre of the display view is used to approximate the focus of the user's gaze.

Device interaction - grab: This concept allows the user to select a playback device with a grabbing gesture. The user first selects an output device by reaching towards it with an open hand and then clenching the fist. If a device was correctly selected, the virtual wristband gives vibration feedback and also changes colour from black to yellow. The device remains selected as long as the user keeps making a fist. The user can then place the hand above the tablet and unclench the fist, which makes the tablet render a UI for media playback.

Device interaction - push: This concept allows the user to first select media content on the tablet and then select an output device by making a flick gesture towards it (Figure 5). The user then interacts with the tablet UI to control the playback of the content.

Figure 5. Device interaction – push.

4. Pilot experiment

In an attempt to create an initial understanding of what it means to evaluate wearable AR interaction with the IVAR method, a basic pilot experiment was conducted. The purpose of the pilot experiment was to let a group of participants carry out tasks with the four interaction concepts while collecting qualitative and quantitative data about their performance.

4.1 Setup

The pilot experiment was conducted in a usability lab with audio and video capturing capabilities. The sessions involved a participant, a test leader and a test assistant (Figure 6).

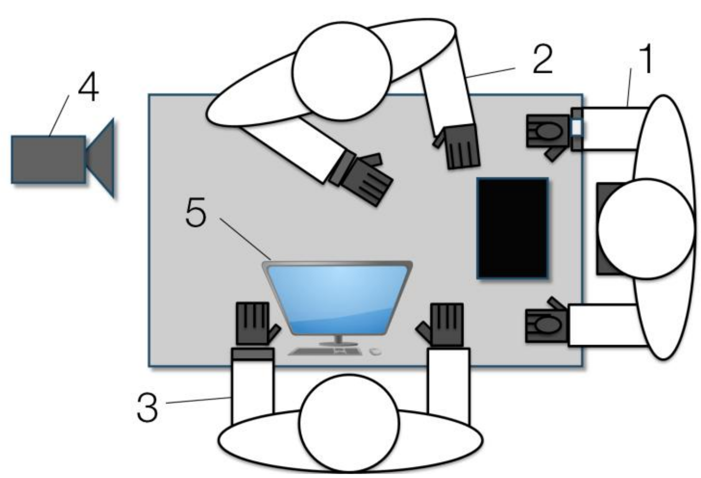


Figure 6. The experimental setup. 1) Participant, 2) Test leader, 3) Test assistant, 4) Video camera, 5) Desktop computer running the VE.

4.2 Procedure

First, the participant was informed about the purpose of the experiment and signed an informed consent form. The participant also filled in a background questionnaire, which included age, gender, occupation and 3D gaming experience.

Next, the participant was given a brief introduction to the idea of future types of interaction beyond the desktop paradigm. Thereafter, after the VR equipment had been put on the participant, he/she was given approximately five minutes to get acquainted with the VR system and the living room VE (Figure 2).

The participant was then informed about the task to perform in the living room VE. The task was to 1) find a device for video playback and then 2) start playback of a film on that device. In the first part of the task, the participant used the two device discovery concepts, i.e. gesture and gaze. In the second part, the two device interaction concepts, i.e. grab and push, were used by the participant. In other words, each participant tried all four interaction concepts. The order in which the concepts were tested was counterbalanced in order to address learning effects. During each test, task completion time, performed errors and error recovery time were recorded and logged by the system. Furthermore, multiple audio/video feeds and one video feed from the VE were captured.

After each interaction concept, the participant filled out a NASA-TLX questionnaire [27]. NASA-TLX is commonly used [28] to evaluate perceived workload for a specific task. It uses an ordinal scale on six subscales (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration) ranging 1 - 20.

The questionnaire was followed by a short semi structured interview.

When the participant had completed all four interaction concepts, he/she was asked questions about perceived differences between the concepts for device discovery and device interaction respectively.

The participant's comments from the experiment were transcribed and analyzed. Total perceived workload was calculated for each participant based on the NASA-TLX data. A Wilcoxon signed rank test for two paired samples ($p < 0.05$) was used to find eventual differences between interaction concepts with regards to the NASA-TLX data.

4.3 Participants

Participants were enrolled among university students; the majority of them studied engineering. 24 persons (9 female) with mean age 24.5 (19 - 37 years) participated in the device discovery part. Only 20 of these participants (9 female) performed the device interaction part of the experiment due to technical problems.

5. Results

This section presents quantitative and qualitative data from the pilot experiment. Overall, all participants managed to complete the tasks and the majority of them showed signs of enjoyment. The following data is divided in to three sections, device discovery, device interaction and possible sources of error.

5.1 Device discovery

All participants managed to perform the device discovery part of the experiment.

5.1.1 Quantitative data

The NASA-TLX data is presented in Table 1 and Figure 7, respectively. The statistical significant differences are marked with an asterisk.

Table 1. Device discovery: Median and Z-value for NASA-TLX values. Perceived workload on top followed by the six sub scales.

Device Discovery	Gesture	Gaze	Z-value
Workload	5.13	3.03	3.40*
Mental Demand	5.00	2.50	1.32
Physical Demand	5.50	1.50	4.07*
Temporal Demand	3.00	3.00	1.16
Performance	5.00	3.00	2.23*
Effort	5.00	2.50	2.81*
Frustration	4.00	2.50	-2.50*

* Significant at the 0.05 probability level

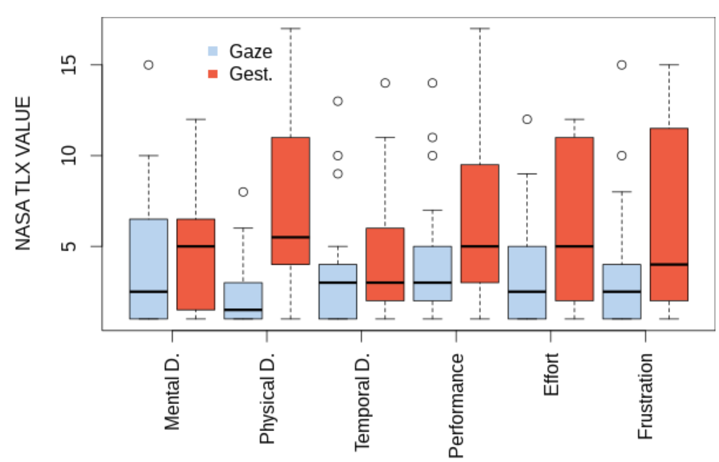


Figure 7. Boxplots for NASA-TLX data, gaze and gesture.

As for task completion time, no statistically significant difference between gesture and gaze could be found (median gesture = 28 and median gaze = 26.5 with range (8 - 142) and (6 - 93) respectively), ($Z = -0.51$, $p < 0.05$).

Table 2 presents the total number of errors and the average recovery time for each device discovery concept. An error was defined as a faulty action made by the participant. E.g. if the participant was asked to play a video on the TV but instead chose the PlayStation it was considered a fault. The recovery time was defined as the time from the faulty action until the correct action was initiated. No statistically significant differences between the two concepts with regards to number of errors and average recovery time could be found (Table 2).

Table 2. Number of errors and average recovery time for the device discovery concepts.

Device discovery	Number of errors	Average recovery time
Gesture	16, $SD = 1.16$	15.06, $SD = 8.21$
Gaze	32, $SD = 1.63$	6.13, $SD = 3.25$

5.1.2 Qualitative data

In general, participants tended to describe gaze as "natural" and "intuitive".

Several participants commented that the augmented information appearing on the discoverable consumer electronic devices might become intrusive. One participant commented that ''When the cards kept appearing I got overwhelmed'' and another related that ''It can be annoying with too much information, popping up all the time''.

Several participants stated that they felt more comfortable and in control with the gesture concept. Two illustrative comments were: ''It is comparable to point towards objects one is interested in'' and ''I feel more comfortable and secure, I am more in control''. Two negative comments about the gesture concept were: ''Stuff like this... look cool in movies but if it was my living room I would not want to move too much... the hand moving is a lot to ask'' and ''Feels weird, not a fan''.

5.2 Device interaction

All participants managed to finish the two device interaction tasks. However, the data of four of the participants could not be used due to technical problems.

5.2.1 Quantitative data

The NASA-TLX data is presented in Figure 8. No significant differences between the two device interaction concepts could be found in the NASA-TLX data. Nevertheless, there was a statistically significant difference between grab and push with regards to task completion time (grab = 28.20, push = 12.88 and median 18 and 10.75 respectively, with a range (8 - 75) and (6 - 28.5) respectively), ($Z = 2.39$, $p < 0.05$).

No statistically significant results were found for number of errors and average recovery time (Table 3). The definition of the errors and the recovery time has been explained in the device discovery section (Section 5.1.1).

Table 3. Number of errors and recovery time by the participants for the device interaction concepts.

Device Interaction	Number of errors	Average recovery time
Grab	4, $SD = 0.70$	8.88 , $SD = 2.51$
Push	6, $SD = 0.66$	4.33, $SD = 1.99$

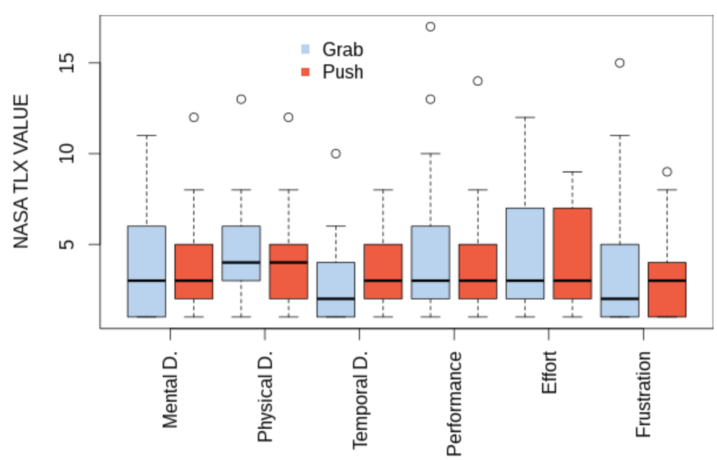


Figure 8. Boxplots for NASA-TLX data, grab and push.

5.2.2 Qualitative data

Overall, the participants seemed to find the push concept easy and convenient to use. Two comments were: ''A very convenient way to select and start'' and ''I think it is comfortable, it is intuitive to send things in the right direction''.

Nevertheless, some of the participants seemed to find the push concept awkward and felt insecure about where the pushed content would end up. One participant reasoned that ''I think it is more logical to start with the (output) device''. Another participant imagined a scenario with a large amount of consumer electronic devices: ''When having many devices I would not have felt comfortable to send things like this''.

Overall, the participants seemed to find the grab concept intuitive. Two comments that illustrate this were: ''Cool idea and it felt very intuitive'' and ''If you just want to watch TV then just grab''. However, also negative comments appeared: ''It felt somewhat unnatural... you want to feel what you have grabbed in the hand, here you grab in the thin air'' and ''This is not as easy as push''.

5.3 Possible sources of error

In this section, some possible sources of errors that could have influenced the pilot experiment data are presented.

Gloves. The Razer Hydra controllers were attached to the participant's hands by fastening them on thin gloves on top of the 5DT Data Gloves. These gloves were only available in one size, which created problems for participants with smaller hands. The problem was that the Razer Hydra controllers could move a bit in relation to the hand, which had a negative effect on the tracking accuracy for these participants. The problem mainly appeared for Gesture and Grab but also to some extent for Push.

Cables. The VR system setup included a number of cables. The cables of the 5DT Data Gloves and the Razer Hydra controllers were grouped together on the participant's arm whereas the cable of the Oculus Rift HMD went behind the participant's back. It could be observed that the cables to some extent restricted the participant's movements. The problem was particularly apparent for participants with smaller hands since the weight of the cables attached to the participant's arm induced even larger movements of the Razer Hydra controller in relation to the hand.

6. Discussion

In this paper we have presented IVAR, a method for prototyping wearable AR interaction concepts in a VE. A basic pilot experiment was performed to create an initial understanding of what it means to evaluate wearable AR interaction with this type of method.

Overall, the results from the pilot experiment can be described as mixed. On the one hand, the participants in general managed to solve the tasks in the VE. Furthermore, the qualitative data suggests that the participants had reached a level of understanding for the interaction concepts, even though they were not familiar with either the concepts or the medium, to be able express preference and evaluative statements. On the other hand, it is very difficult to say something about the effectiveness of IVAR based on the quantitative data. Some very rough tendencies could be observed in the NASA-TLX data, especially when comparing gesture and gaze. This could eventually be interpreted as a result of these two interaction concepts being quite different to their nature. However, it was apparent that some participants experienced problems during the experiment, inducing a considerable effect on task performance time, number of errors and recovery time. In some cases this was clearly due to the somewhat error prone tracking. Also, some participants seemed to occasionally loose track of the experimental tasks and appeared more interested in enjoying the VE and the VR experience.

What it boils down to is that the validity of a method based on participants' perceptions and actions inside a VE must be carefully considered. One could argue that the proposed method constitutes a sort of Russian nested doll effect with ''a UI inside a UI''. This raises the question: are observed usability problems caused by the UI or by the VR technology, or by both? Providing a definitive answer to this question is well beyond the scope of this paper, but based on our results the IVAR method appears to have an adequate potential, making continued exploration and development of the method worthwhile for the purpose of prototyping AR interaction.

To accurately evaluate and compare different AR interaction concepts, a VR setup with higher fidelity than the one presented in this paper would be needed. From a validity point of view, it is crucial that participants behave as they would have done in a corresponding real situation and that their intents and actions are translated to the VE in a precise manner. A key element for achieving this is immersion (as described earlier). Examples of factors that contribute to immersion include field of view, resolution, stereoscopy, type of input, latency etc. In the VR system described in this paper a relatively high degree of immersion was achieved with a wide FOV HMD (100 degrees diagonal) and low latency head tracking. Nevertheless, there is still big room for improvement with regards to immersion factors of the VR system. For example, the version of the Oculus Rift used (DK1), lacks the ability to detect translational movement, restricting the user's head movements to three rotational DOF. Head positional tracking can improve depth perception and thus immersion in a VE due to motion parallax [29]. Furthermore, the display resolution of Oculus Rift DK1 is only 640*800 pixels per eye. Low resolution makes especially text hard to read and it is not possible to faithfully reproduce the crispness of a device UI inside the VE. Also, low pixel density and noticeable pixel persistence inhibit visual fidelity and consequently the degree of immersion.

In the case of user input, there is big room for improvements with regards to tracking hands and fingers. The setup used in this paper was relatively cumbersome with several tracking and mobile devices attached to the participant, resulting in a ''tangle'' of cables and straps. This probably had a negative effect both on immersion and the precision by which participants could perform tasks in the VE. This type of setup is therefore likely to introduce variables that significantly interact with what is being studied using the IVAR method, i.e. different wearable AR interaction concepts. It is also important to note that the precision problem most likely affected the data to a lesser extent in the present study since none of the four concepts involved fine motoric movements. Prototyping and comparing different interaction concepts based on e.g. single-finger flicking gestures would require a much more precise tracking system in order to be meaningful. An alternative setup could consist of Leap Motion's Dragonfly [30] mounted at the front of the Oculus Rift DK2 [31]. Such a VR system would not require the participant to wear any tracking devices on his/her upper limbs, which would lead to much less movement restrictions. It would also facilitate significantly more precise tracking not least by offering several rotational DOFs for each finger. Nevertheless, other problems could appear due to the restricted field of view of the Leap sensor.

The four interaction concepts used in the pilot experiment were for a particular use case that assumes ''interaction from a couch'' with a stationary user. For this use case, the tethered hardware only limits the user's actions to some extent. A use case with more mobility involved would render this VR-based method hard to use. This would either require a portable VR setup or one that allows intuitive VE navigation, e.g. an omnidirectional treadmill. However, this is expensive hardware that can be difficult to use and maintain. Nevertheless, low-cost hardware targeting VR gaming can be expected to appear soon. One example is Virtuix Omni [32] which uses a slippery platform to allow the user to turn, walk or run in place and translate it to similar motion inside a VE.

7. Conclusions

The main conclusion of the research described in this paper is that IVAR shows potential to become a useful prototyping method, but that several challenges remain before meaningful data can be produced in controlled experiments. In particular, tracking technology needs to improve, both with regards to intrusiveness and precision.

8. Future work

To further explore the proposed prototyping method, more controlled experiments should be performed. Most importantly, experiments with a control group are needed in order to learn more about the validity of the method. A possible setup for such an experiment could be to let a control group perform tasks in a real world scenario and compare their performance with that of a group doing the same tasks in a VE.

As has been discussed above, the VR setup described in this paper has big improvement potential in terms of hardware. For example, a new version of the Oculus Rift Development kit, which feature positional tracking, a resolution of 960*1080 per eye and low pixel persistence was released in the summer of 2014 [31]. An alternative VR HMD is also being developed by Sony under the codename Project Morpheus [33]. An alternative display solution for the proposed prototyping method could be see-through displays, such as Microsoft HoloLens [34]. This opens up the possibility to prototype person-to-person interaction as well as multiple user consumer electronics interaction based on AR.

Furthermore, future VR setups for prototyping wearable AR interaction could exploit the benefits of wireless tracking systems such as the upcoming STEM wireless motion tracking system [35] or Myo [36], a gesture control armband.

9. Acknowledgments

This research is partially funded by the European 7th Framework Program under grant VENTURI (FP7-

288238) and the VINNOVA funded Industrial Excellence Center EASE (Mobile Heights).

References

[1] "Google Glass," 2013. [Online]. Available: https://developers.google.com/glass/design/ui. [Accessed: 29-Apr-2015].

[2] "Oculus Rift - Virtual Reality Headset for Immersive 3D Gaming," 2014. [Online]. Available: http://www.oculusvr.com/rift/. [Accessed: 29-Apr-2015].

[3] E. Barba, B. MacIntyre, and E. D. Mynatt, "Here We Are! Where Are We? Locating Mixed Reality in The Age of the Smartphone," in *Proceedings of the IEEE*, 2012, vol. 100, no. 4, pp. 929–936.

[4] R. Azuma, "A survey of augmented reality," *Presence*, vol. 4, no. August, pp. 355–385, 1997.

[5] N. Davies, J. Landay, S. Hudson, and A. Schmidt, "Guest Editors' Introduction: Rapid Prototyping for Ubiquitous Computing," *IEEE Pervasive Computing*, vol. 4, no. 4, pp. 15–17, Oct. 2005.

[6] M. De Sá and E. Churchill, "Mobile augmented reality: exploring design and prototyping techniques," *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile devices and Services.*, pp. 221 – 230, 2012.

[7] A. Oulasvirta, E. Kurvinen, and T. Kankainen, "Understanding contexts by being there: case studies in bodystorming," *Personal and Ubiquitous Computing*, vol. 7, no. 2, pp. 125–134, Jul. 2003.

[8] D. Courter and J. Springer, "Beyond desktop point and click: Immersive walkthrough of aerospace structures," *Aerospace Conference 2010 IEEE*, pp. 1–8, 2010.

[9] S. Carter, J. Mankoff, S. Klemmer, and T. Matthews, "Exiting the Cleanroom: On Ecological Validity and Ubiquitous Computing," *Human-Computer Interaction*, vol. 23, no. 1, pp. 47–99, 2008.

[10] G. Alce, K. Hermodsson, and M. Wallergård, "WozARd: A Wizard of Oz Tool for Mobile AR," in *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI '13*, 2013, pp. 600 – 605.

[11] D. J. Lazar, D. J. H. Feng, and D. H. Hochheiser, *Research Methods in Human-Computer Interaction*. 2010.

[12] R. C. Davies, "Applications of Systems Design Using Virtual Environments," in *Handbook of Virtual Environments - Design, Implementation, and Applications*, K. M. Stanney, Ed. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, 2002, pp. 1079–1100.

[13] E. Ragan, C. Wilkes, D. A. Bowman, and T. Hollerer, "Simulation of augmented reality systems in purely virtual environments," *Virtual Reality Conference 2009. VR 2009. IEEE*, pp. 287–288, 2009.

[14] D. Baricevic, C. Lee, and M. Turk, "A hand-held AR magic lens with user-perspective rendering," *Measurement Science and Technologyixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pp. 197–206, 2012.

[15] C. Lee, G. Rincon, G. Meyer, T. Hollerer, and D. A. Bowman, "The effects of visual realism on search tasks in mixed reality simulation," *Visualization and Computer Graphics, IEEE Transactions*, pp. 547–556, 2013.

[16] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence Teleoperators virtual Environments 6.6*, pp. 603–616, 1997.

[17] J. Freeman, S. E. Avons, R. Meddis, D. E. Pearson, and W. IJsselsteijn, "Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli," *Presence Teleoperators Virtual Environments*, vol. 9, no. 2, pp. 149–164, Apr. 2000.

[18] D. A. Bowman, E. Kruijff, and I. LaViola Jr, Joseph J Poupyrev, *3D user interfaces: theory and practice*. Addison-Wesley, 2004.

[19] Microsoft, “Microsoft Kinect,” 2014. [Online]. Available: https://www.microsoft.com/en-us/kinectforwindows/. [Accessed: 29-Apr-2015].

[20] Sixense, “Razer Hydra | Sixense,” 2014. [Online]. Available: http://sixense.com/hardware/razerhydra. [Accessed: 29-Apr-2015].

[21] I. Leap Motion, “Leap Motion,” 2014. [Online]. Available: https://www.leapmotion.com/. [Accessed: 29-Apr-2015].

[22] 5DT, “5DT Data Glove 5 Ultra,” 2014. [Online]. Available: http://www.5dt.com/products/pdataglove5u.html. [Accessed: 29-Apr-2015].

[23] Sony, “Sony Xperia Tablet Z,” 2014. [Online]. Available: http://www.sonymobile.com/se/products/tablets/xperia-tablet-z/. [Accessed: 29-Apr-2015].

[24] U. Technologies, “Unity - Game Engine.,” 2014. [Online]. Available: http://unity3d.com/. [Accessed: 29-Apr-2015].

[25] W. E. Marsh and F. Mérienne, “Nested Immersion: Describing and Classifying Augmented Virtual Reality,” in *IEEE Symposium on 3D User Interfaces*, 2015.

[26] G. Alce, L. Thern, K. Hermodsson, and M. Wallergård, “Feasibility Study of Ubiquitous Interaction Concepts,” in *Procedia Computer Science, 39*, 2014, pp. 35–42.

[27] N. ARC, “NASA Task Load Index,” 2014. [Online]. Available: http://humansystems.arc.nasa.gov/groups/tlx/. [Accessed: 29-Apr-2015].

[28] S. G. Hart, “NASA-task load index (NASA-TLX); 20 years later,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006.

[29] K. Bystrom, W. Barfield, and C. Hendrix, “A conceptual model of the sense of presence in virtual environments,” *Presence Teleoperators Virtual Environments 8(2)*, pp. 241–244, 1999.

[30] SixSense, “Leap Motion Sets a Course for VR,” 2014. [Online]. Available: http://blog.leapmotion.com/leap-motion-sets-a-course-for-vr/. [Accessed: 29-Apr-2015].

[31] “The All New Oculus Rift Development Kit 2 (DK2) Virtual Reality Headset,” 2014. [Online]. Available: http://www.oculusvr.com/dk2/. [Accessed: 29-Apr-2015].

[32] Virtuix, “Omni: Move Naturally in Your Favorite Game,” 2014. [Online]. Available: https://www.kickstarter.com/projects/1944625487/omni-move-naturally-in-your-favorite-game. [Accessed: 29-Apr-2015].

[33] Sony, “Project Morpheus, exclusive PlayStation TV shows and the best in fluffy dog games,” 2014. [Online]. Available: http://blogs.sonymobile.com/2014/03/21/project-morpheus-exclusive-playstation-tv-shows-and-the-best-in-fluffy-dog-games/?rl=cn. [Accessed: 29-Apr-2015].

[34] “Microsoft HoloLens,” 2015. [Online]. Available: http://www.microsoft.com/microsoft-hololens/en-us. [Accessed: 04-Mar-2015].

[35] S. S. E. Inc, “STEM System | Sixense,” 2014. [Online]. Available: http://sixense.com/STEM. [Accessed: 29-Apr-2015].

[36] T. Labs, “Myo - Gesture control armband by Thalmic Labs,” 2014. [Online]. Available: https://www.thalmic.com/en/myo/.

Paper 4: Feasibility Study of Ubiquitous Interaction Concepts

Alce, G., Thern, L., Hermodsson, K. and Wallergård, M. (2014).

IV

Proceedings of the 6th
International Conference on Intelligent Human-Computer Interaction – iHCI '14.
(pp. 35-42). DOI: 10.1016/j.procs.2014.11.007.

Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Computer Science 00 (2014) 000–000
Procedia
Computer Science
www.elsevier.com/locate/procedia
6th International conference on Intelligent Human Computer Interaction, IHCI 2014

Feasibility Study of Ubiquitous Interaction Concepts

Günter Alce[a,*], Lars Thern[b], Klas Hermodsson[a], Mattias Wallergård[b]

[a]*Sony Mobile Communications, Nya Vattentornet, 221 88 Lund, Sweden*
[b]*Lund University, P.O. Box 118, 221 00 Lund, Sweden*

Abstract

There are all sorts of consumer electronics in a home environment. Using "apps" to interact with each device is neither feasible nor practical in an ubicomp future. Prototyping and evaluating interaction concepts for this future is a challenge. This paper proposes four concepts for device discovery and device interaction implemented in a virtual environment. The interaction concepts were compared in a controlled experiment for evaluation and comparison.

Some statistically significant differences and subjective preferences could be observed in the quantitative and qualitative data respectively.

Overall, the results indicate that the proposed interaction concepts were found natural and easy to use.

Keywords: Natural User Interfaces; Virtual and Augmented Reality; Ubiquitous computing

1. Introduction

In our home environment we surround ourselves with all sorts of devices: TVs, home entertainment systems, gaming consoles, digital photo frames etc. Forecasts predict an "Internet of Things" world where most devices can be remote controlled by users, or other devices through the Internet[1,2].

With constant miniaturization, devices of the future may not even have physical buttons or screens that allow direct interaction.

Imagine a home, with hundreds of connected devices serving the home and its residents. Using "apps" to interact with each device is neither feasible nor practical in such a home[3]. As of today interaction is heavily reliant on users' explicit actions: choosing what content to show, where to show it, how to show it and setting all preferences. Prototyping and evaluating interaction concepts beyond the current paradigm is a challenge, due to components that are undeveloped or not sufficiently advanced[4].

Many researches have studied ubicomp interaction by using paper prototyping[5] or a Wizard of Oz (WOZ) method i.e. where a human simulates missing parts of a system. For example, Lee at al.[6] used a modified WOZ setup, in a

* Günter Alce. Tel.: +46-703-362-017
E-mail address: Gunter.Alce@sonymobile.com

comfortable living room, where two subjects were involved. One of the subjects interacted with gestures while the other subject interpreted and acted upon them.

This paper presents two concepts for device discovery and two for device interaction implemented in a virtual home environment. The implemented concepts are simple but functional for device interaction in an ubicomp environment.

The concepts were carefully designed to make use of devices such as TV, tablet, game console and printer, since Consolvo et al.[7] found these devices to be the most frequently available devices.

The objective of the proposed interaction concepts is to support higher degrees of implicit human - computer interaction, aided by mobile and wearable devices, which the authors think will simplify both discovery and interaction with consumer electronics (CEs). This will yield in a low cognitive workload for the users and more intuitive and natural interaction.

To evaluate and compare the interaction concepts, a controlled experiment was performed. This experiment consisted of participants performing tasks inside a virtual environment (VE) in order to generate quantitative as well as qualitative data.

The main contribution of this paper, is to present four interaction concepts developed with virtual reality (VR) for interaction in ubicomp environments.

In the next section relevant related work is presented. Followed by method, results, discussion, conclusions and future work.

2. Related Work

Mark Weiser's vision stated in the 1990's has inspired many researchers, "The most profound technologies are those that disappear."[8]. This statement has been and still is used as an important guideline for researchers within ubicomp[9]. Currently many researchers in the ubicomp community are targeting aspects such as context-aware computing and ubiquitous intelligence i.e. computational intelligence that is part of both the physical and the digital worlds[9,10]. However, two areas that needs more attention is prototyping and evaluating ubicomp interaction and new input methods[4].

2.1. *Tools for prototyping ubicomp interaction*

Carter et al.[11] highlighted that developers have a limited set of tools including sketches, paper prototype mock-ups and WOZ simulations. Recently, platforms such as Arduino has been used for prototyping ubicomp applications. The integration of software and hardware with Arduino makes it relatively easy for anyone to prototype new concepts[12]. For example Amarino[13] is a useful tool that can be used for communication between smartphone and tangible devices such as clothes, furniture etc.

Examples of research tools that uses a WOZ method include ConWIZ[14], a WOZ tool with a mobile application capable of controlling simulations of contextual objects such as fans, lights etc. Fleury et al.[15] also used a WOZ setup to evaluate four different methods for transferring video content from a smartphone to a TV screen.

Other tools used for ubiquitous prototyping include[16,17,18,6].

2.2. *Evaluating input methods for ubicomp interaction*

There are several input methods suitable for ubicomp interaction that have been studied and evaluated over the past three decades. For example the "put that there" experiment[19] is an early study of using gestural input and voice commands. Wilson and Oliver[20] also used gesture and voice commands in one of their user interface systems. Wilson and Oliver present four systems[20] that uses both explicit and implicit interaction.

Consolvo et al. evaluated ubicomp applications by using the experience sampling method i.e. participants filled out questionnaires every day by responding to alerts. One finding was that it is reasonable to create scenarios for ubicomp applications where the user takes advantage of an available output device, particularly if the device is a television set, desktop computer or printer[7]. Similar output devices have been used in this paper. Other ubicomp interaction research developing and evaluating concepts include[7,21,22].

The listed research based on WOZ have the drawback of relying too heavily on the wizard not accidentally injecting noise in the test result.

The four concepts which are presented in this paper are real working concepts in a VE. Therefore, the results is dependent on the participant and the system during the experiments, allowing for replicapable testing.

3. Method

The four interaction concepts were developed in an iterative design process which included bodystorming, paper prototyping and a focus group session.

To build a system for execution and visualization of the interaction concepts, off-the-shelf input/output devices were used. These were thoroughly tested in different configurations and their advantages and disadvantages were reflected upon.

Through exploratory development the final system components were chosen (these are discussed in 3.2). This setup made it possible to map the physical world to the VE (Fig. 1).

Fig. 1. The VE

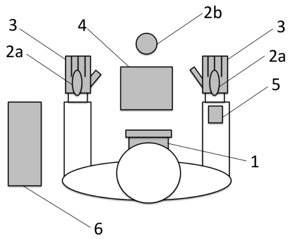


Fig. 2. System components

3.1. Design

The main input for conceptual design was bodystorming, when an idea for a concept was formulated, paper prototyping and wireframing would ensue to formulate use cases and to validate the ideas.

A number of prototypes were implemented using the system components at hand, the implementation of these prototypes made further improvement of the concepts easier since they were somewhat tangible.

3.2. System

The system components (Fig. 2) used to evaluate the four interaction concepts include:

1. Oculus Rift Development Kit[23]. A VR head mounted display.

2. Razer Hydra[24]. A tracking system enabling tracking of position and orientation of the two wired controllers (2a) relative to the base station (2b). The controllers are attached to the back of the user's hands.

3. 5DT Data Glove 5 Ultra[25]. A motion capture glove, enabling tracking of finger joint flexion in real time.

4. Sony Xperia Tablet Z[26]. An Android 10" tablet. The tablet is placed in alignment with a tablet in the VE (Fig. 1). The tablet allows the system to capture and react to touch input from the user.

5. Android smartphone. This device is attached to the wrist of the user's dominant arm and is used to give haptic feedback through vibrations. The location of this feedback device is aligned with the virtual wristband in the VE (Fig. 1).

6. Desktop computer with a powerful graphics card. This computer executes and powers the VE through the use of the Unity[27] game engine.

The VE is a living room (Fig. 1) where the user is sitting down on a couch in front of a table. A tablet, one of the input devices, is lying on a table in front of the user. A TV where media can be played is hanging on the wall in front of the user. Other CEs including speakers, audio receiver, game console and printer, are located throughout the living room to let the user discover more devices. The user is presented with overlaid information while interacting with the VE. This augmentation may originate from a head mounted display or other technologies. The purpose of using augmentation is to give the user spatially placed visual feedback. The input methods in combination with the overlaid information was chosen to allow for prototyping a range of existing, as well as future, device combinations.

3.3. Interaction Concepts

The main objectives for the four interaction concepts were: *a*) Minimizing the disruptive and distracting use of multiple controllers/ terminals, for controlling CEs. This might lower the workload. *b*) Minimizing the overwhelming amount of explicit actions needed to decide what content to display and what device to display it on. This might make the interaction more natural and intuitive. *c*) To bridge the gap between the real and the digital world, making CEs an extension of the users.

3.3.1. Device Discovery

The main inspiration for the device discovery concepts stems from two important notions from Norman[28] that makes for good and consistent device discoverability. "Discoverability: Is it possible to even figure out what actions are possible and where and how to perform them?" and "Understanding: What does it all mean? How is the product supposed to be used? What do all the different controls and settings mean?"

The Gaze concept allows the user to discover the identity and location of CEs in the user's vicinity by looking around (Fig. 3). The hardware system does not include a gaze tracking component, instead the center of the view is used as the focus of the user's gaze. When a discoverable device is in view center for more than one second, a window with interaction possibilities on the focused device is displayed. When the device is no longer in focus, the window disappears.

Fig. 3. Device discovery – Gaze

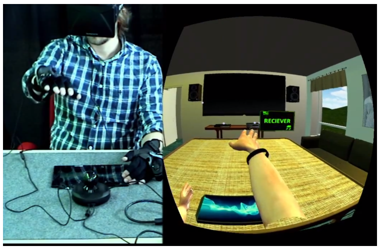


Fig. 4. Device discovery – Gesture

The Gesture concept allows the user to discover the identity and location of devices by moving the dominant hand in a scanning motion (Fig. 4). When a discoverable device is pointed at by the user's hand for more than one second, the user receives vibration feedback to his/her wrist and a window with interaction possibilities of the device is displayed. When the device is no longer pointed towards, the window disappears.

3.3.2. Device interaction

Tablets are attributed to shared ownership[29], therefore a tablet is used in both implementations of the device interaction concepts.

The Grab concept allows the user to select a playback device with a grabbing gesture (Fig. 5). The user first selects an output device by reaching towards it with an open hand and then clenching the fist. If a device was correctly selected, the user receives vibration feedback to his/her wrist and the virtual wristband of the outstretched hand changes color according to the grabbed device. The device remains selected as long as the user keeps making a fist. The next step is to move the closed hand to the surface of a device with a touchscreen and open the closed fist. If the fist is opened in mid-air the grabbed device is deselected and the virtual wristband returns to its original color. By opening the hand on the tablet surface, the tablet renders a user interface (UI) for media playback.

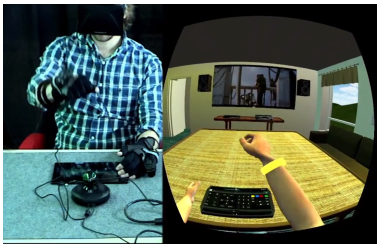


Fig. 5. Device interaction – Grab

Fig. 6. Device interaction – Push

The push concept allows the user to first select media content and then select the output device by flicking the content towards it (Fig. 6). The UI allows the user to select and control the playback of the content.

A video of the interaction concepts can be seen at `http://goo.gl/zewYjm`.

3.4. Evaluation

For evaluation, a controlled experiment was performed in a fully equipped usability lab. The experiment consisted of two parts; one for the device discovery concepts and another for the device interaction concepts. For the purpose of analysis both audio and video was captured in the usability lab.

Participants were mainly university students with engineering background. 24 persons (9 female) participated in the device discovery part. Their mean age was 24.5 (19 - 37 years). There were 20 participants (9 female) in the device interaction part and their mean age was 23.8 (19 - 35 years). The device interaction participants were a subset of the device discovery group (due to technical problems, four participants' data could not be used).

Each test session consisted of five steps (Fig. 7) which in total lasted approximately one hour. The session started with the participant signing an informed consent form and filling out a background questionnaire. The questionnaire included participant age, gender, occupation and 3D gaming experience.

Next, a brief introduction to ubicomp was presented. Thereafter the participant was given approximately five minutes to try out the VR system to get familiarized with the environment.

After the participant was familiar with the system the usability test started. Each participant performed tests on all four interaction concepts. The order in which a concept was evaluated was counterbalanced in order to avoid learning effects. Task completion time, performed errors and error recovery time were recorded, during each test. Furthermore, multiple audio/video feeds from the usability lab and one video feed from the VR system was captured.

Upon completion of each task, the participant filled out a NASA-TLX questionnaire [30]. The questionnaire was followed by a short semi structured interview.

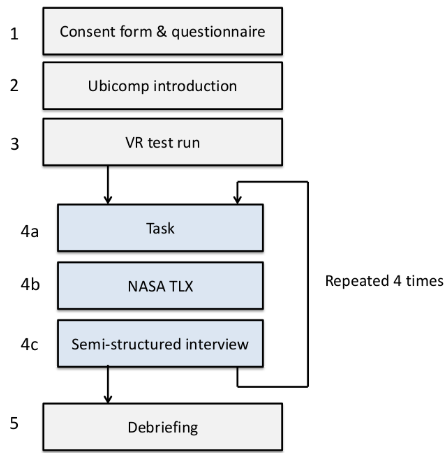


Fig. 7. Test session procedure

Since the tests were of two different characters (Device Discovery and Device Interaction) the participant would be given comparative questions after the second semi structured interview.

After the full test session the participant was debriefed in order to elicit further preference and evaluative statements.

Total perceived workload was calculated for each participant based on the NASA-TLX data. Correlation testing was performed on perceived workload and task completion time ($p < 0.05$). A Wilcoxon signed rank test for two paired samples ($p < 0.05$) was used to analyze the quantitative data and find out whether there were any significant differences between concepts. The participant's comments from the test session were transcribed and analyzed with individual quotes categorized and labeled in a table.

4. Results

This section presents quantitative and qualitative data from the experiments.

4.1. Quantitative

Moderate correlations were found for Gaze and Gesture regarding workload and time to complete task (Table 1). However, only small correlations can be seen for Grab and Push (Fig. 8, 9, 10, 11).

Table 1: Correlation of TLX Workload and Task completion time *p <0.05*

	Gaze	Gesture	Grab	Push
df	22	22	18	18
r	0.47	0.52	0.17	0.26

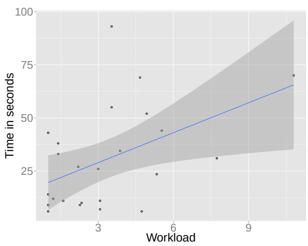


Fig. 8. Scatterplot – Gaze

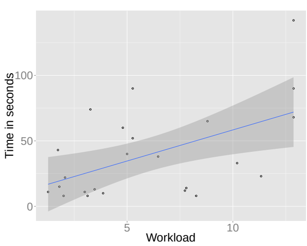


Fig. 9. Scatterplot – Gesture

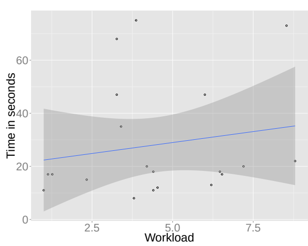


Fig. 10. Scatterplot – Grab

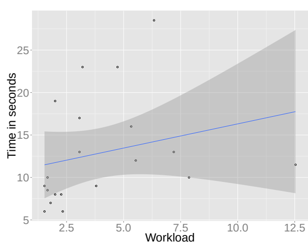


Fig. 11. Scatterplot – Push

For the two device discovery concepts, the mean NASA-TLX workload values (Gaze = 3.34, Gesture = 6.05 with median 3.03 and 5.13 respectively) were notably different (Fig. 12), with a significant difference between them (*Z = -3.40, p <0.05*). It is worth noting that there is a considerable significant difference in the perceived physical demand (*Z = 4.07, p <0.05*). With mean Gaze = 2.33, Gesture = 7.25 and a median of 1.5 and 5.5 respectively. In contrast, there is no statistical difference on perceived mental demand (*Z = -1.32, p <0.05*) with very similar means (Gaze = 4.13, Gesture = 4.92) but with large difference in median (2.5 and 5 respectively). There was no significant difference between Gaze and Gesture with regards to task completion time (*Z = -0.51, p <0.05*). With mean Gaze = 30.58, mean Gesture = 39.58 and median values of 26.5 and 28 respectively. With a range of (6 – 93) and (8 – 142) respectively.

For the two device interaction concepts, the mean NASA-TLX workload values (mean Grab = 4.72, mean Push = 4.19 with median 4.33 and 3.37 respectively) were not notably different (Fig. 13). Overall, the perceived workload was not significantly different between the two interaction concepts (*Z = 1.44, p <0.05*). There was no significant difference of the perceived physical demand (*Z = 1.66, p <0.05*). With mean values Grab = 5.45, Push = 4.45 and median 5.5 and 3 respectively. Furthermore no significant difference was found of the perceived mental demand (*Z = 0.40, p <0.05*). With mean Grab = 4.25, Push = 4.15 and median 4 and 3 respectively. However, there was a significant difference between grab and push with regards to task completion time (*Z = 2.39, p <0.05*). With a mean Grab = 28.20, Push = 12.88 and median 18 and 10.75 respectively. With a range (8 – 75) and (6 – 28.5) respectively.

No significant results were found considering error and error cost for any of the proposed concepts.

4.2. Qualitative

All qualitative data from the experiments were analysed by comparing their strengths and weaknesses. On the one hand the participants perceived the concepts being natural and intuitive. On the other hand, some attributes were stronger emphasised concerning each concept, these attributes are presented in Table 2.

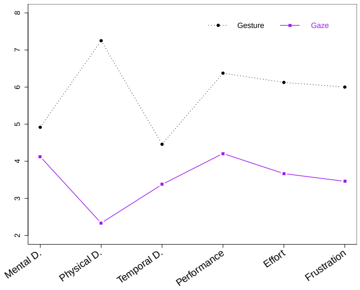


Fig. 12. Mean NASA-TLX values for device discovery

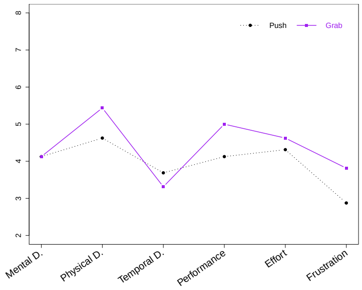


Fig. 13. Mean NASA-TLX values for device discovery

Table 2: Strengths/ Weakness Analysis

	Strengths	Weaknesses
Gaze	Natural, Intuitive, Convenient	Intrusive, Privacy Concerns
Gesture	Control, Precision, Secure	Socially awkward, Physically Demanding
Push	Convenient, Fast, Intuitive	Confusing, Lack of control when having several devices
Grab	Intuitive, In control	Awkward, Unnatural without a tangible device

5. Discussion

In this paper we have presented four interaction concepts for ubicomp, they were evaluated using immersive VR.

Overall, the results indicate that the users found the concepts natural and intuitive. Although statistically there were notable differences regarding how fast participants could finish their task, only small and moderate correlations were found between the task completion time and the perceived workload. This due to task completion time being affected by much more than the six categories in the NASA-TLX. For the device discovery concepts significant differences were found in perceived physical demand. Combining this with the qualitative data, one can make the assumption that higher demand not always is a bad thing. Whilst Gaze felt intrusive and raising privacy concerns, Gesture seemed to yield a feeling of being more in control and having higher security.

The Push and Grab concepts have similar characteristics which also can be seen in the NASA-TLX graph (Fig. 13). Further, in the case of Gaze and Gesture, the NASA-TLX graph(Fig. 13) were much more different, which accords well with the fact that these two interaction concepts are quite different as already discussed.

Nevertheless, the validity of an evaluation based on participants' perceptions and actions inside a VE must be carefully considered. One could argue that the proposed method constitutes a sort of Russian nested doll effect with ”a UI inside a UI”.

Although, while the suggested concepts are basic, they still are functional for ubicomp. System limitations that might have affected the participants are the cables and equipment that the users had to put on but also system limitation such as not having the possibility to lean forward or back. The user need to sit still and only move head, arms and fingers.

6. Conclusions and Future work

The main conclusion of the research described in this paper is that the four interaction concepts seems to be natural and easy to use. The concepts were developed on a very crude level and to make for deeper evaluation they need to be further explored. To further explore the proposed interaction concepts, more controlled experiments should be performed. Most importantly the debriefing needs to consist of more directed questions. And the number of participants needs to be greatly increased.

Acknowledgements

This research is partially funded by the European 7th Framework Program under grant VENTURI (FP7-288238) and the VINNOVA funded Industrial Excellence Center EASE (Mobile Heights).

References

1. Ericsson White paper, . More than 50 billion connected devices. http://www.ericsson.com/res/docs/whitepapers/wp-50-billions.pdf; 2014. Accessed: 2014-05-14.
2. Tudosoiu, B.. Feasibility Study: Minimum Viable Device to support Internet of Things Realization. http://mobileheights.org/wp-content/uploads/2013/10/Feasibility-Study_small.pdf; 2014. Accessed: 2014-05-14.
3. Jenson, S.. Mobile apps must die | blog | design mind. http://designmind.frogdesign.com/blog/mobile-apps-must-die.html; 2014. Accessed: 2014-05-16.
4. Davies, N., Landay, J., Hudson, S., Schmidt, A.. Guest Editors' Introduction: Rapid Prototyping for Ubiquitous Computing. *IEEE Pervasive Computing* 2005;**4**(4):p. 15–17. doi:10.1109/MPRV.2005.78.
5. Carter, S., Mankoff, J.. Prototypes in the wild lessons from three ubicomp systems. *Pervasive Computing, IEEE* 2005;:p. 51–57.
6. Lee, S.S., Chae, J., Kim, H., Lim, Y.k., Lee, K.p.. Towards more natural digital content manipulation via user freehand gestural interaction in a living room. In: *Ubicomp'13*. ACM; 2013, p. 617–626. doi:10.1145/2493432.2493480.
7. Consolvo, S., Walker, M.. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing* 2003; **2**(2):p. 24–31. doi:http://doi.ieeecomputersociety.org/10.1109/MPRV.2003.1203750.
8. Weiser, M.. The Computer for the 21st Century. *Scientific American* 1991;**265**(3):p. 66–75.
9. Rogers, Y.. Moving on from weiser's vision of calm computing: engaging ubicomp experiences. In: *Dourish, P., Friday, A. (eds.) UbiComp 2006: Ubiquitous Computing - 8th International Conference*; LNCS 4206. Springer; 2006, p. 404–421.
10. Sun, Q., Yu, W., Kochurov, N., Hao, Q., Hu, F.. A multi-agent-based intelligent sensor and actuator network design for smart house and home automation. *Journal of Sensor and Actuator Networks* 2013;**2**(3):p. 557–588.
11. Carter, S., Mankoff, J., Klemmer, S.R., Matthews, T.. Exiting the Cleanroom: On Ecological Validity and Ubiquitous Computing. *HUMAN-COMPUTER INTERACTION* 2008;**23**:p. 47–99. doi:10.1080/07370020701851086.
12. Hodges, S., Villar, N., Scott, J., Schmidt, A.. A new era for ubicomp development. *Pervasive Computing,* 2012;:5–9.
13. Kaufmann, B., Buechley, L.. Amarino: a toolkit for the rapid prototyping of mobile ubiquitous computing. In: *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. ISBN 9781605588353; 2010, p. 291–298.
14. Grill, T., Polacek, O., Tscheligi, M.. Conwiz: A tool supporting contextual wizard of oz simulation. In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*; MUM '12. New York, NY, USA: ACM. ISBN 978-1-4503-1815-0; 2012, p. 21:1–21:8. doi:10.1145/2406367.2406394.
15. Fleury, A., Pedersen, J.S., Bo Larsen, L.. Evaluating user preferences for video transfer methods from a mobile device to a TV screen. *Pervasive and Mobile Computing* 2013;**9**(2):p. 228–241. doi:10.1016/j.pmcj.2012.05.003.
16. Salber, D., Dey, A.K., Abowd, G.D.. The context toolkit: aiding the development of context-enabled applications. In: *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*. New York, New York, USA: ACM Press. ISBN 0201485591; 1999, p. 434–441. doi:10.1145/302979.303126.
17. Theofanos, M., Scholtz, J.. A framework for evaluation of ubicomp applications. In: *First International Workshop on Social Implications of Ubiquitous Computing, CHI*. 2005, p. 1–5.
18. Marquardt, N., Diaz-Marino, R., Boring, S., Greenberg, S.. The proximity toolkit: Prototyping proxemic interactions in ubiquitous computing ecologies. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. New York, New York, USA: ACM Press. ISBN 9781450307161; 2011, p. 315. doi:10.1145/2047196.2047238.
19. Bolt, R.A.. "put-that-there": Voice and gesture at the graphics interface. In: *Ubicomp'13*. ACM; 1980, p. 262–270.
20. Wilson, A., Oliver, N.. Multimodal sensing for explicit and implicit interaction. In: *In Proceedings of the 11th International Conference on Human-Computer Interaction (HCII05). Mahwah, NJ: Lawrence Erlbaum, Las Vegas, Nevada*. 2005, .
21. Oliver, N., Horvitz, E.. Selective perception policies for guiding sensing and computation in multimodal systems: A comparative analysis. *Computer Vision and Image Understanding* 2005;**100**(1):p. 198–224.
22. Bowman, D.A., Wingrave, C.A.. Design and evaluation of menu systems for immersive virtual environments. In: *Proceedings of the Virtual Reality 2001 Conference (VR'01)*. IEEE; 2001, p. 149–156. doi:10.1109/VR.2001.913781.
23. Oculus VR, . Oculus Rift - Virtual Reality Headset for Immersive 3D Gaming. http://www.oculusvr.com/rift/; 2014. Accessed: 2014-05-14.
24. Sixense, . Razer Hydra — Sixense. http://sixense.com/hardware/razerhydra; 2014. Accessed: 2014-05-14.
25. 5DT, . 5DT Data Glove 5 Ultra. http://www.5dt.com/products/pdataglove5u.html; 2014. Accessed: 2014-05-14.
26. Sony, . Sony Xperia Tablet Z. http://www.sonymobile.com/global-en/products/tablets/xperia-tablet-z/; 2014. Accessed: 2014-05-14.
27. Unity Technologies, . Unity - Game Engine. http://unity3d.com/; 2014. Accessed: 2014-05-14.
28. Norman, D.A.. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books; 2013. ISBN 9780465072996.
29. Kawsar, F., Brush, A.B.. Home computing unplugged: Why, where and when people use different connected devices at home. In: *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*; UbiComp '13. New York, NY, USA: ACM. ISBN 978-1-4503-1770-2; 2013, p. 627–636. URL: http://doi.acm.org/10.1145/2493432.2493494. doi:10.1145/2493432.2493494.
30. NASA ARC, . NASA Task Load Index. http://humansystems.arc.nasa.gov/groups/tlx/; 2014. Accessed: 2014-05-14.