A Body-and-Mind-Centric Approach to Wearable Personal Assistants

Shahram Jalaliniya

IT University of Copenhagen
Pervasive Interaction Technology Lab

A Thesis Submitted for the Degree of
Doctor of Philosophy
January 2016

SUPERVISOR

Associate professor Thomas Pederson
IT University of Copenhagen, Denmark

Abstract

Tight integration between humans and computers has long been a vision in wearable computing ("man-machine symbiosis", "cyborg"), motivated by the potential augmented capabilities in thinking, perceiving, and acting such integration could potentially bring. However, even recent wearable computers (e.g. Google Glass) are far away from such a tight integration with their users. Apart from the purely technological challenges, progress is also hampered by the common attempt by system designers to deploy existing interaction paradigms from desktop and mobile computing (e.g. visual output, touch-based input; explicit human-computer dialogue) to what is in fact a completely new context of use in which computer users interact with the device(s) on the move and in parallel with real-world tasks. This gives rise to several physical, perceptual, and cognitive challenges due to the limitations of human attentional resources. In fact, while wearable computers in recent years have become smaller and closer to our bodies physically, I argue in this thesis that to achieve a tighter man-computer symbiosis (e.g. such that some everyday decision-making can be offloaded from human to system as in context-aware computing) we also need to tie the computer system closer to the conscious and unconscious parts of our minds. In this thesis, I propose a conceptual model for integrating wearable systems into the human perception-cognition-action loop. I empirically investigate the utility of the proposed model for design and evaluation of a Wearable Personal Assistant (WPA) for clinicians on the Google Glass platform. The results of my field study in a Copenhagen hospital simulation facility revealed several challenges for WPA users such as unwanted interruptions, social and perceptual problems of parallel interaction with the WPA, and the need for more touch-less input modalities. My further exploration on touch-less input modalities such as body gestures and gaze, showed the great potential of using eye movements as an implicit input to WPAs. Since the involuntary eye movements (e.g. optokinetic nystagmus) are unconscious reflections of the brain to external stimuli, analyzing such involuntary eye movements by WPAs opens new opportunities for unconscious interaction (and man-machine symbiosis). Our EyeGrip prototype successfully demonstrated user's complete unconscious control over visual information flows (e.g. Facebook feeds), automatically halting the scrolling when an entity of interest shows up. In this thesis I lay a conceptual framework and demonstrate the potential of some new avenues for achieving tighter integration between wearable computers and human agents.

Acknowledgments

This dissertation would not have been possible without support of others. First of all, I would like to express my special appreciation and thanks to my supervisor Dr. Thomas Pederson. Thank you for your understanding and availability whenever I have needed it.

Besides my advisor, I would like to thank my thesis committee: Prof. Jeff Pelz, Prof. Kasper Hornbæk, and Assoc. Prof. Erik Grönvall for their insightful comments. My sincere thanks also goes to Prof. Oliver Amft, Prof. Peter Eklund, Prof. Jakob Bardram, Prof. Hans Gellersen, Prof. Gerhard Tröster, Assoc. Prof. Tomas Sokoler, and Assoc. Prof. Dan Witzner for their support and constructive comments during my PhD.

I would like to thank all my colleagues in Pervasive Interaction Technology (PIT) Lab for their valuable input at various stages of my PhD. Thank you Steven Houben for being my iCareNet buddy, for debates, dinners, rides to the airport, general help, and friendship. Also I thank Steven Juris, Morten Esbensen, Paolo Tell, Mads Frost, and Sebastian Büttrich for their help through this entire process. A special thanks to Diako Mardanbegi for his amazing support and collaboration in gaze-based interaction during my research.

I am also thankful for the support that I received from Ulrik Kähler Olesen from the Department of Traumatology and Orthopedic Surgery at the Rigshospitalet, Copenhagen.

To my friends: Aidin, Azadeh, Salar, Mana, Sina, Niloufar, Faramarz, Nastaran, Amir, and Shohreh thanks for all the good times, laughs, and parties we have had together.

I would like to acknowledge my parents who have made available their supports remotely in a number of ways. I would also like to show my deepest gratitude to my beloved wife Banafsheh and my sweet daughter Sheida for their constant support during the long hours and lost weekends required to write this thesis.

All the work presented in this dissertation was supported by the EU Marie Curie Network iCareNet grant number 264738, IT University of Copenhagen and the Pervasive Interaction Technology Laboratory (PIT Lab).

Reader's Guide

This dissertation consists of an introduction chapter and two parts. The introduction chapter of the thesis aims to provide an overview of the research context, selected research questions, and the chosen method to answer these open questions. Part I comprises 7 chapters and provides an overview of the conceptual background of the work and description of the system prototypes. Part II consists of a collection of 10 papers that contain concepts, technology and user studies summarized and discussed in Part I. 9 papers have been peer-reviewed and published and one is submitted. The blue colored text in this dissertation is clickable for readers of the electronic version.

List of Publications

Journal & Magazine Articles

1. **Jalaliniya, S.**, & Pederson, T. (2015). Designing Wearable Personal Assistants for Surgeons: An Egocentric Approach. IEEE Pervasive Computing, 14(3), 22-31.
   (10 pages)

2. Bolton, F., **Jalaliniya, S.**, & Pederson, T. (2015). A Wrist-Worn Thermohaptic Device for Graceful Interruption. Interaction Design & Architecture(s), N.26, 39-54.
   (16 pages)

Papers

1. **Jalaliniya, S.**, & Mardanbeigi, D. (2016). EyeGrip: Detecting Targets in a Series of Uni-directional Moving Objects Using Optokinetic Nystagmus Eye Movements. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 5801-5811. 10.1145/2858036.2858584
   (Full paper - 12 pages)

2. **Jalaliniya, S.**, & Mardanbeigi, D. (2016). Seamless Interaction with Scrolling Contents on Eyewear Computers Using Optokinetic Nystagmus Eye Movements. In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16). ACM, New York, NY, USA, 295-298. 0.1145/2857491.2857539
   (Short paper - 4 pages)

3. **Jalaliniya, S.**, & Pederson, T. (2016). Qualitative Study of Surgeons Using a Wearable Personal Assistant in Surgeries and Ward Rounds. eHealth 360: International Summit on eHealth, Budapest, Hungary, June 14-16, Springer International Publishing, 208-219.
(Full paper - 12 pages)

4. **Jalaliniya, S.**, Mardanbegi, D., & Pederson, T. (2015). Magic Pointing for Eyewear Computers. In ISWC '15 Adjunct Proceedings of the 2015 ACM International Symposium on Wearable Computers. (pp. 155-158). ACM. (CIB Proceedings). 10.1145/2802083.2802094
(Short paper - 4 pages)

5. **Jalaliniya, S.**, Mardanbeigi, D., Sintos, I., & Garcia, D. (2015). EyeDroid: An Open Source Mobile Gaze Tracker on Android for Eyewear Computers. In Adjunct UbiComp '15 Proceedings. (pp. 873-879). ACM. 10.1145/2800835.2804336
(Workshop paper - 7 pages)

6. Tell, P., **Jalaliniya, S.**, Andersen, K. S. M., Christensen, M. D., Mellson, A. B., & Bardram, J. (2015). Approximator: Predicting Interruptibility in Software Development with Commodity Computers. In 10th International Conference on Global Software Engineering (ICGSE) 2015. (pp. 90-99). IEEE.
(Full paper - 10 pages)

7. Pederson, T., & **Jalaliniya, S.** (2015). An Egocentric Approach Towards Ubiquitous Multimodal Interaction. In UbiComp/ISWC'15 Adjunct Proceedings. (pp. 927-932). New York, NY, USA: ACM. 10.1145/2800835.2806202
(Workshop paper - 6 pages)

8. **Jalaliniya, S.**, Mardanbeigi, D., Pederson, T., & Hansen, D. W. (2014). Head and eye movement as pointing modalities for eyewear computers. In Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops. IEEE.
(Workshop paper - 4 pages)

9. **Jalaliniya, S.**, Pederson, T., & Houben, S. (2014). Wearable Laser Pointer Versus Head-mounted Display for Tele-guidance Applications? In ISWC '14 Adjunct Proceedings of the 2014 ACM International Symposium on Wearable Computers. ACM. 10.1145/2641248.2641354
(Workshop paper - 8 pages)

10. **Jalaliniya, S.**, Smith, J., Sousa, M., Büthe, L., & Pederson, T. (2013). Touch-less interaction with medical images using hand & foot gestures.

In UbiComp '13 Adjunct Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication. ACM. 10.1145/2494091.2497332
(Workshop paper - 10 pages)

Book Chapter

1. **Jalaliniya, S.**, & Pederson, T. (2015). A Wearable Personal Assistant to Support Clinicians throughout a Workday. Springer International Publishing, Switzerland **(In press - 34 pages)**

Included Papers

1. **Jalaliniya, S.**, & Pederson, T. (2015). Designing Wearable Personal Assistants for Surgeons: An Egocentric Approach. IEEE Pervasive Computing, 14(3), 22-31.
   (Article - 10 pages)

2. **Jalaliniya, S.**, & Pederson, T. (2016). Qualitative Study of Surgeons Using a Wearable Personal Assistant in Surgeries and Ward Rounds. eHealth 360: International Summit on eHealth, Budapest, Hungary, June 14-16, Springer International Publishing, 208-219.
   (Full paper - 12 pages)

3. **Jalaliniya, S.**, Pederson, T., & Houben, S. (2014). Wearable Laser Pointer Versus Head-mounted Display for Tele-guidance Applications? In ISWC '14 Adjunct Proceedings of the 2014 ACM International Symposium on Wearable Computers. ACM. 10.1145/2641248.2641354
   (Workshop paper - 8 pages)

4. **Jalaliniya, S.**, Smith, J., Sousa, M., Büthe, L., & Pederson, T. (2013). Touch-less interaction with medical images using hand & foot gestures. In UbiComp '13 Adjunct Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication. ACM. 10.1145/2494091.2497332
   (Workshop paper - 10 pages)

5. **Jalaliniya, S.**, Mardanbeigi, D., Sintos, I., & Garcia, D. (2015). EyeDroid: An Open Source Mobile Gaze Tracker on Android for Eyewear Computers. In Adjunct UbiComp '15 Proceedings. (pp. 873-879). ACM. 10.1145/2800835.2804336
   (Workshop paper - 7 pages)

6. **Jalaliniya, S.**, Mardanbeigi, D., Pederson, T., & Hansen, D. W. (2014). Head and eye movement as pointing modalities for eyewear computers. In Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops.

IEEE.
(Workshop paper - 4 pages)

7. **Jalaliniya, S.**, Mardanbegi, D., & Pederson, T. (2015). Magic Pointing for Eyewear Computers. In ISWC '15 Adjunct Proceedings of the 2015 ACM International Symposium on Wearable Computers. (pp. 155-158). ACM. (CIB Proceedings). 10.1145/2802083.2802094
(Short paper - 4 pages)

8. **Jalaliniya, S.**, & Mardanbeigi, D. (2016). EyeGrip: Detecting Targets in a Series of Uni-directional Moving Objects Using Optokinetic Nystagmus Eye Movements. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 5801-5811. 10.1145/2858036.2858584
(Full paper - 12 pages)

9. **Jalaliniya, S.**, & Mardanbeigi, D. (2016). Seamless Interaction with Scrolling Contents on Eyewear Computers Using Optokinetic Nystagmus Eye Movements. In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16). ACM, New York, NY, USA, 295-298. 0.1145/2857491.2857539
(Short paper - 4 pages)

10. Bolton, F., **Jalaliniya, S.**, & Pederson, T. (2015). A Wrist-Worn Thermohaptic Device for Graceful Interruption. Interaction Design & Architecture(s), N.26, 39-54.
(Article - 16 pages)

Declaration

This thesis is a presentation of the author's original research. No part of this thesis has been submitted elsewhere for any other degree or qualification. All work is the author's own unless otherwise stated.

Author: Shahram Jalaliniya
Date: 31 January 2016

Copyright ©2016 by Shahram Jalaliniya.
The copyright of this thesis rests with the author. No quotations should be published or information and results derived from this thesis without acknowledgment.

Contents

- Chapter 1 -

Introduction

§ 1.1 Problem statement

Advances in hardware and software technologies has resulted in miniaturizing electronics and emerging new generation of wearable computational and sensory devices, such as functional textiles, smart glasses, smartwatches, and other wearable gadgets. Previous research have shown promising potentials to adopt wearable electronics in a wide range of application areas [111] such as healthcare (e.g. [112, 210]), industry (e.g. [189, 194]), education (e.g. [113]), and maintenance (e.g. [183, 24]). Increasing performance in learning [113], possibility of continuously monitoring patients [112], and reducing human errors [210] are some of the promising findings from previous studies. These findings together with recent advances in wearable technologies have increased expectations on wearable computers to revolutionize the way people use computers in everyday life. In contrast, the application of wearable digital technology in everyday activities is limited to only using smart bands and smartwhatches for mostly self-quantifying applications such as step counters and heal monitoring systems [69]. This is far from what pioneers of wearable computing such as Thad Starner [190] and Steve Mann [118] envisioned for wearable computers as a *constantly* available *assistant* used *in parallel* with real-world tasks [190]. This notion of wearable computer is called *wearable assistant* in this dissertation. Even huge investments by big companies like Google on technological advancement for building unobtrusive smart glasses, as a novel wearable form factor, has not resulted in significant general uptake. This shows that in order to make wearable assistants used in everyday life, it is not enough to only solve technological complexities and work on making devices physically unobtrusive but we need to take a holistic approach including both human-related and technological aspects of interaction into account.

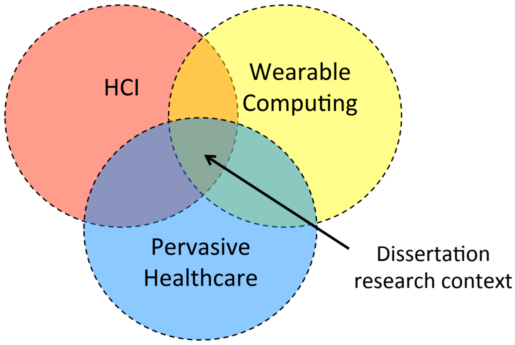


Figure 1.1: *Research context of this dissertation.*

1.1.1 Research context

The research of this dissertation is situated within the area of Human-Computer Interaction (HCI) as an interdisciplinary field about how people interact with computers. More specifically, I focus on HCI challenges of interaction with wearable computers. A wearable computer can be any kind of computing device worn by a user that processes, stores, or communicates data. According to this definition, wearable computers can be found in many different form factors from wearable sensors, to smartwatches and smart glasses. However, the focus of this dissertation is the notion of *wearable assistant* as an always-on device that provides useful information to the user in parallel with performing real-world tasks. Even wearable assistants can be in different form factors such as smartwatches, smart glasses, or even smartphones. In this work, my specific focus is interaction with smart glasses or *eyewear computers* comprised a head-mounted display (HMD) and a processing unit that is connected to the Internet through wireless connection. I define the concept of *Wearable Personal Assistant (WPA)* that represents my conception of such wearable assistants. To concertize the concept of WPA in this thesis, a WPA has been designed, implemented and evaluated for clinicians. This includes pervasive healthcare as another domain to my research context. Figure 1.1 illustrates the research context of this thesis that is situated within *HCI*, *wearable computing*, and *pervasive healthcare*.

1.1.2 Wearable assistants: from vision to reality

One reason the vision of HMD-based wearable assistants have not come true yet, could be the widespread use of smartphones that have been evolved to play the role of personal assistants by providing phone assistance, e-mail assistance, web assistance, and etc. Nowadays, the smartphone is the most common computing device to support users in mobile scenarios by providing useful services such as navigation, transport planning and etc. However, they do not support interaction on the move and concurrent interaction while a user is performing real-world tasks [125]. Existing smartphones, like their predecessors personal digital assistants (PDAs), force users to stop, pick the device from their pocket, and devote at least one hand and full visual attention to the interaction with the device [112]. This form of interaction, interferes with most of the activities people do in the real world. Aside from the physical constraints that holding smartphones creates for mobile users, interaction with computing devices in a dynamic physical environment (e.g. while walking) implies extra cognitive and perceptual load, which can negatively affect the performance of users in the real-world tasks. There is a good potential for wearable computing devices such as smart glasses to solve some of the physical challenges of mobile interaction. Since the display is always in front of the user's eyes, the users do not need to hold the device in their hands. Previous work on designing wearable assistants [112] listed four main issues for implementing wearable assistants:

1. Interaction with the environment through different sensors (context-awareness)

2. Providing hands-free input modalities

3. Acting based on the context without any explicit input by the user

4. Unobtrusive form factor

Most of the state of the art smart glasses such as Google Glass provide the possibility of hands-free interaction which leaves users' hands free for real-world tasks. Furthermore, the unobtrusive form factor and promising functionalities of state of the art smart glasses addresses second and fourth challenges mentioned in previous studies. This inspired me to investigate whether the modern eyewear computers can be good enough platforms for envisioned wearable assistants.

RQ1: *Given the unobtrusive form factor of state of the art smart glasses such as Google Glass and possibility of hands-free interaction, how much closer can the modern smart glasses bring the visionary wearable assistants to reality?*

To answer the above-mentioned question, first of all the concept of wearable assistant needs to be defined (section 2.4-page 24). The idea of *wearable assistant* [37] or *intelligent assistant* has been proposed as an alternative design metaphor to the desktop metaphor by wearable computing community. We all know that desktop metaphor has played the most profound role in evolving computer user interfaces from green-screen terminals to more advanced graphical user interfaces. However, designing for modern mobile and wearable computing devices requires moving beyond desktop metaphor

[132]. Since one of the main assumptions of classical desktop metaphor is the fact that users devote all their perceptual and cognitive resources to the interaction with a desktop computer while in mobile scenarios part of the perceptual and cognitive resources is occupied by the real-world task at hand. But the definition of a wearable assistant provided by wearable computing community [193, 37, 118] stays on a high-level visionary picture [37, 193] without providing a clear step-by-step method for designing a wearable assistant; therefore, it is a big challenge to design a wearable assistant based on only the previous work. Furthermore, taking a user-centered approach, such as participatory design [135], is not easy since the future users of wearable assistants have no idea about the potentials and limitations of the wearable computers. This led me to the second research question in my thesis.

RQ2: *How can we design a wearable personal assistant that accommodates for interaction on the move and parallel interaction with real-world tasks?*

1.1.3 Interruption management

As mentioned in the previous section, moving from smartphone-based assistants to wearable assistants might solve some of the physical constraints of interaction on the move. However, the cognitive and perceptual challenges of concurrent interaction with both real world and computing devices could even be more serious in wearable systems. Since wearable devices are closer to both the body and mind of the "users" than classical personal computing devices ever have been. Therefore, they can more easily distract users' attention from the real world. For instance, ignoring a visual notification on the head-mounted display can be much harder for a user while driving compared to notifications on usual smartphones. This means interruption management is a crucial issue for wearable assistants to avoid negative effects of unwanted interruptions on ongoing tasks [2]. The classical approach to reduce unwanted interruptions in mobile scenarios is to design context-aware systems where the interruptibility of a user is determined based on the user's context including user's activity, surrounding environment, people, and etc. (e.g. [151, 166, 68, 57, 186]). But tracking and modelling physical phenomena in the vicinity of a user have also proven to be very hard indeed. An alternative to the context-awareness approach can be to present notifications in the periphery of user's attention. We know that human attention is a natural mechanism to filter out irrelevant cues to the task in hand. This mechanism has been used in peripheral interaction [15, 67] where information with lower priority targets the periphery of user's attention. This motivates my third research question.

RQ3: *How can we establish a symbiotic interplay between the existing interruption management infrastructure in our brain and wearable personal* *assistant, approaching graceful interruption management?*

1.1.4 Touch-less interaction with wearable assistants

The need for interaction on the move and in parallel with real-world tasks requires wearable assistants to support hands-free interaction techniques. For example, when hands are busy with real world tasks or in sterile environments such as operation theatre, providing touch-less communication channels to the users can be a big advantage for wearable assistants. In fact there is no perfect input modality for wearable assistants that works in all situations. In fact, in mobile scenarios, users usually move from one situation to another one frequently, and each situation requires different input and output modalities. For example, in a classroom voice commands are not the best way of communication, or when a user is performing a manual task, hands are not free for interacting with the wearable assistant. Voice commands and head movements are the only touch-less input modalities supported by state of the art smart glasses such as Google Glass, and Vuzix smart glass. However, HCI community has already explored other touch-less modalities such as eye gaze and body gestures mostly for interaction with desktop computers but less explored for wearable devices. This drives the fourth research question in my thesis.

RQ4: *How can body gestures and eye movements be used as touch-less input modalities for interaction with wearable personal assistants? More specifically, what are the technical and usability challenges of touch-less interaction through body gestures and eye movements?*

As it is mentioned in section 1.1.1, the specific focus of this thesis is interaction with smart glasses or *eyewear computer*. The most important output channel in an eyewear computer is head-mounted display (HMD). While the technological advances in developing unobtrusive and high performance Head-Mounted Displays (HMDs) have opened new horizons for applications of wearable computers in everyday life, fundamental challenges still remain. Eye fatigue, small field of view, swimming effects, limited resolution, and multiple focus planes, are some of the most famous problems associated with HMDs [102, 162]. My last research question is about an alternative visual output for eyewear computers.

RQ5: *Given the known user challenges associated with Head-Mounted Displays (HMDs), is motor-controlled laser pointer technology a viable alternative?*

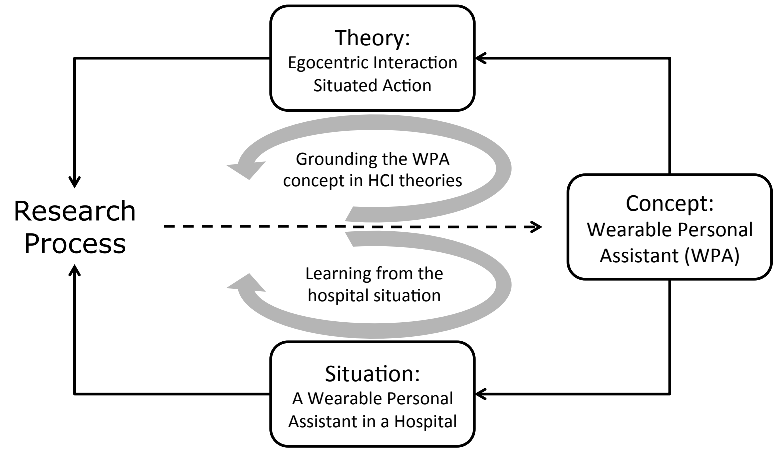


Figure 1.2: *Concept-Driven approach [195] adapted to the wearable personal assistant concept and the hospital setting.*

§ 1.2 Method

To answer the research questions divided and formulated in the previous section, I have followed an approach termed as "concept-driven interaction design research" [195]. The concept-driven interaction design is a complementary approach to more traditional user-centered interaction design approaches such as participatory design. The main difference between concept-driven approach and the traditional user-centered methods is the point of departure in the design process. The traditional approaches are mostly situation-driven which means they find inspiration and restrictions in the (empirical) situation while the concept-driven approach finds inspiration and restrictions in earlier theories and concepts. Moreover, while traditional situation-driven approaches has a client and a problem to solve, the concept-driven approach is an exploratory investigation of previously established theories with the aim of contributing to those theories. This is done by creating artifacts to manifest the theoretical concept in different forms. Through this act of creating artifacts, the designer explores the possible scenarios about interaction that the theoretical concept opens up and, thereby, contributes to the constitute theories about the proposed concept.

The theoretical concept in my thesis is *Wearable Personal Assistant (WPA)*. Therefore, the point of departure in my research is the definition of WPA concept grounded in the HCI theories (the upper cycle in Figure.1.2). The second part is an exploratory process including hands-on design and development of concrete artifacts inspired by futuristic scenarios about possible interaction and design situations (the lower cycle in Figure.1.2). In this dissertation, the design situation is a hospital setting.

1.2.1 Grounding the concept of wearable personal assistant in Human-Computer Interaction theories

To construct the wearable personal assistant concept, first of all, the existing definitions of the wearable personal assistant and other similar concepts are reviewed. Then the relevant HCI theories are used to explain the interaction between the visionary WPA and users (Chapter 2).

1.2.2 Exploring the possible scenarios

Based on the constructed concept of WPA informed by HCI theories and previous work, the possible future scenarios of interaction with the visionary WPA are explored through crafting artifacts that manifest the concept in an experiential and interactive form. This includes designing a wearable laser-pointer as a novel visual output mechanism for the WPA (Chapter 4), exploring body-gestures (Chapter 5) and eye movements (Chapter 6) for touch-less interaction with the WPA, and developing a novel interruption management approach for WPAs (Chapter 7). To investigate the utility and usability aspects of the designed artifacts for future possible scenarios, I mainly conducted lab experiments.

1.2.3 Learning from hospital design situation

Healthcare is one of the most studied application areas for wearable computers. In fact, due to the spatial distribution of departments, wards, meeting rooms, and offices in clinical settings, clinicians move between different locations all the time [17]. Visionary WPAs could support clinicians in mobile scenarios by providing them access to the patient information in different situations. Visionary WPAs could also provide touch-less input modalities in sterile environments. These are good reasons to select healthcare work domain as a design situation for concertizing the WPA concept. A WPA is designed for clinicians through a field study at the Rigshospital in Copenhagen. To evaluate the prototype of the WPA, I conducted a clinical simulation study at ITX simulation facilities of Herlev hospital in Copenhagen. The results of my empirical study in the hospital served as input to advance the interaction theory development, more specifically: the framing of the concept of WPA (Chapter 3).

§ 1.3 Contributions

The main outcome of this thesis falls into three main categories: 1) conceptual, 2) technological, and 3) empirical contributions. The definition of Wearable Personal Assistant is mainly informed by the egocentric interaction paradigm [150], as well as recent findings in cognitive science, perception psychology and neuropsychology. Various scenarios about possible ways of interacting and engaging with the WPA are envisioned, experienced, and enacted through a range of hardware and software prototypes that are evaluated in the lab or in the hospital design situation.

1.3.1 Conceptual

1. ***The Wearable Personal Assistant Concept:*** Apart from visionary descriptions of wearable assistants [193, 37, 118], there is no solid theoretical definition for the wearable personal assistant concept which makes it challenging to design wearable personal assistants. Wearable personal assistants are closer to both the body and mind of the "users" than classical personal computing devices ever have been. Therefore, our definition of wearable personal assistant (section 2.4-page 24) is inspired by egocentric interaction paradigm [150] as a human-body-centric approach. We propose an information flow model which allows us to discuss perception, cognition, and action of users with the specific twist that (inspired by recent findings of cognitive science, perception psychology and neuropsychology) conscious and unconscious cognition is dealt with in separation. This enables us to take a more holistic view of the role of WPAs making it very evident that the explicit interaction taking place between WPAs and user happens in a very information-rich context in which our brains processes much more than we traditionally model as system designers. It allows us also to start speculating about functionalities that could be offered by WPAs that interface directly to the unconscious part of our cognitive processes, something which is undoubtedly still very hard to implement in practice, even if successful attempts have indeed been made, e.g. [45]. (Chapter 2)

2. ***The self-mitigated interruption concept:*** There are many ways in which interruption management could be implemented in a wearable personal assistant. In this thesis, we present our first investigation into using the thermohaptic modality for perceptually and cognitively graceful integration of notifications originating from digital services with ongoing tasks that the human agent is performing. The main idea behind the mechanism which we call *self-mitigated interruption* is to let the existing supervisory attentional system play an important part in deciding whether a notification is to interrupt the current primary task or not, instead of primarily let the wearable personal assistant rely on sensing and interpreting the context and then rather brutally call for attention. In the self-mitigated interruption the system knows the importance of the notification; therefore, the system can adjust the intensity of the stimuli according to the importance of the notifications.

If a user is busy with real world tasks, a low-priority message with lower intensity has a lower chance to interrupt the user compared to an important notification with high intensity. (Chapter 7)

1.3.2 Technological

1. ***A multimodal WPA for orthopedic surgeons on the Google Glass Platform:*** To evaluate the utility of Google Glass as a state of the art eyewear computer for wearable personal assistants, I designed and implemented a multimodal system on the Glass platform to support orthopedic surgeons throughout a workday. The system includes three main modules: 1) touch-less interaction with X-ray images; 2) tele-presence; and 3) Electronic Patient Records (EPR). To develop the system, all possible modalities on the Glass such as voice commands, touch gestures, and head movements are used to support users in different situations. Also to investigate the utility of the touch-less interaction module, I modified an open-sources medical system (Invesalius) for viewing 3D and 2D medical images. The modified system is able to communicate with the Glass application to adjust the view of medical images. Furthermore, to set up the remote collaboration scenario, I developed another application on an Android Tablet which is able to call or receive calls from the Glass application to facilitate a vocal and visual communication between the local surgeon (the Google Glass user) and a remote person (the Tablet user). The Tablet user is able to draw sketches on the image that is superimposed in real-time on the Glass head-mounted display. Integration of all these systems on different platforms is necessary to evaluate the WPA concept in a real scenario. (Chapter 3)

2. ***A wearable system-controlled laser pointer:*** One of the functionalities envisioned for WPAs is to direct users' attention towards particular objects in the real-world. Due to the limitations of existing technologies, I designed and implemented a motor-controlled wearable laser pointer which is able to direct a laser beam in vertical and horizontal directions by rotating two mirrors. Since the motors can move very fast (20KHz) the system can draw sketches or texts superimposed on real-world objects. My prototype is the first wearable version of such galvo-scanner-based laser pointers. (Chapter 4)

3. ***An open-source client on Google Glass for gaze-based interaction (GlassGaze):*** eye gaze is among less-explored input modalities for wearable computers. In fact there is no eye tracking technology available on state of the art eyewear computers. To explore eye movement as a touch-less input modality for wearable personal assistants, we developed and released the first open-source eye tracking hardware

and software module for Google Glass [123]. The software system comprises a client application on the Google Glass platform [123], and a remote gaze tracking server on a Windows machine [121]. The hardware module consists of a remote infrared camera which sends the eye image over wireless connection to the remote server. The remote server analyses eye image and sends back the gaze coordination to the Google Glass clinent through wifi connection. (Chapter 6)

4. ***A gaze tracker server on Android phone (EyeDroid):*** The gaze tracker client we developed for Google Glass lets the user be mobile in the range of wireless network between infrared camera and the remote gaze tracking system. To increase the mobility of the user we had to remove the dependency of the system on the remote server. Therefore, we developed the server side of the gaze tracking system on an Andorid device [80]. The new gaze tracking system includes an infrared camera that connects to the Android smartphone through a USB cable. The eye image is analysed by the Android application (EyeDroid) and the gaze coordination is sent to the Google Glass client through a WIFI connection. (Chapter 6)

5. ***A wearable thermohaptic device for self-mitigated interruption:*** We developed and evaluated a prototype of a wrist-worn thermal haptic system able to provide notifications to the person wearing it. Temperature differences are produced by the Peltier principle by passing electrical current through a Thermoelectric cooler. The objective of this prototype is to generate wearable thermal haptic notifications whilst maintaining a somewhat realistic wearable device weight. The wrist device holds a micro-controller board (Arduino) and four thermistors (to measure the hot and cold sides of the Peltier, the skin temperature of the person wearing the device, and the room temperature). Thermal contact with the skin is made by a anodised aluminium heat spreader. Residual heat is dissipated with a heat sink fitted to the back side of the Peltier. (Chapter 7)

6. ***Hand & foot gesture recognition using wearable inertial senor & capacitive floor sensor :*** We designed and implemented a system for gesture-based interaction with medical images using a wearable inertial sensor to detect hand gestures and capacitive floor sensors for foot-gesture recognition. The wearable sensor [65] sends the acceleration and rotation data to a remote server through a wireless connection (ANT protocol [71]). I developed a driver for ANT-based communication between the wearable sensor and a Windows machine during my 3-month visit at ETH Wearable Computing Lab. The remote server runs a MATLAB application to detect the hand gestures using a Neural Network classifier. The same server receives and analyzes

the data from capacitive floor sensors to detect foot gestures. In our implementation, the foot gestures are used as a clutch mechanism for activating/deactivating medical image systems to receive hand-gesture commands. Also the foot gestures are used to switch interaction mode between different systems in the operation theatre. (Chapter 5)

7. ***EyeGrip: a calibration-free method for seamless interaction using eye movements*** EyeGrip is a novel eye movement-based interaction technique for seamless interaction with the scrolling contents on a computer screen. EyeGrip shows the potential use of optokinetic nystagmus (OKN) for eye-based interaction with dynamic user interfaces. In the EyeGrip method, we analyze the saw-teeth like OKN signal to detect which object among other sequential objects draws user's attention. System uses this information to automatically stop the scrolling contents on the screen without any additional explicit input. We developed two different classifiers to detect objects of interest: 1) a Neural Network-based classifier, 2) a threshold-based classifier that measures the length of smooth pursuits in the OKN signal and compares it with a defined threshold. We investigated the influence of speed and maximum number of visible images in the screen on the accuracy of the system. We also empirically evaluated the utility of the proposed interaction technique through implementation and evaluation of four different applications: 1) a menu scroll viewer on eyewear computers and 2) a Facebook newsfeed reader on a HMD, 3) a mind reading game, and 4) an image explorer system for desktop computers. (Chapter 6)

1.3.3 Empirical

1. ***A framework for designing wearable personal assistants:*** I started the empirical study, by conducting an ethnography in a hospital. I tried to observe and understand the nature of the work in the hospital, and I decided to focus on orthopedic surgeons since in the hospital that I have been studying, they are among the most mobile clinicians. I shadowed a surgeon for a whole day and I observed them in different situations from operation theatre to wards and X-ray conferences. But to design a WPA, it is challenging to take a situation-driven approach such as participatory design where users play the main role in the design process. Since users do not have a clear idea about potential scenarios where a WPA can be used. In such a design space, you as designer should come up with some design initiatives and drive the design process. But the question is how the initial design ideas can be found. How can the designer be sure if he/she has covered all possible scenarios at the initial step? This is the main reason why I went

through a concept-driven interaction design research where the point of departure is the WPA concept. I developed a design framework for a WPA to support clinicians and in particular orthopedic surgeons. In the proposed design framework which is basically a 2 dimensional matrix, the main characteristics of hospital work are core elements (rows) while the columns correspond to the three kinds of assistance defined based on the WPA concept. By crossing rows and columns, I developed an analytical framework where in each cell I had to answer the question: what kind of assistance (e.g. perception assistance) is needed to support each aspects of the work (e.g. collaboration). Taking this approach, I proposed 12 initial ideas for the functionalities of the WPA for orthopedic surgeons. The feasibility and usefulness of the initial 12 ideas have been discussed with three orthopedic surgeons in semi-structured interviews and based on the interview results, we focused on three main ideas: 1) touch-less interaction, 2) tele-presence, and 3) mobile access to the patient records (EPR). (Chapter 3)

§ 1.4 Dissertation Overview

The dissertation is organized into two main parts. Part I provides an overview of the conceptual background of the work and description of the system prototypes. Figure 1.3 illustrates mapping between the structure of the chapters in Part I to my research method explained in section 1.2. The arrows in Figure 1.3 do not show the sequence of steps in the method. They only illustrate the direction of the two iterative processes in my research. Each chapter in Part I consists of a related work section providing a big picture of the research context, a brief description of my contributions to the introduced research context, and discussing how my contributions fit into the WPA conceptual model and how they address some of the challenges of the WPA in hospital. Part II consists of a collection of 10 papers that contain concepts, technology and user studies summarized and discussed in Part I. To maintain the consistency, readability and presentation of the original publications, the papers are included in their publicly available published formats.

Part I consists of 7 chapters:

Chapter 2 *The Wearable Personal Assistant Concept*
In this chapter, the concept of wearable personal assistant is defined based on the previous work, the egocentric interaction paradigm, and recent findings from cognitive science.

Chapter 3 *A Wearable Personal Assistant for a Hospital Setting*
In this chapter, I briefly discuss the results of an empirical study

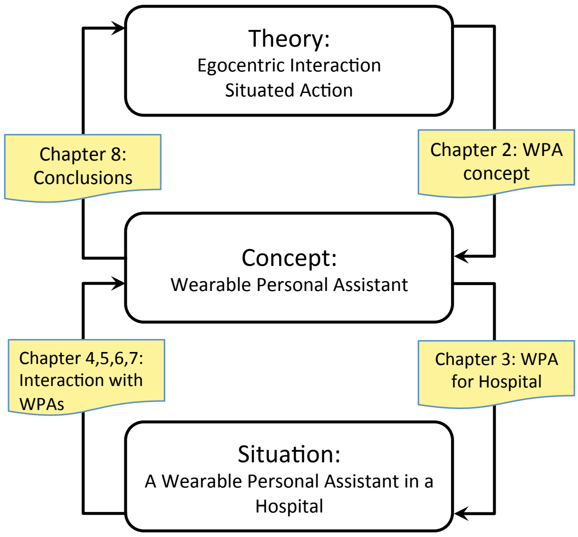


Figure 1.3: *Mapping the structure of the dissertation to the research method (concept-driven approach).*

on design, implementation, and evaluation of a wearable personal assistant for orthopedic surgeons in a Danish hospital.

Chapter 4 *A Wearable laser pointer as a visual display for WPAs*
This chapter compares different technologies of visual output such as head-mounted display, wearable projector, contact lens displays, etc. It also introduces my wearable laser pointer as a novel alternative.

Chapter 5 *Body gestures as an input modality for WPAs*
This chapter discusses technological and usability aspects of using body (hand, foot, and head) movements as a touch-less modality to provide discrete and continuous commands to a wearable personal assistant.

Chapter 6 *Exploring Eye Movements for Interaction with WPAs*
This chapter is about utility of using eye movement as an explicit or implicit input for interaction with wearable personal assistants. It also discusses technological and usability challenges of detecting and transforming eye movements for interaction with eyewear computers as a platform for wearable personal assistants.

Chapter 7 *Interruption Management & WPAs*
This chapter describes the previous work on interruption management. It also introduces our self-mitigated interruption concept as a novel approach.

Chapter 8 *Conclusions & Future Work*

This chapter concludes my findings in this thesis and proposes some future research directions.

Part II consists of 10 papers:

Paper 1 *Designing Wearable Personal Assistants for Surgeons: An Egocentric Approach* (published in IEEE Pervasive Computing Magazine, July 2015)
In this paper, we propose a human body-and-mind centric (egocentric) design framework (as opposed to device-centric) and present initial findings from deploying it in the design of a wearable personal assistant for orthopedic surgeons. The result is a Google Glass-based prototype system aimed at facilitating touchless interaction with X-ray images; the browsing of Electronic Patient Records (EPR) on the move; and synchronized ad-hoc remote collaboration.

Paper 2 *Qualitative Study of Surgeons Using a Wearable Personal Assistant in Surgeries and Ward Rounds* (published in proceedings of eHealth 360: International Summit on eHealth 2016)
In this paper, we report on the utility of a wearable personal assistant (WPA) for orthopedic surgeons in hospitals. A prototype of the WPA was developed on the Google Glass platform for supporting surgeons in three different scenarios: (1) touch-less interaction with medical images in surgery room, (2) tele-presence colleague consultation during surgeries, and (3) mobile access to the Electronic Patient Records (EPR) during ward rounds. We evaluated the system in a simulation facility of a hospital with two real orthopedic surgeons.

Paper 3 *Wearable Laser Pointer Versus Head-Mounted Display for Tele-Guidance Applications?* (published in adjunct proceedings of ISWC '14)
In this paper, I introduce a wearable motor-controlled laser pointer and quantitatively determine if wearable laser pointers are viable alternatives to Head-Mounted Displays for indicating where in the physical environment the user should direct her/his attention.

Paper 4 *Touch-less Interaction with Medical Images Using hand & Foot Gestures* (published in adjunct proceedings of Ubicomp '13)
In this paper, we present a system for gesture-based interaction with medical images based on a single wristband inertial sensor and capacitive floor sensors, allowing for hand and foot gesture input.

Paper 5 *EyeDroid: An Open Source Mobile Gaze Tracker on Android for Eyewear Computers* (published in adjunct proceedings of ISWC '15)
This paper is about development and evaluation of a video-based mobile gaze tracker for eyewear computers. Unlike most of the previous work, our system performs all its processing workload on an Android device and sends the coordinates of the gaze point to an eyewear device through wireless connection. We propose a lightweight software architecture for Android to increase the efficiency of image processing needed for eye tracking.

Paper 6 *Head and eye movement as pointing modalities for eyewear computers* (published in adjunct proceedings of BSN '14)
In this paper, we examined using head and eye movements to point on a graphical user interface of a wearable computer. The performance of users in head and eye pointing has been compared with mouse pointing as a baseline method.

Paper 7 *MAGIC Pointing for Eyewear Computers* (published in proceedings of ISWC '15)
In this paper, we propose a combination of head and eye movements for touch-lessly controlling the "mouse pointer" on eyewear devices, exploiting the speed of eye pointing and accuracy of head pointing. The method is a wearable computer-targeted variation of the original MAGIC point- ing approach which combined gaze tracking with a classical mouse device.

Paper 8 *EyeGrip: detecting targets in a series of uni-directional moving objects using optokinetic nystagmus eye movements* (published in proceedings of CHI '16)
In this paper, we introduced EyeGrip as a novel and yet simple technique of analysing eye movements for automatically detecting the users objects of interest in a sequence of visual stimuli that is moving horizontally or vertically in front of the user's view. We assess the viability of this technique in a scenario where the user looks at a sequence of images that is moving horizontally in the display while his/her eye movements are tracked by an eye tracker. We investigated the influence of speed and image width on the accuracy of the system. We also demonstrated the rich capabilities of EyeGrip with two example applications: 1) a mind reading game, and 2) an image explorer system.

Paper 9 *Seamless interaction with scrolling contents on eyewear computers using optokinetic nystagmus eye movements* (published in proceedings of ETRA '16)
In this paper, we report on using EyeGrip for seamless interaction with scrolling contents on eyewear computers. We empirically

evaluated the usability of EyeGrip through implementation and evaluation of two different applications for eyewear devices: 1) a menu scroll viewer and 2) a Facebook newsfeed reader.

Paper 10 *A Wrist-Worn Thermohaptic Device for Graceful Interruption* (published in Interaction Design & Architecture(s) Journal December 2015)
In this article, we present our self-mitigated interruption concept (essentially a symbiosis of artificial external stimuli tuned to existing human attention management mechanisms) and perform a first exploratory study using a wrist-worn thermohaptic actuator. We frame our empirical thermohaptic experimental work in terms of Peripheral Interaction concepts aimed at supporting the design of envisioned Wearable Personal Assistants intended to, among other things, help human perception and cognition with the management of interruptions.

Part I

Research Contextualization

- Chapter 2 -

The Wearable Personal Assistant Concept

In this chapter, first of all the previous work and history of wearable assistants is briefly reviewed. Then the current state and initiatives for the future of wearable assistants is discussed. Finally, I introduce our Human-body-and-mind-centric (Egocentric) approach to WPA. There is an overlap between the content of this chapter and Paper 1 in Part II; however, in this chapter there are more details about our Human-body-and-mind-centric approach to the WPA compared to Paper 1.

§ 2.1 Beyond the desktop metaphor

The desktop metaphor has been the earliest and probably most profound concept determining our present day experience of computer systems. The first general-purpose systems for digital work have emerged based on the desktop metaphor in the early 1980s [184]. Desktop systems provided a digital work environment to support an individual user of a stand-alone computer in an office environment mainly for storing and retrieving documents. One of the main goals of introducing the desktop metaphor was to bring the computers out of the research labs where users knew how to interact with computers mostly through text commands. The desktop metaphor enabled non-specialist users to interact with the computers through graphical user interfaces. It also helped both users and the designers of interactive desktop systems to have a common understanding of interaction with desktop computers. In fact using the desktop metaphor as a unifying framework (see Figure 2.1) between designers and users of interactive systems has proved to be a huge success.

However, computing technologies are evolving rapidly; people often use a range of computing devices, such as laptop computers, tablets, smartphones, and even wearable devices. Interaction with these new device ecologies requires integrating information across multiple users, applications, and

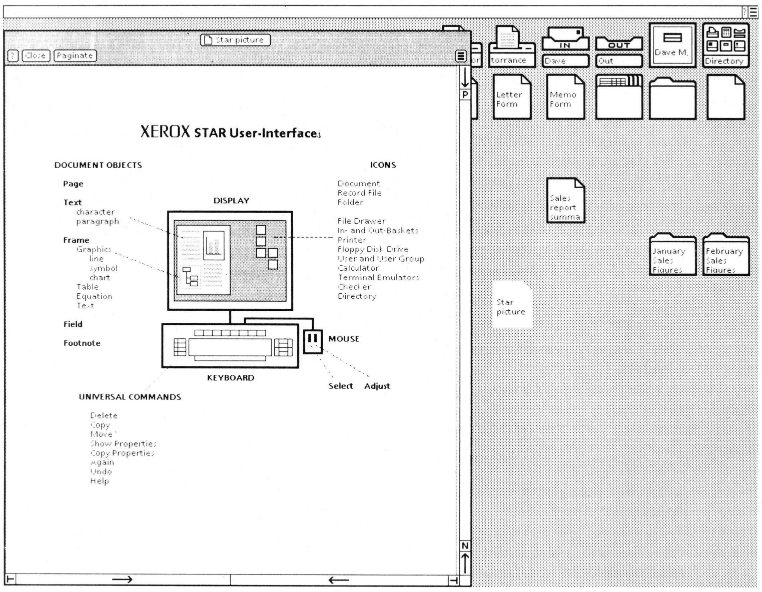


Figure 2.1: *Xerox introduced the desktop metaphor for interaction with personal computers. This figure shows a "Desktop" as it appeared on the Star screen [184].*

devices. This means that one "desktop" is no longer enough. Moreover, sometimes people use their mobile devices on the move e.g. to find a particular location while interacting with a location-aware system on their smartphones or buying an electronic ticket for public transportation while walking towards a bus station. There is no natural place for "desktops" in such scenarios. In short, these new ways of using computers violates many assumptions of the desktop metaphor where an individual user devotes his/her full attention to the interaction with a stand-alone computer. Empirical studies (e.g. [154, 90, 46, 42]) have shown that traditional desktop systems do not effectively support collaboration, multitasking, and multi-device configurations. This means that in order to design systems that are able to support collaborative and mobile work across multiple devices we need to move beyond the desktop metaphor [132].

Different approaches have been proposed as alternatives to the desktop metaphor [132]. Some of them [58, 91] have introduced new ways of accessing to the personal information that in traditional desktop systems is stored in folders. For example, Lifestream [58] introduces a time-ordered stream of documents, as an underlying storage system for dynamically organizing personal data. Another example is Haystack [91] which is an information management tool. Haystack enables end users to define properties for information objects and relationships between them using a semistructured data model. Robertson et al. [161] introduced GroupBar, Scalable Fabric, and Task Gallery to modify the existing task management mechanisms in Microsoft Windows for a smoother and easier switching between tasks by users. *Activity-Based Computing* [16] is another alternative concept that proposes the notion of *activity* (versus application and device) as a conceptual tool for designing integrated digital environments. Activity-Based Computing allows users to preserve continuity in their work when switching between different computing devices.

§ 2.2 The Personal Assistant metaphor

In line with the other post desktop metaphor research, the wearable computing community [37] adopted an alternative interaction metaphor for wearable systems: The Personal Assistant. Based on the Personal Assistant metaphor, a wearable system should act like a human personal assistant:

"A good (human) personal assistant is unobtrusive, predicts what information is needed and prepares it in anticipation of its need, schedules meetings and appointments" [37]

However, there is no consensus term or generally accepted definition for the Wearable Personal Assistants. For instance, Thad Starner [193], one of the pioneers of wearable computing, uses *Wearable Intelligent Assistant* term for his envisioned general-purpose wearable system. While some others use the Personal Assistant metaphor versus the desktop metaphor [37].

The idea of designing wearable systems like a personal assistant is mainly inspired by earlier studies on intelligent interface agents [115]. The interface agents (1) observe the user's actions and imitate them, (2) receive user feedback on the systems' actions, and (3) learn from users feedback. Based on this approach a *personal digital assistant* could help users by handling emails, scheduling meetings, filtering news, or recommending books or music [114].

Steve Mann, another pioneer in wearable computing, defines eight attributes for his envisioned wearable computer [118]:

1. Constant (always ready)
2. Unrestrictive (you can do other things while using it)
3. Unmonopolizing of the user's attention (Users can attend to both physical world and the wearable system)
4. Observable by the user (it should be able to notify the user)
5. Controllable by the user (The user can grab control of the system at any time)

6. Attentive to the environment (context aware, multimodal, multisensory)

7. Communicative to others (can be used as a medium of expression or as a communications medium)

8. Personal (the user and the system are inextricably intertwined)

Thad Starner's *wearable intelligent assistant* "augments memory, intellect, creativity, communication, and physical senses and abilities" [190]. He defines four ideal attributes for his envisioned intelligent assistant as follows: [190]

1. *Persist and provide constant access to information services (designed for everyday and continuous use)*

2. *Sense and model context (the wearable must observe the user's environment, the user's physical and mental state, and its own internal state)*

3. *Adapt interaction modalities based on the user's context (the wearable should adapt its input and output modalities automatically to those that are most appropriate)*

4. *Augment and mediate interactions with the user's environment (the wearable should provide information support in both the physical and virtual realms)*

Thad Starner's definition of the wearable intelligent assistant is mainly inspired by the concept of a *cyborg* or a *man-computer symbiosis* [38] where the futuristic human is envisioned to be tightly coupled with machines. The result of such a partnership is imagined to think and process data in a completely different way [107].

Both Thad Starner's and Steve Mann's definitions mentioned above stress the vision of a wearable assistant supporting users in mobile scenarios and in parallel with real-world activities by providing relevant information to the task at hand through an appropriate modality. To understand what makes wearable assistants special for mobile scenarios, we can compare them with smartphones and their predecessors Personal Digital Assistants (PDAs) which are also designed to support mobile users. We can notice that wearable assistants (in theory) should have an advantage over PDAs/smartphone-based assistants in cases of interaction on the move and parallel interaction. Wearable computers are physically closer to the human body more than any other computing device ever. This reduces the time between intention and action for wearable computers [191]. To interact with a smartphone, the user needs to pick it from his/her pocket that increases the time between intention and action. Moreover, touch-based user interface of smartphones and PDAs requires full visual attention and significant hand-eye coordination which cuts the user's attention from surrounding environment. While in wearable computers, there are more possibilities for microinteractions [9] (interaction that takes less than 4 seconds) through head-nudge or voice commands.

But the question is that given these great potentials of wearable assistants for supporting users in mobile scenarios, why are they not present in everyday life as smartphones are?

§ 2.3 Why wearable assistants are not pervasively used

Obviously, all above-mentioned requirements for ideal wearable assistants make them very hard to design. Some important technological and interaction-related challenges are discussed by Starner [190]: power use, heat dissipation, and networking are among the main technological challenges of building wearable assistants while privacy and interface design can be seen as interaction-related issues. But is the lack of widespread use caused only by the engineering challenges involved in packaging something quite big and stationary (the desktop PC) into something smaller and unobtrusive? The smartphone is maybe the optimal solution to this challenge and the need to look further is minimal.

Previous studies have shown the challenges of interaction with smartphones on the move [125]. First of all, physical constraints raised by holding smartphones in the user's hands and providing touch input interferes with real-world tasks where the user most of the time needs his/her hands. Secondly, limitations of human perceptual and cognitive resources makes it impossible for users to attend to more than one task at the same time. Most of the applications on smartphones require users' full attention for interaction which interferes with moving and performing real-world tasks. Therefore, apparently current user interfaces of smartphones do not support users' mobility, and there is a need for alternative technologies such as wearable assistants.

For a long time, the general picture of wearable computers was a bulky, cumbersome, and uncomfortable computer worn only by computer gigs. But recent advances in hardware and software technologies resulted in emerging relatively unobtrusive eyewear computers such as Google Glass and M100 Vuzix Smart Glass. These devices are equipped with a wide range of hardware technologies: a head-mounted display, a high definition front-view camera, motion sensors, microphones, headphones, WIFI and Bluetooth connections, and etc. These hardware and software advancements solves at least two important issues addressed in the early studies on wearable computers

[112]: 1) unobtrusiveness of wearable computers, and 2) supporting hands-free interaction. These two factors are the main drivers for my first research question RQ1 (page 3) to investigate whether now is the time for wearable assistants to take off.

§ 2.4 Wearable Personal Assistant (WPA)

The concept of wearable assistant in this thesis is very in line with the *Intelligent Assistant* defined by Thad Starner [190]. However, I use the term *Wearable Personal Assistant (WPA)* instead to emphasize the tight integration between a single mind, body, and computer as Steve Mann [118] also mentioned *personal* as the 8th attribute of a wearable computer.

The idea of integrating computers with the human brain might sound unrealistic when Thad Starner [190] used the concept of a cyborg or a man-computer symbiosis to describe an ideal wearable assistant. But nowadays, we can find several examples of such systems that show the practicality of the *man-computer symbiosis* idea. For instance, the BrainPort system [44], as a sensory substitution device, helps visually impaired people to see the world through their tongue. The BrainPort system captures the real world through a front-view camera and converts the captured image to an electrotactile image on a tongue display. Another example is a vibrotactile sensory substitution vest [139] which converts auditory information to haptic through small vibratory motors distributed on a vest. The vibrotactile vest helps deaf people hear surrounding environment through their skin. Furthermore, researchers in the field of neuroscience are trying to use this vibrotactile vest for communicating more complex information e.g. stock market data directly to the user's brain (unconsciously) through the vest [140]. These examples illustrate the flexibility of the human brain in perceiving the world through computing devices.

§ 2.5 Human body-and-mind centric (Egocentric) approach to Wearable Personal Assistants

Wearable Personal Assistants (WPAs) are intended to extend the user's body and mind. It is close at hand therefore to adopt a human body-and-mind centric design approach, such as the one adhered to in this thesis: Egocentric Interaction [149]. This approach complements existing largely technology driven efforts in addressing challenges of designing Wearable Personal Assistants.

Our study on developing a WPA for clinicians (Chapter 3) revealed the fact that the main bottleneck for what can be computed by combined human and computer systems stem from the limitations of human perception, cognition, and action, *not* in limitations of current computer hardware and software. The human-body-and-mind-centric approach helps us as designers to make sure that our WPAs talk to our pretty static biological setup in the way it was designed by evolution to interpret and act in the world.

§ 2.6 Egocentric Interaction

The Egocentric Interaction Paradigm extends and modifies the classical user-centered approach in HCI[138], on several points, including:

1. Situatedness. Acknowledges the primacy of the agent's current bodily situation at each point in time in guiding and constraining the agent's behavior.
2. Attention to the complete local environment. Makes it a point to take the whole environment into consideration, not just a single targeted artifact or system.
3. The proximity principle. Makes the assumption that proximity plays a fundamental role in determining what can be done, what events signify, and what agents are up to.
4. Changeability of environment and agent–environment relationship. Takes into account agents' more or less constant movements of head, hands, sense organs, and body.
5. The physical-virtual equity principle. It pays equal attention to both interaction with virtual (digital) objects (classical HCI) and physical objects (classical ergonomics).

The term "egocentric" signals that it is the human body and mind of a specific human individual that (literally) acts as center of reference to which all modeling is anchored in this interaction paradigm. The term is analogously often used in for instance psychology and virtual reality (e.g. –[156]) to denote the conceptual and spatial frames of reference which humans by necessity rely on when thinking and acting in the world, also when collaborating with others. [149]

2.6.1 Action & Perception instead of Input & Output

In the egocentric interaction paradigm, the modeled human individual needs to be viewed as an agent that can move about in a mixed-reality environment (an environment consisting of both directly accessible everyday "real" entities and virtual/digital objects accessed through mediating digital devices), not as a "user" performing a dialogue with a computer. Adopting

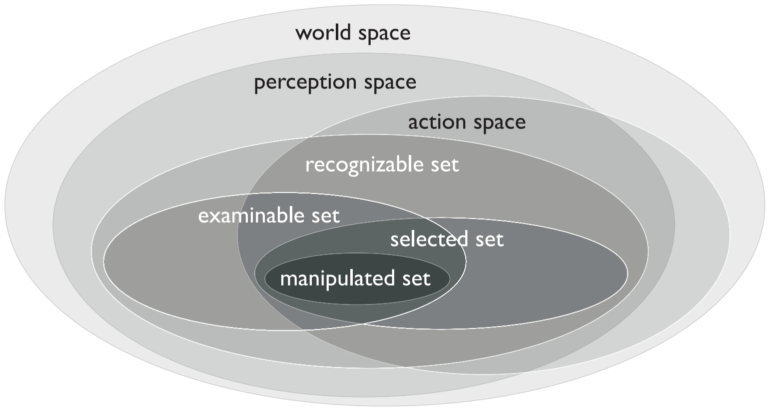


Figure 2.2: *The Situative Space Model. [150].*

the physical-virtual equity principle mentioned before, the Egocentric Interaction paradigm substitutes the concepts of (device) “input” and “output” with (human agent) “action” and “perception”.

2.6.2 The New "Desktop"

In order to facilitate the design of egocentric interaction systems, Pederson et al. [150] have proposed a situative space model (SSM) to capture what a specific human agent can perceive and not perceive, reach and not reach, at any given moment in time (see Figure 2.2). This model is for the emerging egocentric interaction paradigm what the virtual desktop is for the PC/WIMP (Window, Icon, Menu, Pointing device) interaction paradigm: more or less everything of interest to a specific human agent is assumed to, and supposed to, happen here. Only the perception and action spaces are described here and the reader can find more details elsewhere [150].

- *Perception Space:* The part of the space around the agent that can be perceived at each moment. Like all spaces and sets of the SSM, it is agent-centered, varying continuously with the agent's movements of body and body parts. Different senses have differently shaped perception spaces, with different operating requirements, range, and spatial and directional resolution with regard to the perceived sources of the sense data. Compare vision and hearing, e.g.

- *Action Space:*The part of the space around the agent that is currently accessible to the agent's physical actions. Objects within this space can be directly acted on. The outer range limit is less dependent on object type than perception space and is basically determined by the physical reach of the agent.

§ 2.7 The Perception-Cognition-Action Loop

While the Situative Space Model helps in the analysis of spatial relationships between nearby entities and a given human agent, it says nothing about temporal aspects. And while it indicates the availability and suitability for a WPA to make use of a given modality and device for interacting with the user, it does not say anything about timing. E.g. when is a good time to play the not-so-important message notification beep such that it does not interfere with ongoing tasks? The perception-cognition-action flow model developed as part of this thesis (Figure 2.3) intends to inform such WPA system design decisions. Figure 2.3 shows a very simplified model of information flows occurring as result of a human agent acting in the world. The purpose of this model is not to provide a completely true account but a good-enough model for designing future interactive systems. [81]

2.7.1 Perception

By and large, our perception of the world (pathway 2-4) and our perception of our body state (arrow 5) is beyond our conscious control. However, conscious cognitive processes influence unconscious processes (arrow 7), as in the case when we deliberately address our attention to a certain speaker in a crowd and we automatically (thanks to subconscious processing), to some degree, can single out the voice we are interested in. We can also consciously and indirectly affect unconscious processing by orienting our body sensors (e.g vision) towards phenomena of interest (pathway 8-10-2-4).

2.7.2 Cognition

In our model, human cognition is divided into unconscious and conscious processing (arrow 12 and 13 in Figure 2.3 respectively), receiving input from sensors capturing in-body phenomena (e.g. proprioceptive information about limb positions; information for maintaining homeostasis) and from sensors capturing information from the external world. No world phenomena or in-body phenomena is subjected to conscious cognitive processing before having been unconsciously processed (pathways 2-4-6 and 5-6 respectively).

Conscious processing is slower than the unconscious. For instance, muscular reactions to immediate threats are initiated subconsciously (pathway 4-9) long before conscious processes are engaged. We protect our faces with our hands "instinctively" from approaching projectiles like hockey pucks even when consciously aware of the fully protective shields of transparent material in front of us.

Dividing cognition to conscious and unconscious processing helps us explain two different types of interaction between human mind and the external world (including WPAs). There is a wealth of knowledge about how

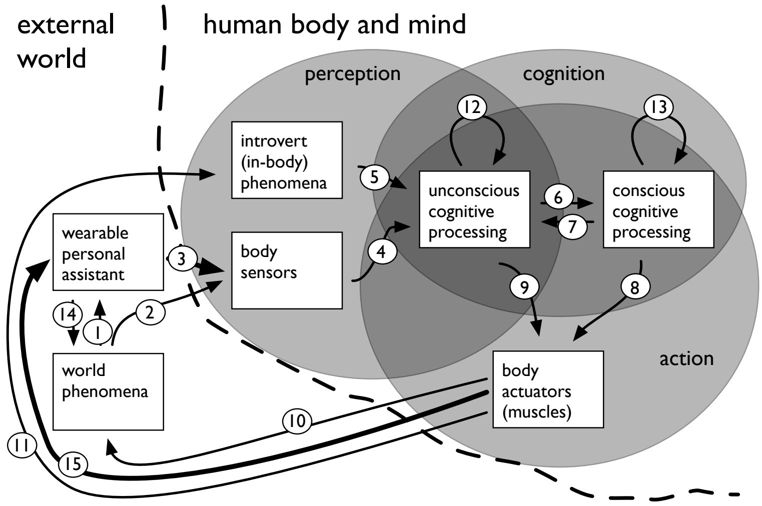


Figure 2.3: *A body-and-mind centric model of how Wearable Personal Assistants fit into the flow of perception, cognition, and action of human agents.*

humans function at various level of abstraction from low-level unconscious signal processing to high level conscious attention. As system designers, we should draw on this knowledge to design WPAs able to interact with us both consciously and unconsciously. As Thad Starner also pointed out, the interface of the WPAs can be a natural extension of the user's body that would not necessarily require conscious attention [190].

2.7.3 Action

Human action is initiated and controlled by a mix of conscious and unconscious cognitive processes. An example of an activity mostly driven by an *unconscious* perception-cognition-action loop could be walking along a well-known road with no exposure to obstacles (pathway 2-4-9-10 & 11). An example of activity that makes use of a combination of conscious and unconscious cognition could be when attempting to direct the tip of a thread through the hole of a needle, demanding focused visual and tactile conscious attention (pathway 2-4-6-8-10) in parallel with unconscious detailed control of hand, finger, and arm muscles (pathway 5-9-11).

§ 2.8 The Wearable Personal Assistant in the Loop

In this section we will discuss the role of a WPA in the perception-cognition-action loop shown in figure 2.3. When appropriate, we will use Google Glass as a concrete representative of WPAs in order to examplify interaction mechanisms.

By including unconscious cognitive processing and by including all perceivable world phenomena (including everyday objects such as coffee cups and footballs), the model shown in Figure 2.3 allows us to get a more complete perspective of the context where a WPA operates compared to classical HCI models typically concerned only with pathway 15-3 (bold arrows in the figure) and how it relates to conscious cognitive processing. It becomes evident that any information generated by a WPA (arrow 3) is just *one* source of information among many others that hit the unconscious and conscious parts of our brains which together do their best to make sense of it all. Heuristics for serving that information timely, arriving at successful "attention management", is probably best based on knowing what else is hitting the senses in parallel. The Context Aware systems community has investigated this for years but often using a system- or device-centered approach. We believe that a human-centric approach towards determining what matters in a given situation (e.g. using the SSM, Figure 2.2) will reveal interesting complementary information.

Note: Conceptually, the hardware user interface of a WPA (e.g. the HMD, microphone, loudspeaker, and touchpad on Google Glass) receives input from the human agent, and provides output which the human agent can sense, in the shape of "world phenomena" (arrows 10 and 2). In figure 2.3 these information flows have been re-represented as separate information flows (arrows 15 and 3), only to facilitate the discussion below.

2.8.1 Implicit Input and Output

Although seldom very clearly defined in HCI literature, the distinction between explicit and implicit input and output [170] is useful for discussing some important properties of WPAs:

- *Explicit input: action intentionally and consciously directed towards the WPA.* Example: the human agent navigates the GUI presented on the Google Glass HMD by swiping the touch area of Google Glass (pathway 8-15).

- *Implicit input: action performed by human agent without the conscious intention of communicating with the WPA.* Example: The human agent acts in the real world (moves about; manipulates objects; interacts with human agents), partially sensed by the WPA (pathways 9&8-10&11). In this thesis (Chapter 7), I introduce EyeGrip [76] as an implicit input for the eyewear computers. EyeGrip uses Optokinetic Nystamus eye movements which is an involuntary reflection of the eyes elicited by the tracking of a moving field.

- *Explicit output: WPA addressing the conscious.* The WPA creates a change in the perception space of the agent (Figure 2.2) which the human agent cannot avoid consciously perceiving (3-4-6), inviting the human agent to act.

- *Implicit output: WPA addressing the unconscious.* The WPA generates a phenomena in the perception space of the agent (Figure 2.2) that reaches the unconscious part of cognition (pathway 3-4) but *not* the conscious part (pathway 6), e.g. through ambient displays such as the "Dangling String" by Jeremijenko [206]). In this thesis (Chapter 4), I introduce self mitigated interruption concept which is about displaying a stimuli at a certain intensity which can be perceived if the user is cognitively free and not perceived if the user is cognitively busy.

By having actuators and sensors placed on or very close to the body, and by being there for large parts of the day, we would argue that the WPA has the possibility to sense and affect several of the information flows shown in Figure 2.3 with more precision than more traditional interactive systems (e.g. PCs, smartphones). This leads to the intriguing idea of future WPAs able to facilitate the transition from "felt sense" tacit knowledge generated as the human agent experiences the world to knowledge which the agent can consciously reflect on and articulate [59], augmenting human cognition at the core. Some more concrete potential mechanisms for making use of implicit input and output in the context of WPAs are listed as follows (all currently explored by the pervasive computing community but not necessarily in the context of designing WPAs).

- *Situation identification:* By implicitly monitoring the body state through pathway 10-1 (e.g. body posture, galvanic skin response, heart beat), and correlating it with the state of the nearby world (arrow 1), the WPA has a reasonable platform for determining the current situation in which the human agent is.

- *Subliminal cueing:* Certain phenomena measured best very close to the body (e.g. eye movements, facial expressions, EMG) can provide important insights into ongoing conscious and unconscious cognitive processing and therefore also for determining the intensity level and type of stimuli that could be used for subliminal cueing (e.g. [8]), and for the WPA to subliminally direct the human agent's gaze in a certain direction (pathway 3-4-9).

- *Mediated reality:* If the WPA gets sufficiently integrated into the visual perception flow, beyond Google Glass see-through and partially covering monocular HMDs and more towards the EyeTap vision of Steve Mann [120], the WPA could act both as a "filter" and "highlighter" which directly altered the perception of the surrounding world (arrow 2) so as to facilitate tasks.

2.8.2 WPAs & three types of assistance

Theoretically, WPAs can assist users in all three steps of the perception-cognition-action loop.

- *Perception Assistance:* Perception assistance includes both augmenting the world by adding needed information to the perception space or simplifying the world by filtering out potentially distracting/irrelevant phenomena from the perception space.

- *Cognitive assistance:* Many ideas for perception assistance ultimately target cognitive assistance. Hence, there is some overlap in the assistance mechanisms targeting perception and cognition. What we list here are support mechanisms that are targeted to affect cognitive processes more directly.

- *Action assistance:* The main focus of the action assistance is to make the world easier to manipulate by providing information relevant to the task in hand through appropriate modalities.

To investigate the utility of the Egocentric approach towards designing WPAs, I conducted an empirical study in a hospital where I designed a WPA for clinicians to support them throughout a workday. Chapter 3 and Paper 1 in the second part of the thesis, explain the design process and practicality of the Egocentric approach to the design of the WPA for clinicians.

§ 2.9 WPA interaction challenges

Explicit interaction with WPAs requires devoting parts of the user's physical, perceptual, and conscious cognitive resources to the interaction. The WPA is supposed to support the user on the move and in parallel with real-world tasks. Since humans cannot consciously attend to more than one task at a time, there will be a continuous competition between the WPA and the real world to grab users' attention. The limitation of humans physical, perceptual, and cognitive resources for interaction seems to be the main challenge of using WPAs in everyday life. These limitations force us to stop performing one task to attend to another one. But these interruptions can negatively affect the performance of users [2].

Previous studies have taken different approaches to mitigate the problems of multitasking and interruption. One strategy to reduce the negative effect of interaction with WPAs is to minimize the interaction duration: "Microinteractions are interactions with a device that take less than four seconds to initiate and complete."[9]. Another approach is minimizing the need for providing explicit input to the WPA by sensing context (e.g. human activities) and feeding implicit input to the WPA [173, 172]. To minimize the unwanted interruptions, the output of WPAs can also be displayed implicitly via subliminal cueing [8] or peripheral interaction [15, 67]. Another common strategy to avoid unwanted interruptions is to sense and model the context and predict the interruptibility of the user in any given context [18]. However, designing intelligent systems for managing interruptibility is not an easy task due to the complexities in types and objectives of interruptions, the diversity of contextual elements that need to be analyzed [202], the concept drift problem [185], and the challenges of detecting importance of massage based on e.g. the communication history between the two parties [61]. In Chapter 7 and Paper 10 of this dissertation, I introduce our self-mitigated interruption concept based on a symbiotic interplay between the existing interruption management infrastructure in our brains and the WPA.

To reduce the physical challenges of interaction with wearable computers, previous work has explored touch-less (e.g. [39]), hands-free, and eyes-free [159] interaction techniques. Those studies combined various input modalities such as body gestures, eye gaze, and speech commands to support users in different situations. In this dissertation, I focus on body gestures (hand, foot, and head) (Chapter 5 and Paper 4) and eye movements (Chapter 6 and Papers 6, 5, 7, 8, 9) as touch-less input modalities for WPAs. My empirical study in a hospital showed that there is a serious need for touch-less interaction due to the sterility restrictions in hygiene environments. The results of the empirical study in the hospital is represented in Chapter 3 and Paper 1, 2.

- Chapter 3 -

A Wearable Personal Assistant for a Hospital Setting

In this chapter, the result of my empirical study in a Danish hospital is described. First of all, the result of ethnography in the hospital is explained (section 3.2). Next my proposed framework (section 3.3) for designing a clinical WPA based on the WPA concept introduced in Chapter 2 is explained. Then the implementation of the WPA on Google Glass platform is discussed (section 3.5), and finally the evaluation of the WPA through a clinical simulation is described (section 3.6). The content of this chapter is related to Paper 1 (design and implementation) and Paper 2 (evaluation). The first part (before section 3.6) is about design and implementation of the WPA which includes more details compared to Paper 1. Section 3.6 briefly describes the method and results of the WPA evaluation which is explained in more detail in Paper 2.

§ 3.1 Why a WPA for the healthcare work domain?

Mobility is one of the main characteristics of work in hospitals [17]. Due to the spatial distribution of departments, wards, meeting rooms, and offices in clinical settings, clinicians, patients, and other resources need to move between different departments all the time. This kind of mobility, sometimes referred to as local mobility [17], gives rise to several work challenges. Aside from the considerable time that clinicians spend in moving between different places and waiting at elevators in hospitals, having access to the needed information in multiple situations is a challenge. The majority of previous work on providing remote access to the information has focused on providing mobile access to electronic patient records through wireless networks and mobile devices (e.g. PDAs, smartphones, laptops). However, most mobile devices do not support interaction on the move, which means users often need to stop, pick up their device, and direct their attention away from the task in hand and surrounding real world to interact with mobile devices [125]. This lag between "intention and action" indicates how effective a system can be used in mobile scenarios[191]. Moreover, for most mobile devices a touchscreen is the main user interface for interaction, which is not always the best channel for interaction in hospitals due to the sterility restrictions in some places such as operation theatres. Due to the above-mentioned challenges of using mobile devices in hospitals and potentials of wearable computers to support interaction on the move and touch-less interaction, wearable computers constitute a viable alternative to existing mobile platforms in healthcare work domain. However previous attempts [210, 3, 109, 48, 201] to bring wearable devices as clinicians' assistants to the hospital settings never took off because of technical limitations, complexities of healthcare work domain, and human related issues of interaction with wearable systems.

Emerging new generation of relatively unobtrusive eyewear computers such as Google Glass that support various hands-free modalities (e.g. head motion and voice commands), raises hopes for addressing some of the technical challenges in using wearables in hospitals. However, the design challenge of bringing wearables to the hospital context is still there. The first question (RQ1-page 3) is that given the unobtrusive form factor of state of the art smart glasses such as Google Glass and possibility of hands-free interaction, are the modern smart glasses a good enough platform for wearable personal assistants? The second question (RQ2-page 4) is that how can we design a wearable assistant for clinicians to support them in different situations in a hospital?

To answer the above questions, I developed a conceptual design framework (section 3.3), which helps designers of wearable systems focus on both human related issues of interaction with wearable systems and characteristics of the healthcare work domain. Moreover, to investigate the practicality of the proposed conceptual framework, I used the framework to design and implement a wearable personal assistant prototype for orthopedic surgeons on the Google Glass platfrom through an empirical study at Rigshospital in Copenhagen (section 3.2). To evaluate the hardware platform and the design of the WPA, I conducted a user study at ITX simulation facilities of Herlev Hospital in Copenhagen where I asked real orthopedic surgeons to use the WPA in three different scenarios: 1) access electronic patient records (EPR), 2) tele-guidance, 3) touch-less interaction in operation theatre.

To study the characteristics of work in hospitals I conducted an ethnographic study in a Danish hospital (section 3.2). Since surgeons are among the most mobile clinicians in the hospital, in my study, I focused on orthopedic surgery department. The ethnographic study provided valuable first-hand insight on how work is performed in hospitals. Furthermore, in order to address the human related concerns of designing wearable systems, I took a human-body-and-mind-centric (Egocentric) approach to develop the framework (section 3.3). The proposed conceptual framework (Figure 3.2) is shaped based on findings from the observational study at hospital and our human body-and-mind-centric approach.

§ 3.2 Ethnography in a hospital

In order to understand the healthcare work domain, I conducted an observational study at Rigshospitalet in Copenhagen. Our initial observations showed that surgeons' mobility is among the highest mobilities in the hospital; therefore, I studied the department of traumatology and orthopedic surgery of the hospital, and the specific focus of the ethnography was observing orthopedic surgeons throughout a workday. As part of the ethnographic study I shadowed an on-call orthopedic surgeon during a workday. Moreover, several orthopedic surgeons were observed in different types of orthopedic surgeries such as scoliosis surgery and fluoroscopic suregry. In order to record the observations for the further analysis, I took notes, pictures, and short videos of surgeons during the surgeries, ward rounds, and other occasions. During our study, one of the surgeons was the main contact point and coordinator of the project. Based on our observations, the main situations where the surgeons perform medical tasks can be summarised as follows:

1. Regular meetings
2. radiology conferences
3. surgeries
4. acute and trauma cases
5. ward rounds
6. stops in corridors and offices

A brief description of our observations in each situation is presented in the following sections.

3.2.1 Regular meeting

To observe an orthopedic surgeon in all situations, I shadowed an on-call orthopedic surgeon throughout a workday. I started the workday at 6:45 in the morning together with the surgeon. First, I dressed up like other surgeons due to the sterility requirements. The surgeon's workday started with a meeting with other surgeons. During the meeting, they discussed some general topics such as important administrative issues and special patient cases. For instance, during the meeting the shortage of vacant beds for new patients was discussed. A computer connected to a projector was used to present slides, medical pictures and other documents.

3.2.2 Radiology conference

After the meeting, all surgeon met at a radiology conference to review the latest X-rays, MRIs, and CT-scans of the patients. The radiology conference room was located in the 4th floor of the hospital while the orthopedic surgery and meeting room are located in the 15th floor. So that they needed to take the elevator to arrive at radiology conference room where the medical images were presented on the large screens. During the conference, surgeons discussed the important cases and took notes in their special notebooks. In the notebook, there was a visual tag for each patient. The tag helps surgeons retrieve the electronic patient records accurately in the EPR system.

3.2.3 Surgery

After the radiology conference, some of the surgeons who have been booked for different types of orthopedic surgeries went to the surgery rooms. Usually, the surgeries are planned in advance, and surgeons are aware of the schedule of the surgeries. Operation room is selected and prepared based on the type of the surgery since each type of surgery needs particular medical infrastructures. For example, in some of the complex orthopedic surgeries, the navigation system is needed to monitor the 3D model of the surgical site and the position of the operation instruments. The surgeon, I was shadowing during the workday, showed us three different operation theatres briefly. He did not have any booked surgery during the day since he was an on-call surgeon. Usually the patients are prepared for surgery before surgeons enter the surgery room. The surgical team need to sterilize their hands before entering the operation theatre. In the operation theatre, there are several screens to monitor patient records, X-rays, MRIs, and CT-scans (see left Figure.3.1)

Normally, the surgical team consists of at least an anesthesiologist, a surgeon, a surgery assistant, and a nurse who operates the computing devices. Since computers and their peripherals are difficult to sterilize, usually during a surgery, an assistant or nurse operates the mouse and keyboard for surgeons. In more complex surgeries, the surgical team can include more clinicians. I observed a complex orthopedic surgery (Scoliosis surgery) in which the surgical team used the navigation system [197] to increase accuracy of the operation. In surgeries with the navigation system, a CT-scan of the patient is displayed on a large screen, and at the same time the

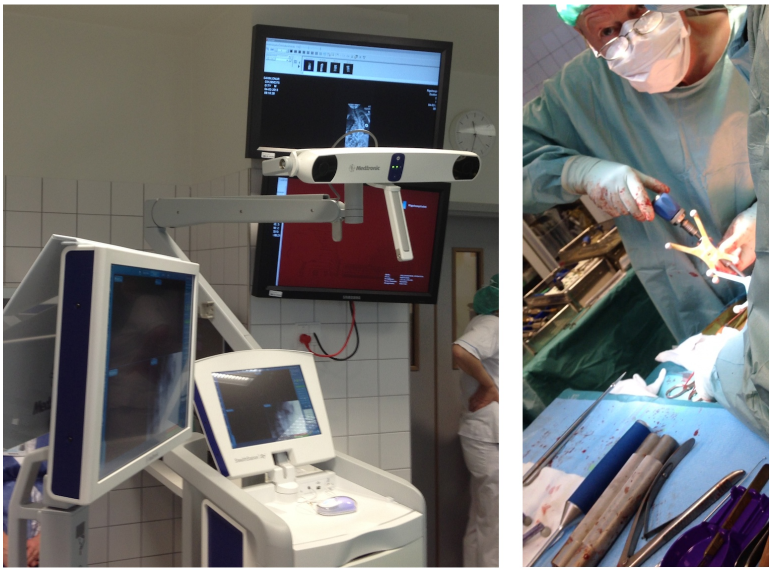


Figure 3.1: *The left picture illustrates the navigation system and several large screens to monitor medical images in the operation room. The right picture shows how surgeons look at the screens and at the same time operate on the patient [81].*

navigation system tracks the positions of surgical tools in relation to the patient's coordinate systems. Which means the surgeon needs to look at the screen and at the same time use the surgical tools to operate (see right Figure.3.1). In such situations, the surgeon needs to frequently switch the visual focus between the surgical site and the screen. The same challenge can be observed in other types of the surgeries where a live view of the patient's X-rays (fluoroscopy) is needed. During some complex surgeries, the surgeon might call an experienced colleague for help. In such cases, the experienced surgeon provides guidance either over the phone or personally in the operation room.

3.2.4 Acute and trauma case

When I was shadowing an on-call orthopedic surgeon, the acute department called the surgeon, and the surgeon had to go to the acute department immediately. The surgeon used a special card to take the elevator as soon as possible. The acute department sent a short text message to the surgeon on his mobile phone describing a brief history of the patient, and the surgeon read the message before arriving at the acute department. In the acute department, there was a woman injured in a bicycle accident and the emergency team were working tightly together to investigate if there is risk of trauma (an injury that has the potential to cause prolonged disability or death) for the patient or not. They used different sources of information to treat the patients: they asked the details of the accident from the ambulance personnel, they reviewed the health records of the patient, measured vital signs of the patient, took X-rays of the head, and etc. Fortunately, after couple of minutes they found no sign of traumatic injuries. The emergency team slowed down procedure gradually and sent the patient to the ward for further investigations.

3.2.5 Ward round

Visiting patients in the ward is one of the daily routines of the surgeons. They do their ward rounds in collaboration with a nurse by moving from patient to patient in their bed wards. The surgeon makes a diagnosis and prescribes treatment, and the nurse has an overview of the patients and updated knowledge about their condition. During the day I shadowed the surgeon, he visited two patients in the ward. Before ward round, the surgeon reviewed the patient records and recent medical images of the patient on a computer in the nurse office. After visiting the patient, the surgeon recorded his voice by a Dictaphone to report the ward round results. The recorded voice was transferred to the patient record system through a desktop computer. To save the doctors' time, the administrative personnel transcribe the recorded files later.

3.2.6 Stops in corridors & offices

Due to the high mobility of clinicians in the hospital, they usually bump into each other spontaneously in corridors, wards, and other locations in the hospital. In these ad hoc collaborative situations, sometimes they talk about a particular patient or medical task. When I was following the on-call surgeon, one of his colleagues saw him in the corridor and asked him to have a look at one of the patients X-rays. They moved to a stationary computer to have a look at the medical information of the patient. Then they moved together to the ward to visit the patient and discuss the appropriate treatments.

§ 3.3 A design framework for wearable personal assistants in a hospital setting

To ensure that the special requirements of the healthcare work domain is considered in design of the WPA, I developed a conceptual framework (Figure 3.2) in the form of a two-dimension matrix. The rows of the matrix describe the main characteristics of working in the clinical environments

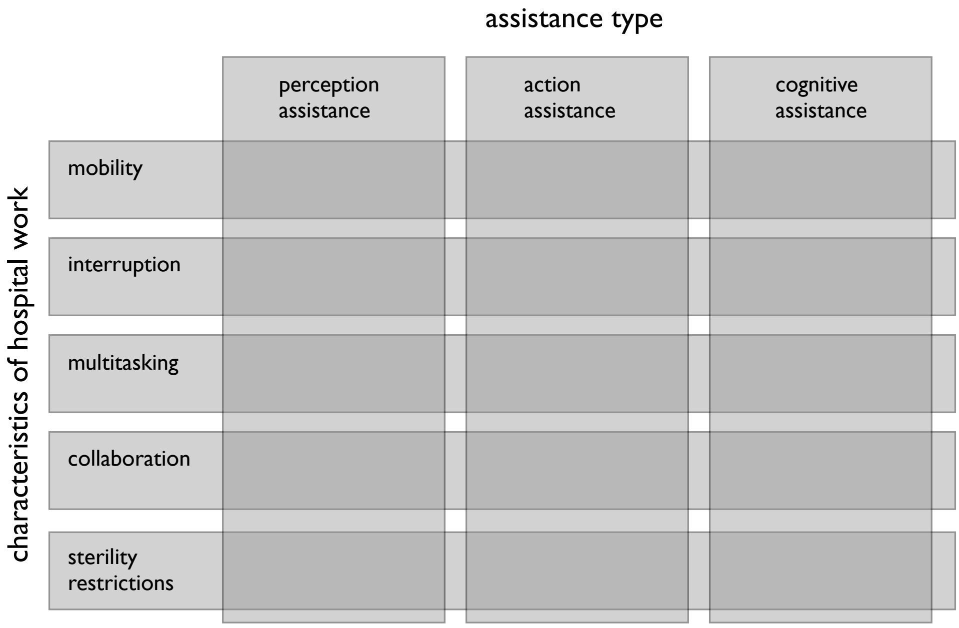


Figure 3.2: *A design framework for Wearable Personal Assistants in hospital settings [81].*

while the columns explain the main types of assistance that WPA can provide to support the hospital work.

3.3.1 The characteristics of work in hospitals

Based on my observations in the hospital and according to the previous studies on hospital work, I categorized the specifications of the hospital work domain into 5 main characteristics as follows.

3.3.1.1 Mobility

Mobility is the main characteristics of the work in hospitals [17]. My observations also showed that clinicians move between different departments and floors all the time. The main reason for mobility in hospitals is the need for being in different physical places, getting in contact with a particular person, and need to access knowledge and different shared resources [17]. The challenge arises from mobility in a hospital setting is locating and moving people, resources, and knowledge between different places. The digital resources and explicit knowledge can be shared among clinicians through computing devices (e.g. PDAs) connected to the electronic patient record system [182, 55]. But sharing physical resources, people, and tacit knowledge is still a challenge.

3.3.1.2 Interruption

Due to the high mobility of clinicians in the hospitals, they use pagers, mobile phones, and other devices to find each other; however, these devices are also inherently interruptive and interfere with a smooth flow of work [17]. For example, in my filed study at the hospital, during the trauma case, I observed that someone called the surgeon's phone and he stopped the treatment for a short time to answer the phone call. Problems caused by interruption include the slowing of work progress, a reduction in task performance [12, 41], and negative impact on mental state [13].

3.3.1.3 Multitasking

As I observed in different orthopedic surgeries, clinicians need to switch their attention frequently between computer screens displaying patients medical images and the surgical site (see right Figure.3.1). This kind of multitasking divides the limited cognitive and perceptual resources of the clinicians between different tasks, and could negatively affect the performance of clinicians in each task [88].

3.3.1.4 Collaboration

My observations in the hospital and previous studies [158] indicated that healthcare is a highly collaborative work domain. Almost every single medical task in hospital is accomplished in collaboration. The ward rounds require a collaboration among physicians, nurses, and patients. Moreover, surgeries and handling emergency cases require a tight collaboration between clinicians. In such synchronous collaborative works, each team member needs to be aware of the other team members' actions. This awareness is not always easy to achieve due to the limitations of human perception and cognition.

3.3.1.5 Sterility Restrictions

In the sterile environments, like operation theatre, clinicians should not touch unsterile objects. Since computing devices are hard to sterile, usually it is not possible for clinicians to touch computing devices when their hands need to be sterile. In the surgeries I observed, a dedicated nurse was responsible for controlling computers for the surgeon. In fact, the previous studies on WPA for clinicians [3, 35] also explored touch-less input modalities. In order to comply with sterility restrictions, all of the possible interactions with WPA should be doable through touch-less techniques which is not always easy to achieve.

3.3.2 Three types of assistance

In order to deploy the proposed human body-and-mind-centric approach introduced in the previous chapter, we defined three types of assistance for WPA to support above-mentioned characteristics of healthcare work domain.

3.3.2.1 Perception Assistance

Perception assistance includes both augmenting the world by adding needed information to the perception space or simplifying the world by filtering out potentially distracting/irrelevant phenomena from the perception space. One of the main perceptual assistance that can be provided by a WPA is extending human perception beyond limitations of human perception. For example, the WPA can project an invisible object in the eyes of the wearer.

3.3.2.2 Cognitive Assistance

Many ideas for perception assistance ultimately target cognitive assistance. Hence, there is some overlap in the assistance mechanisms targeting perception and cognition. What we targeted here is to affect cognitive processes more directly through memorizing, learning, etc.

3.3.2.3 Action Assistance

The main focus of the action assistance is to make the world easier to manipulate by providing information relevant to the task in hand through appropriate modalities. Action assistance mainly targets to help human agents overcome physical constraints caused by performing real world tasks.

§ 3.4 Ideas for WPA functions to support orthopedic surgeons

The introduced conceptual framework helps us understand what types of assistance is needed to support each characteristics of hospital work. For example, what kinds of perceptual, cognitive, or action related assistance that can support mobility in the hospital by addressing mobility challenges.

Based on our conceptual framework and observational study, I proposed the initial ideas for functions of WPA listed in the cells of the matrix (Figure 3.3). The feasibility and utility of the initial ideas have been discussed with three orthopedic surgeons in semi-structured interviews, and based on the interview results I focused on three main ideas: 1) mobile access to the electronic patient records, 2) telepresence, and 3) touch-less interaction. The selected ideas are highlighted in the Figure 3.3. In the following sections, I briefly explain the initial 13 ideas for WPA to support above-mentioned characteristics of the hospital work.

3.4.1 Mobility Support

3.4.1.1 Briefing on the move

Previous studies have shown that clinicians spend considerable part of their time on the move[17]. To utilize this time the WPA could provide them relevant information to the next situation where they are moving. For example, in some emergency cases in hospital, they need to call surgeons, and sometimes it takes a lot of time until the surgeon arrives at the acute department. In such cases, the WPA can provide a brief state of the patient to the surgeon on the move through appropriate modalities (e.g. visual when the human agent is standing in the elevator or aural when the human agent is walking)

3.4.1.2 Mobile access to the patient records

The main reason for mobility in hospitals is the need for being in different physical places, getting in contact with a particular person, or need to access knowledge and different shared resources. A WPA connected wirelessly to the electronic patient record system could facilitate information and knowledge sharing, potentially further simplified by automatic retrieval based on location (history) of the clinicians (e.g. data for the currently nearest patient), or bookmarked X-ray images for a surgeon attending a radiology conference.[17]

3.4.1.3 Telepresence

While digital resources can be easily shared through mobile devices it is still a challenge to share physical resources, people, and tacit knowledge. WPAs could help sharing tacit knowledge among clinicians by offering ad-hoc telepresence sessions between a remote specialist and local clinicians.

3.4.1.4 Data entry on the move

In a clinical setting, everything needs to be properly recorded, for legal reasons and for ensuring continuity of care. For instance, after visiting a patient in ward a round, reviews and decisions should be recorded by the clinicians. The WPA could support data entry on the move by automatically recording visited patient id, date, time, and identifiable author. Furthermore, multimedia content such as vocal reports, short videos, and pictures of patients could be added to the patient records through WPAs.

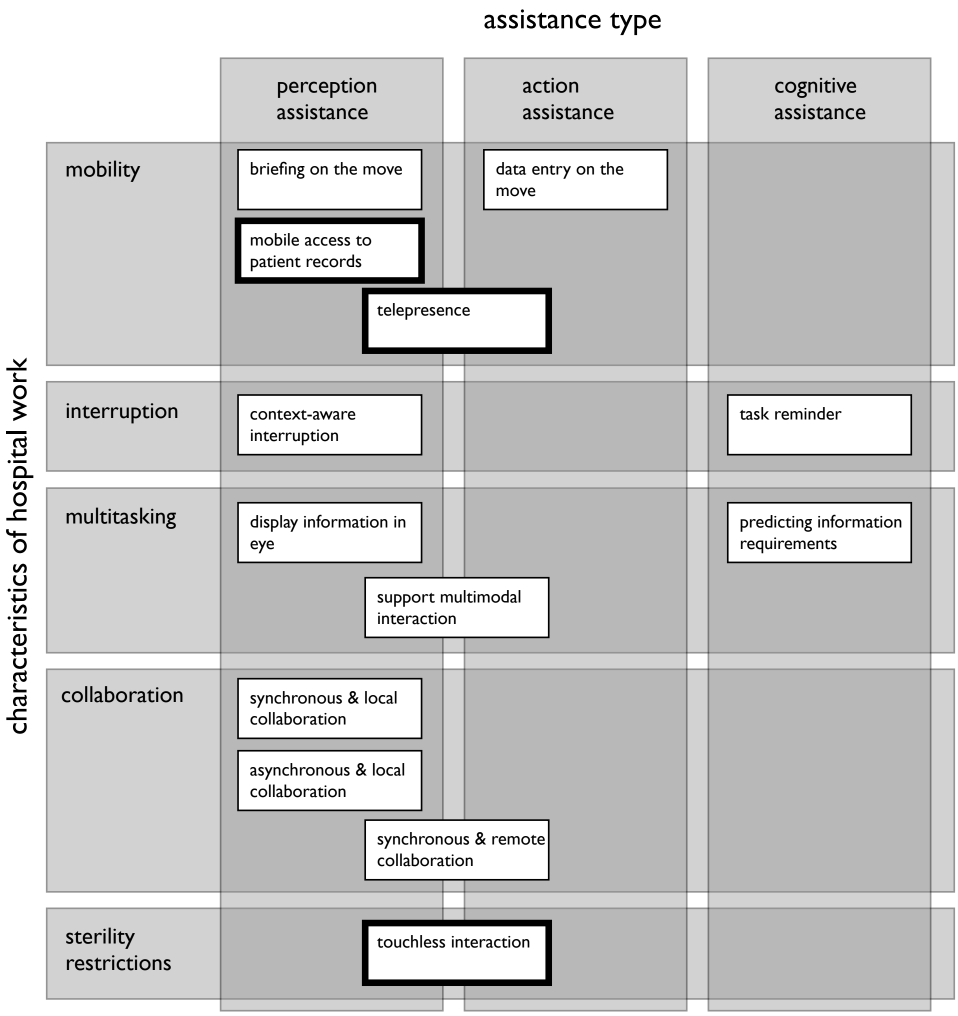


Figure 3.3: *Using the design framework to define initial ideas for WPA functions [81].*

3.4.2 Interruption Support

3.4.2.1 Context-aware interruption

In a hospital setting, clinicians move all the time, and sometimes they need to find each other since the healthcare work domain is highly collaborative. They usually use smartphones, pagers, or other mobile technologies to find each other. Since these devices are inherently interruptive, there is a contradiction between being available through mobile devices and performing medical activities smoothly[17]. If clinicians carry an always-on WPA all the time, WPAs could determine current context and determine whether its wearer is interruptable in a given situation or not, following one of at least two approaches: 1) the context of the target clinician could be shared among other clinicians, allowing callers to make the decision whether to interrupt or not[18], 2) the WPA could intelligently choose interruption strategy by e.g. adjusting the time, modality, and intensity of the notification. Having detailed information about the perceptual focus of the target clinician would allow the WPA to offload the decision of exactly when to interrupt, to the unconscious processes (Figure 2.3, page 28) of the user.

3.4.2.2 Task reminder

Forgetting the interrupted task is one of the implications of interruption. To mitigate this risk, the WPA could remind about the interrupted tasks after interruption.

3.4.3 Multitasking Support

3.4.3.1 Display information in eye

During surgery, the surgical team currently needs to monitor important information on a display. For example, during some of the orthopedic surgeries, it is necessary to take periodically visualize X-rays of the patient (flouroscopy). Fluoroscopic surgeries forces the surgeon to frequently switch focus between the surgical site and the screen. The WPA could display this information directly to the surgeon's eyes, allowing the surgeon to maintain focus on the patient, reducing surgery time and complications from X-radiation exposure.

3.4.3.2 Predicting information requirements

The WPA would be able to recognize clinicians' activities and allow them to access relevant information quickly, without sacrificing their connection to the patient or procedure at hand.

3.4.3.3 Support multimodal interaction

WPAs could facilitate parallel activities by providing appropriate input and output modalities according to action and perception restrictions of the situation in which clinicians perform medical tasks. For instance, when the human agent is performing a visual task, the WPA could switch automatically or manually to aural modality for displaying helpful information.

3.4.4 Collaboration Support

3.4.4.1 Synchronous & local collaboration

Almost every single medical task in hospital is accomplished in collaboration. The ward rounds require a coordinated collaboration of physicians, nurses, and patients. Moreover, doing a surgery or handling an emergency case require collaboration of a medical team. In synchronous and local collaborative scenarios, WPAs could increase awareness among team members. For example, during a surgery the attention point of a surgeon on the surgical site, tracked by a gaze tracker, could be shared among the surgical team. Another example scenario would be to control sterility in operation theater. In the surgery room, non-sterile objects (e.g. hands) should never touch sterile objects. Assuming that all surgery team members have their own WPA, the WPAs could together maintain a common picture of the sterility state of objects based on who touches what and issue warnings if sterility requirements are about to be violated.

Furthermore, Ad-hoc collaboration sessions occur frequently due to the high mobility of clinicians. WPAs could automatically but gracefully inform the participants about tasks and patients in common in these spontaneous meetings. Also, the WPAs could form a virtual shared space to enrich the collaboration with multimedia content.

3.4.4.2 Synchronous & remote collaboration

As mentioned earlier, specialists are sometimes called with short notice to specific locations for emergency judgements. WPAs could reduce the need for physical colocation by offering quick-to-setup hands-free video communication where the remote specialist could see what the local clinician sees and provide valuable medical feedback. Such a telepresence application could support also emergency teams in the field. [201].

3.4.4.3 Asynchronous & local collaboration

Apart from synchronous collaboration, many medical tasks are coordinated through asynchronous collaboration. For example, clinicians update the time schedule of the personnel on whiteboards, descriptions of performed medical tasks on patients during a working shift are entered into the computer systems, etc. A problem is that most of the mechanisms for handling these information resources are passive, demanding clinicians to specifically act in order to share. Ubiquitous WPAs would allow for binding virtual objects to physical objects, locations, or situations. For example, a night shift nurse could leave a voice message close to a patient which would be played by the take-over colleague's WPA when close to that patient.

3.4.5 Sterility Support

3.4.5.1 touch-less interaction

In most operation rooms, several computers and large displays monitor different medical information before and during surgery(see left Figure 3.1). Due to sterility restrictions, input devices are handled by an assistant or a nurse instructed by the surgeon, sometimes causing misunderstandings and delays. A WPA could act as an interface between stationary computers and the surgeon through touch-less modalities such as speech, body gestures, gaze, etc.

§ 3.5 WPA Prototype on Google Glass for three selected scenarios

I discussed the utility of the initial ideas with three orthopedic surgeons in semi-structured interviews, and based on the interview results we focused on three main ideas: 1) mobile access to the electronic patient records, 2) telepresence, and 3) touch-less interaction. In the following sections, we briefly explain the initial 13 ideas for WPA to support above-mentioned characteristics of hospital work. Due to the unobtrusiveness of the Google Glass and also providing several input channels such as voice commands, head motion, and touchpad, we decided to develop the WPA prototype on the Google Glass platform (Figure 3.4). The first prototype of the WPA supports three selected scenarios.

3.5.1 Mobile Access to the Patient Records

According to the interviewed surgeons, mobile access to the patient records through a WPA could be a valuable support in several situations such as ward rounds, operation theatre, and ad hoc collaborations. However, each situation requires the WPA to provide different modalities for interaction. For example, the WPA should support touch-less interaction in the operation theatre due to the sterility restrictions while in the ward rounds surgeons can use Google Glass touchpad to provide input to the device. In fact supporting multimodal interaction is a crucial requirement for WPA in order to succeed

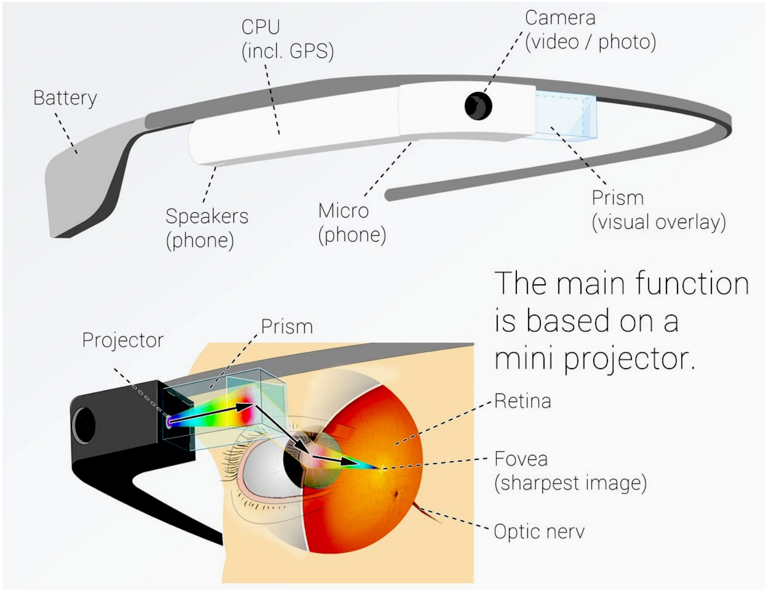


Figure 3.4: *Google Glass architecture [130].*

in providing mobile access to the patient records. The main steps of mobile access to the patient records on the Google Glass application are as follows (Figure 3.5).

3.5.1.1 Finding the patient records

To find and retrieve the health records of a patient the system provides two main channels: voice and QRCode. Users are able to filter the patients list by saying either name or person number of the patient. They can also find the patient records by reading the visual tags (QRCode of the patient) through front view camera. The latter method is faster and more accurate where the QRCode is available.

3.5.1.2 Switching between textual data and medical images

Patient records are distributed in several pages (cards) on the Google Glass. Due to the limitation of display size in Google Glass the textual part of the patient records are shortened. However, medical images (X-rays, CT-Scans, etc.) are the most important part for the orthopedic surgeons. Users can switch between different cards through voice commands or performing swipe gestures on the touchpad of Google Glass.

Figure 3.5: *Screen-shots of the main cards of three applications [81].*

3.5.1.3 Interaction with medical images

Users can zoom in/out, rotate, and navigate through a zoomed in view of the medical images by providing voice commands, touch gestures on the touchpad, and head movements. For example, using the head tracker of the Google Glass, surgeons are able to navigate through an enlarged image in real-time by head movements (see Figure 3.5).

3.5.2 touch-less Interaction

We designed the touch-less interaction module of WPA which enables surgeons to provide touch-less input to other computers in the operation theatre through voice commands and head gestures. The surgeon can choose to interact with two different systems: 1- a 3D medical imaging system (InVesalius [1]), 2- a 2D image viewer to review the X-rays and other 2D images. The surgeon can switch between 2D images, zoom in/out, and navigate through an enlarged image on the stationery screen via voice commands and head gestures through WPA. To interact with 3D imaging system (InVesalius), user should first choose the desired view between Axial, Sagittal, and Coronal. The depth view of the 3D model can be adjusted by either

[1] InVesalius is a free open-source software to construct 3D medical images. http://svn.softwarepublico.gov.br/trac/invesalius/

voice commands or head movements. Aside from supporting touch-less interaction, this module enables users to take a snapshot from the stationery screens and display it in the surgeons' eye (see Figure 3.5).

3.5.3 Telepresence

Based on our interview with three surgeons, during complex surgeries, the surgeon might need help from an expert colleague. In such situations, the surgeon asks the expert colleague to personally help or provide guidance through phone call. To enhance the effectiveness of such collaborations, the telepresence module of the WPA is designed to share a still image taken by the local surgeon with a remote expert. The remote expert can use a mobile application to see the shared image and provide guidance through vocal communication and also by adding sketches on the still image. The graphical content provided by remote expert is superimposed on the head mounted display of the local surgeon in real-time.

§ 3.6 Clinical simulation study at ITX facilities

3.6.1 Method

Since deploying the WPA in a real clinical setting is technically and legally unfeasible, we evaluated the WPA in a clinical simulation in a separate simulation facility. In the medical work domain, conducting clinical simulations is a popular method to train clinicians in critical clinical scenarios, such as surgeries and emergency cases, and the usefulness of such studies is proved for training clinicians [4]. The clinical simulation approach has also been used as an effective method for evaluating clinical systems with representative users performing representative tasks [85]. Our study is conducted at a clinical simulation and training facility in a Danish hospital. This simulation facility includes different hospital departments from patient wards to surgery rooms. In our study, we set up the facility for the above-mentioned three scenarios. In the touch-less interaction and tele-presence scenarios we set up the surgery room, and for the mobile access to the patient records scenario we set up a patient room with two beds in the ward. [82]

During one-day simulation, two orthopedic surgeons, a senior nurse, and two human actors (to play the role of patients) participated in the study. Both surgeons performed all three scenarios. After completing each scenario, the surgeons were asked to complete a questionnaire polling their experiences completing the task and using the system. The participants were also interviewed to get deeper insights into their experience of using the WPA. [82]

Figure 3.6: *A surgeon uses the WPA for touch-less interaction with X-rays (left screen) and MRI images (right screen) [82].*

We took a scenario-based approach in evaluation of the WPA. Scenarios included:

- **Scenario 1: touch-less interaction with medical images:** In the surgery room, the surgical team including a surgeon and a nurse are about to start the surgery. Before starting the surgery, the surgeon needs to have a look at X-rays and MRIs. But his/her hands are sterile and s/he cannot touch the mouse or keyboard. Therefore, the surgeon uses the WPA for browsing X-rays and MRIs on two different screens in the operation theatre through voice commands and head movements. The surgeon might need to zoom in, rotate, or navigate through the medical images until s/he finds a good view and gets ready to start the surgery. The surgeon can also take a snapshot of the screens and see the content on the HMD of the Google Glass. [82]

- **Scenario 2: tele-presence during surgery:** After adjusting the medical images on the screen (in the previous scenario) during the surgery, the surgeon encounters a complex situation and needs help from an expert colleague. The surgeons uses the WPA to start a tele-presence session with the remote colleague. The local surgeon takes a picture of the surgical site and calls the remote surgeon using the Glass. The remote surgeon answers the call. Then the local surgeon explains the situation and shares the taken picture with the remote surgeon. The remote surgeon provides some guidance while marking the shared photo on his tablet. The local surgeon sees the content provided by the remote surgeon on the Glass in real-time. [82]

- **Scenario 3: mobile access to the EPR in ward rounds:** It is one

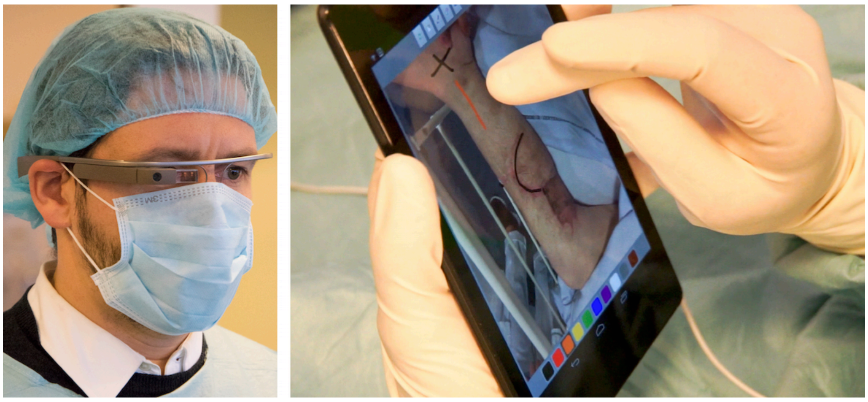


Figure 3.7: *A remote surgeon (right picture) uses a tablet computer to provide guidance to the local surgeon (left picture). The local surgeon sees the visual guidance on the HMD in real-time [82].*

day after surgeries. Patients are lying down in the hospital bed in the ward. The surgeons should visit two patients who got surgery. The surgeons use the WPA to review the new X-rays and the latest state of the patients while walking to the ward together with a nurse. The surgeon will search for the patient records on the Glass by saying the patients name. After finding the patients records, the surgeon reads EPR text explaining the latest state of the patient on the Glass and also looks at the X-rays and MRI pictures on the Glass. The surgeon can zoom in/out, rotate or navigate through the medical images. The nurse has the latest state of the patients (last blood test, etc.). The nurse answers the questions that the surgeon might ask during the ward round. The surgeon visits the patients and asks some questions about his/her pain, etc. Also the surgeon might need to answer the patients' questions during the ward round for which the EPR system might be used. After visiting the patients, the surgeon prescribes the next treatments and the nurse writes down the prescriptions. [82]

3.6.2 Results & Discussion

The result of our simulation study indicates that using the WPA for touch-less interaction with medical images can save surgeons time and energy for the surgery. Moreover, by using the WPA for touch-less interaction, there is no need for a dedicated nurse to control the mouse for surgeons. However, there are some limitations in both voice commands and head movements for touch-less interaction. Using voice commands is a relatively reliable input modality but due to the slow speed of the discrete voice commands, it is not appropriate for providing a lot of commands in a short period of time.

Figure 3.8: *A surgeon uses the WPA to browse electronic patinet records and X-rays in the ward round scenario [82].*

In contrast to the voice commands, the head movements can be useful for continuous interactions; however, due to the perceptual overlap between seeing the large screen (X-ray and MRI systems) and seeing the pointer on the HMD, it is not easy to use the head movements as a mouse to control the pointer on the HMD for interaction with other systems. The lowest scores in Figure 3.9-a are related to the accuracy of head tracking specially by P2 that reveals the challenge of using head movements for touch-less interaction with medical images. [82]

Apart from the low quality of the image on the HMD of Google Glass which is indicated in both usability questionnaire (Figure 3.9-b) and the complementary interview, the WPA was successfully used for tele-presence scenario. According to the both surgeons tele-presence was the best application for the WPA. However, in this scenario we observed the problem of overlapping between human to human conversation and voice commands to the system. This indicates the need for more touch-less input modalities (e.g. gaze) in Google Glass. [82]

The most challenging scenario was to use the WPA in ward rounds which revealed the social problems of using Google Glass in parallel with human to human interactions. Apart from the social problems, the Google Glass small display turned out to be a limitation for intensive text readings which is in line with the concept of microinteractions [9] where interacting with the device should not exceed 4 seconds. To achieve such fast interactions, the WPA needs to prepare the information for the surgeons in a way that the surgeon can get what s/he needs at a glance. Using context to predict the information requirements of the user, and using visualizing techniques to communicate maximum amount of information in the shortest time period are some of the approaches that can be taken to minimize the interaction duration. [82]

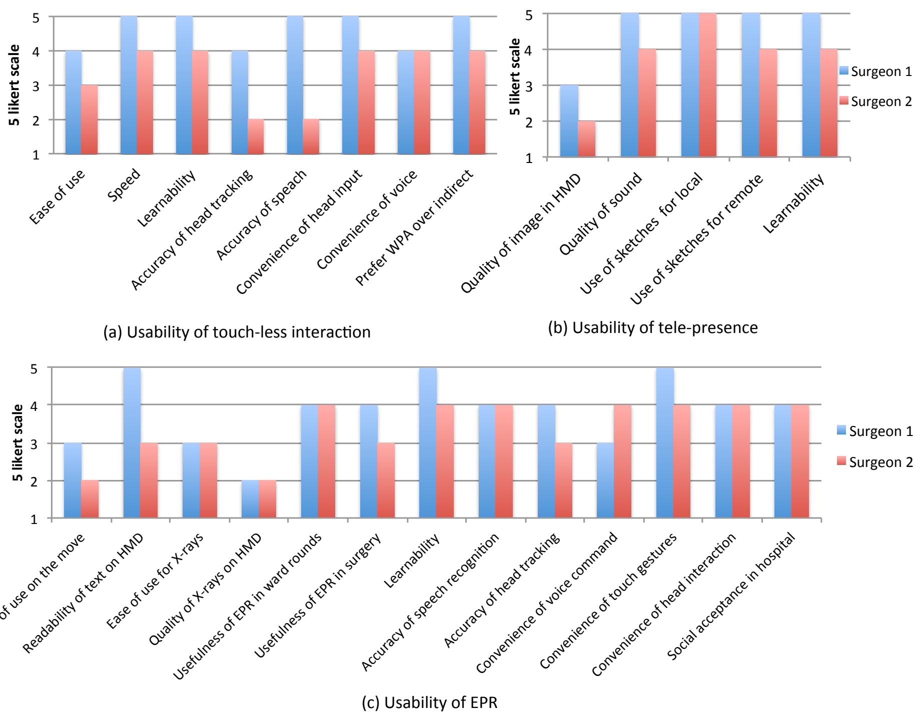


Figure 3.9: *a) Usability of the touch-less interaction module of the WPA, b) usability of the tele-presence module of the WPA, and c) usability of the EPR module of the WPA [82].*

The three studied scenarios are representatives of three different configurations for interaction with WPAs as follows. [82]

1. **WPA as an interface to other computers:** The touch-less interaction scenario defines the WPA as an interface between the user and other computers. In this type of scenarios, the human agent interacts with two different computers in parallel.

2. **WPA as an interface between two human agents:** In the tele-presence scenario, the WPA is defined as an interface between two human agents which means the user interacts with another human agent through the WPA and there is no parallel interaction.

3. **WPA in parallel with human to human interaction:** In the ward round scenario, the user interacts with another human agent and with the WPA in parallel. If we look at the results of the usability questionnaires and interviews, we can conclude that the WPA got the best scores in the tele-presence application where there was no parallel interaction, and the user interacts sequentially with the WPA and the other human agent. In the touch-less interaction scenario, the usability of the WPA is evaluated as average. In this scenario, the user interacts with two computers in parallel: the WPA and X-ray/MRI systems. The most challenging scenario is the ward round where the user needs to interact in parallel with the WPA and a human agent.

Our observations indicate that the bottleneck for what can be computed by combined human and computer systems stem from the limitations of human perception and cognition *not* in limitations of current computer hardware and software. As system designers, we need to more than ever make sure that our WPAs talk to our static biological setup in the way it was designed by evolution to interpret and act in the world.

- Chapter 4 -

A Wearable laser pointer as a visual display for WPAs

Vision is the dominant sense for humans. This means that visual perception plays a crucial role in experiencing the real world [155]. That is probably the main reason why vision-based user interfaces (e.g. GUI) have mostly been used as the dominant communication channel between humans and computers. Even the early versions of wearable computers included head-mounted [119] or wrist-mounted [22] visual displays. In the related work part of this chapter, I briefly review different form factors of wearable visual displays. Then I briefly introduce a wearable laser pointer as a novel visual display for wearable personal assistants. More details about the wearable laser pointer can be found in Paper 3.

§ 4.1 Related Work

Head-mounted displays (HMDs) are the most famous form factor for visual displays in wearable computers. However, other forms of visual displays such as wrist-mounted displays, wearable projectors, or recently developed contact lenses have also been used or studied as an output modality for wearable computers.

4.1.1 Head-mounted displays (HMDs)

HMDs are display devices that are mounted on a helmet or glasses with one (monocular) or two (binocular) small displays in front of the user's eye(s), and user can see the visual contents such as images, textual information, or videos in the display. Head-mounted displays can be monocular with monoscopic view or binocular with stereoscopic view. Some HMDs are see-through which means that the wearer can see also the real-world through

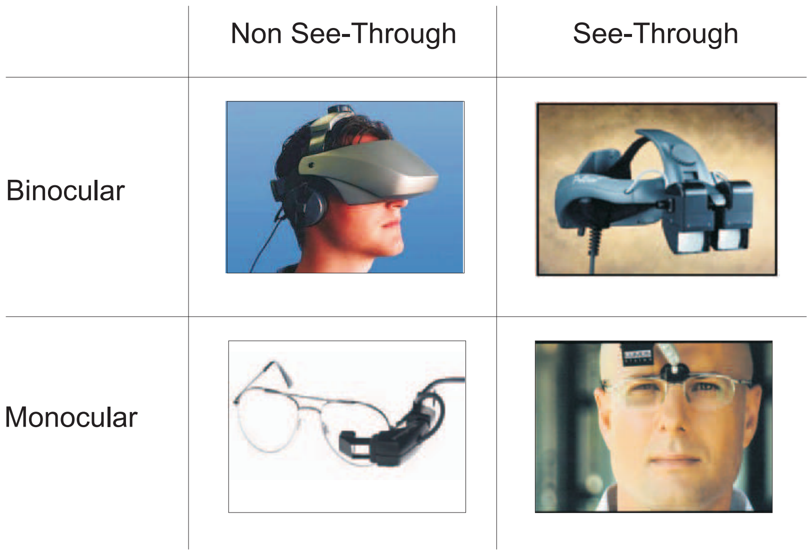


Figure 4.1: *Different types of head-mounted displays [211].*

the display, while some others just show the virtual world to the wearers and block the real-world view. See-through HMDs are useful in augmented reality and mixed reality applications in which the computer-generated image is superimposed on the real-world view. In the see-through HMDs, the computer-generated image can be projected on a semi-transparent mirror and user can see the real world directly (optical-see through HMDs). In the video see-through HMDs, a head-mounted front-facing video camera streams the real-world image, which is combined electronically with computer-generated image in front of the viewer's eyes. [196, 129, 74]

In Figure 4.1 different types of head-mounted displays are shown.

In general, monocular HMDs are easier to use for wearer compared to the binocular ones due to their smaller weight [11]. They also offer the free eye of users a complete real-world view; however, NCR's study on HMDs showed that sharing attention between real world and computer-generated image could be challenging for wearers of the monocular HMDs [23].

While the technological advances in developing unobtrusive and high performance HMDs have opened new horizons for applications of wearable computers in everyday life, fundamental challenges still remain. Eye fatigue, small field of view, swimming effects, limited resolution, and multiple focus planes, are some of the famous problems associated with HMDs [102, 162].

Another important issue with monocular HMDs is the negative effects of monocular HMDs on face-to-face communication between humans. Monocular HMDs reduce the quality of interaction between humans and affect eye contact with other humans when the HMD is displaying information. As

Figure 4.2: *Apple watch (right side) and Samsung Gear S2 (left side) are two famous brands in smartwatch market.*

it was discussed in the Chapter 3, I also observed the negative effect of using Google Glass in the ward round scenario where the surgeons needed to look at the information displayed on the Google Glass and at the same time communicate with the patients.

4.1.2 Wrist-mounted displays

Wristwatch is another form factor for wearable computers, which is easier to access compared to other forms of mobile devices such as PDAs and smartphones usually kept in the pocket. A wrist-mounted display can be easily viewed by flicking the wrist while PDAs or smartphones need to be picked up and opened before use. However, due to the small size of the device it is not easy to display large amount of text or interact with the touchscreen using touch modality. Smartwatches are becoming more popular these days for mostly self-monitoring applications and microinteraction. Figure 4.2 illustrates smartwatches from Apple [1] and Samsung [2] companies).

4.1.3 Wearable projectors

Through emerging pocket-size Pico projectors, the vision of augmenting the physical world with interactive projection came true [66]; however, the concept of wearable projection has been studied earlier using bulky projectors [92]. In the Wuw-wear ur world system by MIT Media Lab [131] a gestural

[1] http://www.apple.com/watch/

[2] http://www.samsung.com/us/mobile/wearable-tech/all-products?filter=smartwatches

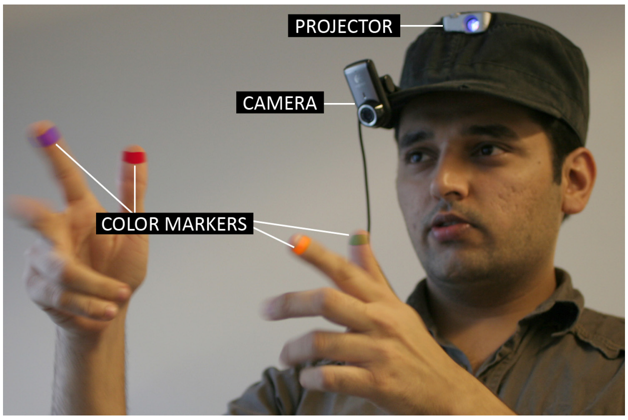


Figure 4.3: *Wuw-wear ur world: a wearable gestural interface [131].*

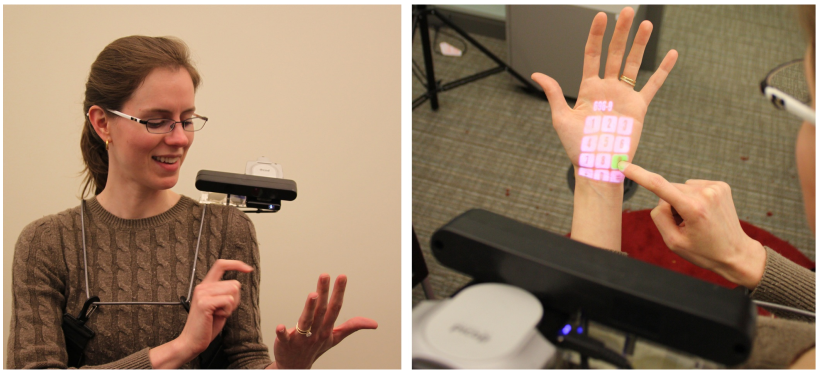


Figure 4.4: *The OmniTouch system enables graphical, interactive, multi-touch input on arbitrary, everyday surfaces.*

interface using a wearable projector and a camera was developed and evaluated (Figure 4.3). This system has a front-facing camera to detect the users' hands and project information onto the surrounding surfaces or physical objects. OmniTouch [66] system is another study on wearable projection by Microsoft Research Group. The OmniTouch system targeted to extend mobile interaction beyond the limitations of existing mobile devices by using ad hoc surfaces around users instead of display of mobile devices. The OmniTouch system has a depth sensor creating a 3D model of the surrounding environment that helps the system to detect user's hands for different kinds of input (Figure 4.4). One of the main advantages of wearable projectors is the possibility of sharing information in collaborative settings [128] which is not possible in head-mounted displays; however, there are some limitations for projection technology such as contrast challenges in bright environments.

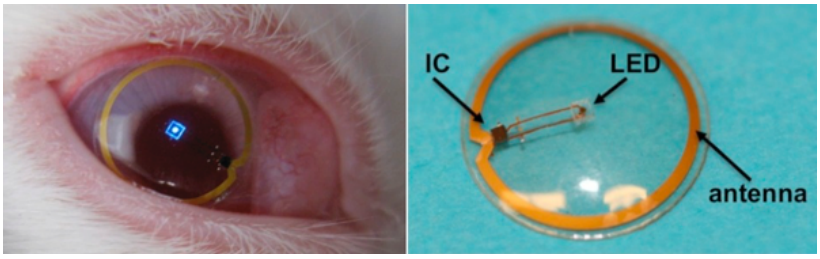


Figure 4.5: *The right picture illustrates the first prototype of the contact lens display, and the left picture shows the contact lens in a rabbit eye [108].*

In addition, the quality of the projected image depends on the texture and color of the background surface.

4.1.4 Contact lens displays

There are some studies on feasibility of developing and using wireless contact lens displays as a future display technology for augmented reality applications [108]. The first prototype of the contact lens display comprises a single pixel LED, a miniaturized IC, and an antenna for electricity induction (Figure 4.5). This single pixel prototype has been tested on a live rabbit successfully.

4.1.5 Wearable laser pointers

Wearable motor-controlled laser pointer is another technology for superimposing information (e.g. sketches, text, point, etc.) onto physical objects and surfaces around human agents. Stationary laser pointers have been used for augmented reality applications as an alternative technology to HMDs [175]. In [167], a combination of shoulder-mounted laser pointer and HMD has been evaluated to guide a local worker by a remote expert in a teleguidance scenario. The shoulder-mounted laser pointer was controlled by a remote guide to direct the attention of the wearer to particular objects at the local site. Since they used servos to control the laser beam, the only content could be provided to the local user through the laser pointer is just a point.

§ 4.2 The proposed wearable laser pointer

To answer the RQ5 (page 5)), I extended the previously implemented stationary laser pointer for augmented reality applications [175] to a wearable version [83] (see Figure 4.6). The utility of my wearable laser pointer was in-

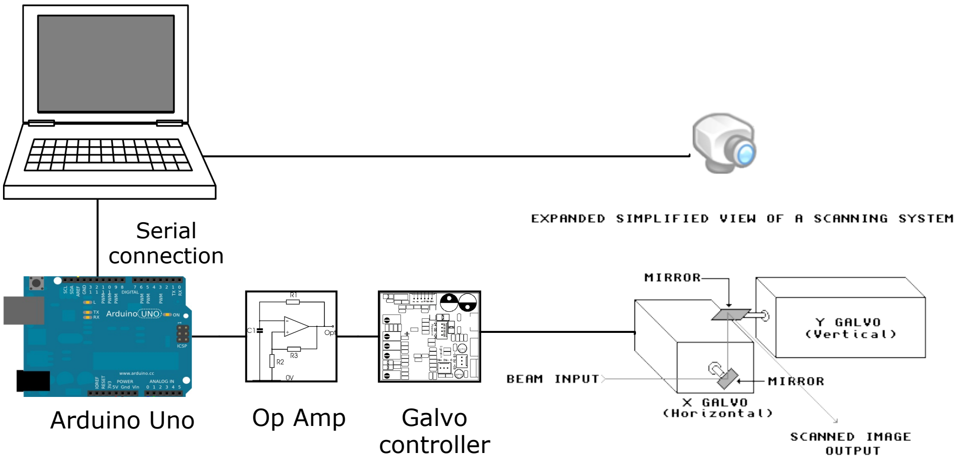


Figure 4.6: *System architecture of the wearable laser pointer.*

vestigated through a lab study where the performance of the participants is measured in a remote collaboration task. I compared the user performance in two different conditions: 1) the user had a HMD to see the provided guidance by a remote person, 2) the user had a helmet-mounted laser pointer drawing sketches directly on the physical objects in order to direct the user's visual attention towards those physical objects (see Figure 4.7). Details of the lab experiment is discussed in Part II - Paper 3.

4.2.1 Advantages of the wearable laser pointer

First of all, in projection-based augmented reality, the computer generated content is superimposed directly onto physical objects. Instead, in HMD-based augmented reality systems, the user needs to view the computer generated graphics and the real world image through the HMD. This indirect view of the real world causes some problems such as eye fatigue and focusing problems for users of the HMDs.

The second advantage of wearable laser pointers is the possibility of sharing information in collaborative scenarios an advantage it shows with projection-based displays. However, the laser beam is more visible compared to the state of the art projectors' light. This is an important strength for wearable laser pointers specially in outdoor applications in daylight.

4.2.2 Limitations of the wearable laser pointer

Apart from the above-mentioned advantages there are some technical limitations for wearable laser pointers. The first challenge is the form factor. The existing motor-controlled wearable laser pointers are bulky and hard

Figure 4.7: *The right picture illustrates a closer view of the galvo scanners, the middle image shows a user wearing the helmet-mounted motorized laser pointer while the left picture shows a user with the HMD [83].*

to carry. The second technical challenge of using wearable laser pointers is instability of the laser beam: the projected content moves when the wearer's head does. This is a problem if the intention is to rest the point of the beam on a specific nearby object. Similar to augmented reality user interfaces on HMDs and other mobile displays, the visual content in laser pointers also needs to be registered in the real world.

Another important issue is the limitation in displaying complex visual content. Since in laser pointers the visual content should be drawn by moving two motor-controlled mirrors, in order to display complex symbols (e.g. long textual information) the laser pointer needs to move faster, otherwise the quality of the displayed visual content decreases. In my prototype, I used a pair of 20 KHz galvanometers controlled by an Uno-Arduino microcontroller. Using this setting, there was no problem to display polygons with the size of 30 degrees and about 10 edges. Figure 4.8 illustrates some sample contents projected by the wearable laser pointer on the wall.

In laser pointer systems, a front-facing (scene) camera needs to stream the image of the environment in order to register the virtual content in the real world. due to the distance between the scene camera and the laser pointer mirrors, there is always a potential displacement between the targeted points by the system and the actual laser position in the real world. This displacement is called *parallax error*. One strategy to minimize the parallax error is to calibrate the system for different distances and use a depth sensor to adapt the calibration. Our approach was to place the camera very close to the laser pointer mirrors (<1cm) to minimize the parallax error and calibrate the system for an average distance (2m) resulting in an accuracy of about <1 degree in the range of 1 to 5m.

Another limitation of wearable laser pointers is the significant power consumption and need for heat dissemination. In our prototype we used a 15 volts power supply and a heat sink to get rid of the extra heat generated

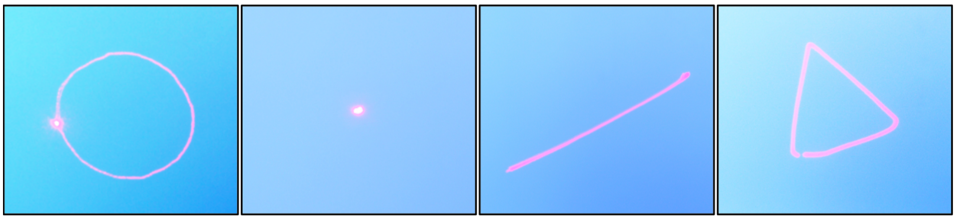


Figure 4.8: *Visual content displayed by the wearable laser pointer on the wall.*

by the controlling unit. In state of the art eyewear computers the device works with a 5 volts battery (for about 45 minutes in Google Glass) which is much lower than my wearable laser pointer.

Finally, the strong laser beams are dangerous for nearby people. The eye safety restrictions forces us to use the low power laser which is not bright enough for the daylight condition.

§ 4.3 Discussion

4.3.1 The WPA concept

The simple and flexible laser projection mechanism in laser pointers enables WPAs to direct users' visual attention to a certain object in the real world. Aside from providing explicit visual content (path 14-2-4-6-8 in Figure 2.3 - page 28), by decreasing the intensity of the visual stimulus the laser pointer would be able to provide subliminal cues to direct the visual attention unconsciously (path 14-2-4-9 in Figure 2.3, page 28). This *unconscious interaction* could reduce the negative effects of explicitly interrupting users (section 2.8.1). Providing such subliminal visual cues can be technically much harder on HMDs since to display subliminal cue the refresh rate of the display needs to be relatively high (about 100 Hz) which is not available in state of the art HMDs.

4.3.2 The WPA for clinicians

Our comparative study between a motor-controlled laser pointer and a HMD in a remote collaboration task (Part II - Paper 3) [83] showed the potential of the laser pointer to solve some of the well-known challenges of HMDs such as the focusing problem and eye fatigue. In fact laser pointers can be used as a complementary output device by WPAs for displaying simple visual contents in situ. For example, a wearable laser pointer could be used as an additional visual display in our tele-presence scenario in the surgery room. The laser pointer could be controlled by the remote expert to point the laser beam directly onto the surgical site; however, for such remote pointing we need to have a live video stream between local and remote surgeons. Such a pointing mechanism would let other clinicians in the surgery room also see the laser point. This can be an advantage over the HMDs which are visible only for one person.

- Chapter 5 -

Body gestures as an input modality for WPAs

In this chapter, I briefly review the previous work on gesture-based interaction with wearable computers. Then I discuss some of the usability challenges of gesture-based interaction and explain how our project on gesture-based interaction with medical images in the operation theatre addresses some of those challenges. This chapter tries to answer the body gesture part of the RQ4 (page 5).

§ 5.1 Related work

We use body gestures to communicate with each other, and even before we learn how to talk as newborn we are able to express ourselves through body gestures. "A gesture is non-verbal communication made with a part of the body" [20].

The use of gestures for human-computer interaction has always been an interesting topic in the HCI community (e.g. [26, 159, 217, 95, 39]). From a user-experience perspective, gesture-based interaction has been widely accepted as an intuitive and robust input mechanism in gaming applications [1]. Still new technologies, such as the MYO Gesture Control Armband[2], the Leap Motion Sensor[3], and compact TOF cameras [4] are being explored for gesture-based interaction (Figure 5.1). While researchers have proved that 3D gestural input is as effective as touch input for mobile devices [86], the effectiveness of the gesture input for everyday interaction with computers is still under research.

Even though the most popular application for gestural control systems is computer gaming, the possibility of touch-less interaction makes gestural

[1]Microsoft Kinect Sensor. www.xbox.com/en-US/kinect
[2]https://www.myo.com/
[3]www.leapmotion.com/product
[4]http://www.opticsbalzers.com/

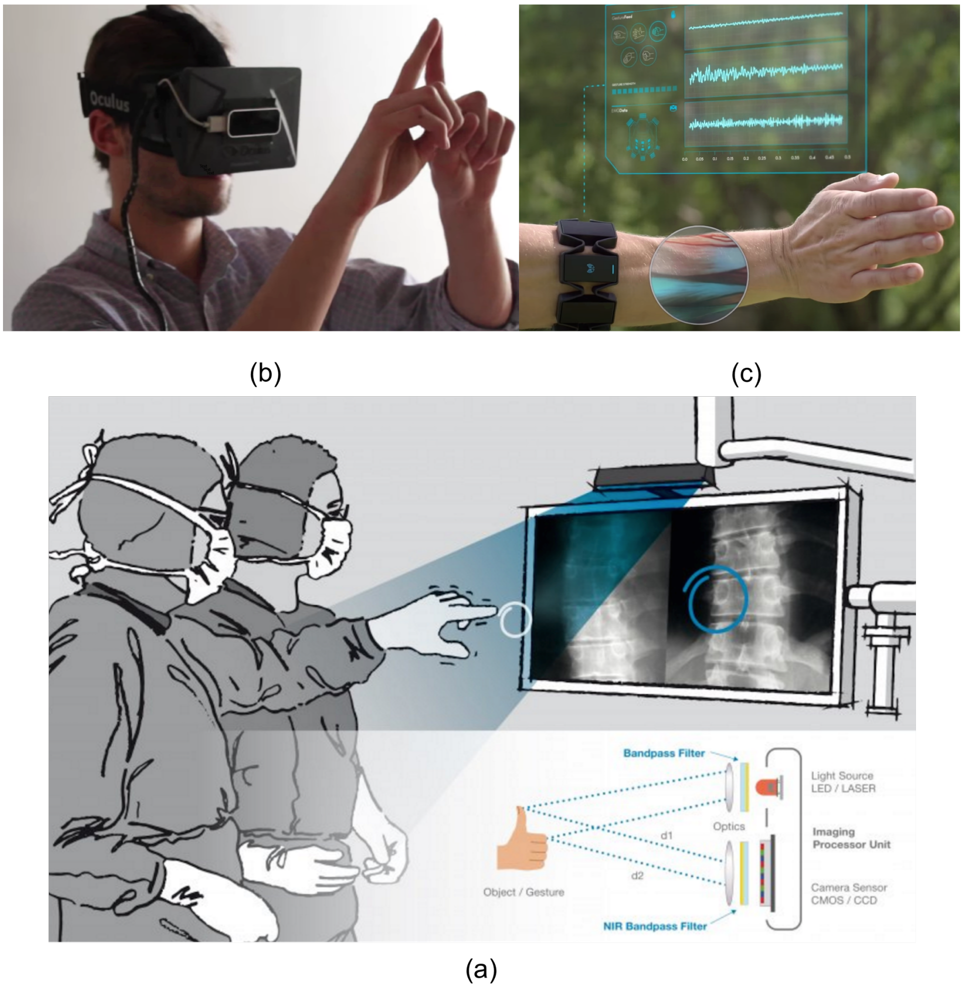


Figure 5.1: *a) Opticsbalzers TOF camera for gesture recognition in surgery rooms, b) Leap Motion sensor for interaction with virtual reality applications, and c) Myo sensor detects hand gestures by analyzing electrical signals from muscles (EMG).*

input a viable modality for interaction in sterile environments (e.g. [205, 51, 165, 141]) and interaction with wearable computers (e.g [39, 95, 97, 14]) because the gesture-based interaction leaves users' hands free for performing real-world tasks.

Body gestures can be used either the only input modality (e.g. [39, 95, 159]) or they can be combined with other modalities such as speech commands [26] and eye gaze [98] for multimodal interaction. In the following sections of the related work, the possible technological approaches to gesture recognition is briefly explained, and previous work on three types of body gesture (hand, foot, and head gestures) is reviewed.

5.1.1 Gesture recognition approaches

In general, body gestures could be detected in different ways from using wearable sensors to environmental sensors.

5.1.1.1 Wearable sensors

The wearable sensors can be inertial sensors such as accelerometer[214], gyroscope[219], or combination of them [21]. Another approach to detect body movements is to use wearable cameras on the wrist [95], foot [14], head [131], shoulders [66], or the whole body [179]. It is also possible to detect touch-based gestures based on wearable capacitive sensing [159, 169].

5.1.1.2 Sensors in the environment

The sensors embedded in the environment can be vision-based (e.g. [218, 10, 205, 60, 187, 165, 51]), pressure sensors [146], capacitive sensors [218], or light refraction [10]. The vision-based gesture recognition can be done through regular webcams [205], a stereo camera [60], a time of flight camera [187], or the Microsoft Kinect [165, 51]. The latter is becoming increasingly popular thanks to its low cost and easy implementation.

In the vision-based approach with cameras embedded in the environment, the user does not need to wear any additional device. However, a direct line of sight is needed for the interaction, where the users typically have to hold their hand in an unnatural position in order for the system to detect the gestures. Non-vision-based sensing approaches such as inertial wearable sensors, capacitive, and pressure sensors pose a good alternative to vision-based systems, as they do not require a direct line of sight. Using inertial wearable sensors for gesture recognition only allows a designated person to interact with the system, avoiding the potential confusion associated with having multiple people in the room the system is deployed in. However, wearable sensors need to physically be mounted on the user's body which might decrease acceptability of such systems by users.

In gesture-based interaction, it is important to differentiate between gestures performed as intentional input the interactive system and gestures that take place as part of other activities (and should be ignored by the interactive system). One approach is to detect dedicated, easily distinguishable, gestures out of a continuous data stream [87] or enabling the system with another modality (clutch mechanism) such as user voice commands [21].

5.1.2 Hand gestures

A survey [20] on body-gesture-based interaction indicated that using hand gestures are the most dominant type of gestures used in gesture-based interactive systems. Using gloves to recognize hand gestures is among the earliest gesture recognitions approaches. The Data Glove [217] was able to monitor ten finger joints and the six degrees of freedom (6 DOFs) of the hand movements , SCURRY glove [97] enabled users to type on a virtual keyboard, and Acceleration Sensing Glove (ASG) [152] was able to recognize pseudo static gestures and be used for mouse pointing.

As it was explained in the previous section, using wearable inertial motion sensors is another popular approach for detecting hand gestures (e.g. [214, 219, 21]).

A big part of the hand gesture recognition systems have been developed for interaction with wearable computers. For instance, the GestureWrist system [159] was a capacitive sensor placed around the wrist to detect change of the arm shape and hand gestures for interaction with wearable computers. GesturePad [159] was another form of capacitive sensor for interactive clothing that allowed users to provide commands to a wearable computer by touching different parts of their clothes. Since wearable computers have limited processing resources and battery, using vision-based gesture recognition is not easy on wearable computers since these approaches usually demand heavy processing and a lot of power. One approach to minimize the processing and power consumption burden on a wearable computer and increasing the accuracy of the gesture recognition is to add a gesture recognition component to the wearable computers. Mime [39] is a low-power 3D sensor component suitable for gesture-based interaction with HMDs and eyewear computers. Mime [39] includes a front-view time-of-flight (TOF) module with a standard RGB camera to position the users hand in a 3D space (see Figure 5.2). In the Mime system [39], the scene is illuminated by a pulsed LED and the reflected light is time-sampled. The samples are processed to calculate hand coordinates in 3D which is used to recognize hand gestures.

Another example for such gesture recognition systems is Digits [95] that uses a wrist-worn camera to recover the 3D pose of the user's hand. Finally, ShoeSens [14] uses a shoe-mounted Microsoft Kinect as a sensor for detecting hand gestures.

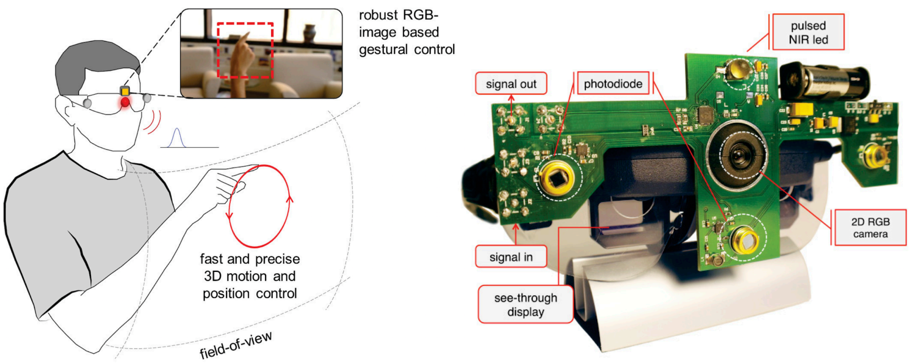


Figure 5.2: *Mime can be added as a peripheral component to eyewear computers for gesture-based interaction [39].*

5.1.3 Foot gestures

The utility of foot gesture-based interaction can be understood easier if we think about a mobile scenario, where the user's hands are dirty or pre-occupied (with e.g. shopping bags) and the user wants to interact with a mobile or wearable device. In non-mobile scenarios, we use our feet for interaction with machines (e.g. while driving a car). Moreover, modern computer games such as Microsoft Kinect and Nintendo Wii have brought the foot-based interaction to everyday life. While previous studies [144] showed that human's foot is capable of performing complex movements and can be quick in performing coarse level actions, it is also proven that foot is less precise than hands and fingers [148].

In HCI community, the foot gestures have been used for entertainment and gaming [147, 143], ambient awareness [163], user identification via gait analysis [72], navigating through documents [174], or providing input to interactive systems [215, 40, 176, 64].

Foot gesture recognition can be accomplished through either on-body sensors (e.g. accelerometer [6] or insole embedded sensors [200]) or environmental sensors (i.e. sensitive floors). The latter category features different technologies including pressure [146], light refraction [10], or capacitive [218]. While capacitive approaches usually have less resolution, when compared to the others, these do not require for a soft or transparent surface, and easily allow for swipe-like foot gestures.

5.1.4 Head gestures (movements)

Humans use head nods and head shakes in their face to face conversations to express their agreement or disagreement [134]. Head movement has also been used for interaction with computers as the only input modality (e.g.

[122, 133]) or in combination with other modalities such as gaze [216, 188, 78] or hand gestures [7]. It is proven that using head movements for interaction is easy to learn [19] and accurate [79] way for interacting with computers. However, the mass of the head can reduce the speed of interaction, and it can be tiring for the neck muscles [53]. The head-based interaction can broadly be categorized into into discrete and continuous. One example for continuous use of head movements in interaction is head pointing where the horizontal and vertical head movements are translated to the horizontal and vertical movements of the cursor in a graphical user interface [216, 188, 78]. While head nods and shakes can be interpreted as discrete commands to the system for dialog boxes and document browsing in Windows [133], controlling a wheelchair [62], interaction with smart home [122], or interaction with automotive information systems [7].

Head movements can be detected by a camera [133, 122, 96] or other wearable inertial sensors [52, 36]. Most of the camera-based approaches track eye position for head gesture recognition. Different classifiers have been used for head gesture recognition such as Hidden Markov Model [89] and SVM [133]. Mardanbegi et al. [122] used an eye-tracker to detect head gestures by tracking pupil rotation that happens during head movements. Most of the modern eyewear computers such as Google Glass and Vuzix Smart Glass are equipped with 9 degree of freedom inertial sensors that enables them to detect head movements accurately.

§ 5.2 Challenges of gesture-based interaction

While using body-gestures can be an interesting touch-less modality for interaction with computers on the move or in parallel with real-world tasks, there are some well-known usability challenges associated with gesture-based interaction. First of all in gesture-based interaction similar to all in-air interaction techniques there is no haptic feedback which has a negative effect on user performance and usability of the system [73]. The second problem is the need for a clutch mechanism to differentiate the user's actions that are intended to be a command to the system from those that are not. Another challenge is related to the social aspects of the interaction where a user needs to use the system in presence of other people (e.g. in collaborative scenarios). Since body gestures are used in human to human interaction, performing gestures for interaction with computers might be confused by other people or might seem awkward and reduce the acceptability of the system [3].

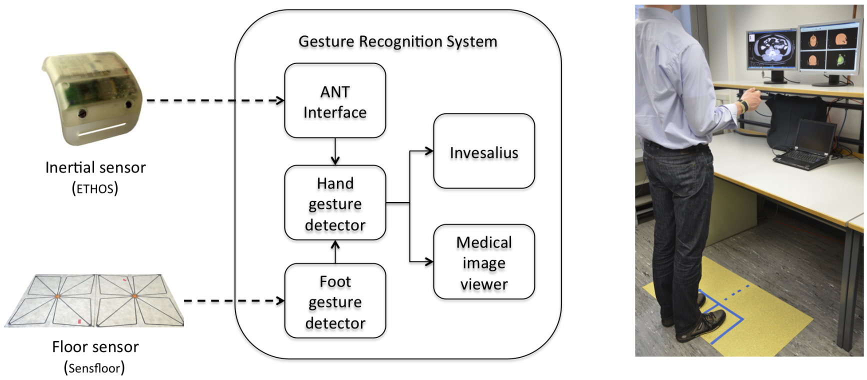


Figure 5.3: *A participant interacting with medical images [84].*

§ 5.3 Hand & foot gestures for touch-less interaction in hospitals

Due to the sterility restrictions in operation theatres, gesture-based interaction with medical systems in surgical settings can be a big advantage. To investigate the utility of using hand and foot gestures for interaction with medical images in the surgery room, we designed and implemented a system that helps orthopedic surgeons interact with 2D X-ray and 3D CT scans displayed on two different displays through hand and foot gestures. We used a single wearable sensor including accelerometer and gyroscope to detect 6 different hand gestures. To address the clutch mechanism problem, we investigated the utility of foot gestures for starting and finishing the interaction, and switching the interaction mode from one system to another one. To detect the foot gestures we used capacitive sensors embedded into the floor. Figure 5.3 represents the system architecture and a user interacting with the system through hand and foot gestures. Assuming that in the future, places like hospitals would be equipped with such sensors, we investigated the feasibility of using such sensors for mobile interaction in hospitals. Since in a hospital clinicians mostly move within the building of the hospital (*local mobility* [17]) if the whole building is covered by such capacitive sensors, future WPAs for clinicians can utilize such an infrastructure for foot-gesture-based interaction. More details about the study can be found in Part II-Paper 4.

§ 5.4 Discussion

5.4.1 The WPA concept

Hand-gesture-based interaction can be a viable input modality for WPAs when the user's hands are dirty (e.g. during a surgery or while cooking) or when the user needs to hold a tool in his/her hands; however, the user's hands are still involved temporarily in the interaction. Foot-based interaction can be a good complement for hand gestures when the user is standing still and not walking (e.g. during a surgery or while cooking). An context-aware adaptive user interface can switch the interaction modality from foot gesture mode to hand gesture mode when the user starts walking.

5.4.2 The WPA for clinicians

Since eyewear computers are already equipped with relatively accurate head tracking sensors, it is feasible to use head movements for interaction with eyewear-based WPAs without adding any equipment. In the hospital study explained in Chapter 3, the head movements of the surgeons is used for continuous interaction with the computers in the surgery room. In this case, the WPA acts like a proxy for interaction between the user and other computing devices. If the floor of the room was equipped with embedded capacitive sensor (e.g. SensFloor [104]), the WPA could also provide a foot-gesture-based input modality for interaction with the WPA or other computers in the hospital.

- Chapter 6 -

Exploring Eye Movements for Interaction with WPAs

Using eye movements for interaction with computers goes back to the early 1990 when Robert Jacob [75] explored eye gaze as an input modality for pointing, moving screen objects, and menu selection in desktop computers. But eye-based interaction is not limited only to the desktop computers. Mobile eye trackers are developed to bring eye-based interaction to people's everyday life where mobility is an inseparable part of everyday activities. Since wearable personal assistants are supposed to be used on the move and in parallel with real-world tasks, utilising eye movements as a touch-less input modality can be a big advantage for WPAs. Aside from using eye movements as an input modality, the eye movement data can be used for context recognition [29]. The context inference from eye movements can be a classical activity recognition [32] or more complex aspects of the user context like cognitive state [31]. These contextual clues can be extremely valuable for WPAs to provide relevant information to the task in hand through an appropriate modality in a right moment. In this chapter, I briefly review previous work and discuss challenges of eye-based interaction with wearable computers. Then I explain the contributions of this thesis to address some the discussed challenges. This chapter answers to the eye-movement part of the RQ4 (page 3).

§ 6.1 Related work

In the related work section, first of all the previous work on gaze-based interaction with HMDs is reviewed since the focus of this dissertation is eyewear-based WPAs which mostly include an HMD as the main visual output device. Moreover, the existing technologies for mobile eye tracking are reviewed since for eye-based interaction with WPAs we need a mobile eye tracker integrated with the WPA. Finally, previous work on eye-based interaction techniques are reviewed to reveal the strengths and weaknesses of each interaction technique for the envisioned WPAs.

6.1.1 Gaze-based interaction with HMDs

Eye gaze have been previously used for interaction with HMDs in virtual reality (e.g. [198, 142, 43]) and augmented reality (e.g. [137, 105, 5]) applications. The main purpose of implementing gaze in an augmented reality system is to detect users' point of interest in the field of view. This information can be used for predicting the user's intentions and anticipating what the user will request from the system [137, 105, 5]. In virtual reality applications, gaze can be used for selecting virtual objects in a virtual environment [198] or adapting virtual environment based on the user's gaze direction [142, 43].

6.1.2 Mobile eye tracking techniques

The most popular approach for both mobile (head-mounted) and remote (stationary) eye tracking is video-based method. A typical setting for video-based eye tracking includes relatively inexpensive cameras and image processing hardware and software to calculate the point of regard in real-time. To compute the point of regard, the corneal reflection of the light source which is typically infra-red is measured relative to the location of the pupil center. [50]

During the mid-1970s, electrooculography (EOG) (e.g. [30]) was the most popular approach for tracking eye movements [50]. A typical EOG system includes electrodes placed around the ocular cavity to measure changes in the static fields due to the eye's dipole potential [29]. By analyzing the signal generated from changes in the electric potential field, the system can detect relative eye movements.

6.1.3 Eye-based interaction

Using eye gaze as an input modality for computing devices has long been a topic of interest in HCI community, and it is due to the fact that humans naturally tend to direct eyes toward the target of interest. Eye gaze can be used both as an explicit and implicit input modality. "Implicit input are actions and behaviors of humans, which are done to achieve a goal and are not primarily regarded as interaction with a computer, but captured, recognized, and interpreted by a computer system as input [171]". While explicit input are our intended commands to the system through mouse, keyboard, voice commands, body gestures, and etc.

6.1.3.1 Eye gaze for explicit input

One of the most explored explicit ways of using gaze to interact with computers is to use eye gaze as a direct pointing modality instead of mouse in a target acquisition task [75]. The target can be selected either by fixating the gaze for a while on a particular area (dwell-time) [180] or using a mouse click [79]. However, controlling cursor with eye movements is limited to pointing towards big targets due to the inaccuracy of gaze tracking methods and subconscious jittery motions of the eyes[216]. Eye-gesture is another explicit approach for gaze-based interaction where user performs predefined eye-strokes[47]. Previous studies [19, 79] have shown that using eye gaze as an explicit input modality is not always a convenient method for users. In fact, overloading eyes as hummans' perceptual channel with a motor control task is not convenient [216].

6.1.3.2 Eye gaze for implicit input

In implicit method of using gaze in user interface design, natural movements of the eyes can be used to detect context, for example looking at certain objects in an environment can reveal interest of humans to those objects [116]. Eye gaze can also be used to infer about user's behaviour, for instance which objects attracts user attention during an everyday activity like cooking [164]. Another example of using eye gaze as an implicit input is to detect user's attention point and react to the users eye contact [178], or adapt user interface behavior [70] accordingly. The gaze data can also be used indirectly for interaction purposes [216, 78, 122, 203]. For instance, in the MAGIC pointing technique [216, 78], eye gaze data is used to move the cursor as close as possible to the target. Mardanbegi et al. [122] proposed a gaze-based interaction technique where the gaze data is used indirectly for head-gesture recognition. Another implicit way of using gaze data is Pursuits interaction technique [203] which enables users to select an object on the screen by correlating eye pursuit movements with objects moving on the screen.

§ 6.2 Challenges of gaze-based interaction with eyewear computers

Previous work indicates three main challenges for both stationary [117] and mobile [29] gaze-based interaction: 1) eye-tracking accuracy, 2) calibration drift, and 3) the Midas touch problem (the problem of recognizing the user's intentional gaze input from other eye movements). These challenges are also applicable to the gaze-based interaction with WPAs. In addition to these three classical eye interaction challenges, integrating eye trackers with

WPAs have some hardware and software complexities which are explained as follows.

6.2.1 Hardware & software challenges

As I explained in section 4.1.1, HMDs can be monocular, binocular, video-see-through, optical-see-through, or non-see-through. Since being able to see the real world is very important factor in real world tasks, binocular non-see-through HMDs which cover the whole field of view of the user do not seem to be a good option for WPAs. Also due to the floating effect and latency in the live video feed, video-see-through HMDs are not still appropriate for real-world applications. Therefore, the only acceptable form factors for a HMD as part of a WPA are monocular and binocular optical-see-through HMDs or monocular non-see-through HMDs. Optical see-through HMDs are relatively expensive and not easy to use in daylight condition. Thus the only feasible form factor for HMD of a WPA is monocular non-see-through HMDs which have been used in most of the recently produced eyewear computers such as Google Glass and Vizux Smart Glass. In the latter form factor, one eye is left free to observe the real world, and the HMD covers only a small part of the field of view (e.g. about 14 degrees in Google Glass) of only one eye. None of the commercially available eyewear computers are equipped with eye tracking systems. Only in Google Glass a near-eye infrared proximity sensor is installed to recognize blinking and if the user is looking at the HMD or not (see Figure). This proximity sensor cannot be used for calculating point of regard which is crucial for gaze-based interaction.

6.2.1.1 GlassGaze: an open-source gaze-tracker for Google Glass

To integrate an unobtrusive eye tracker into the Google Glass, we developed a hardware and software platform which is called GlassGaze [123]. The system architecture of the GlassGaze is illustrated in Figure 6.1. To detect the eye gaze, an external infrared camera was added to the Glass under the display. The camera sends the eye image wirelessly to a remote server [121]. The server analyzes the eye image in real-time and calculates the eye gaze coordination based on the user calibration data. The server sends the calculated gaze coordination to the client application on the Google Glass through WiFi connection. The client application receives the gaze data and adjusts the user interface accordingly. The GlassGaze client application on the Google Glass is released as an open-source Android application [1]. The Android application enables the user to calibrate the system for both gaze tracking on the display and gaze tracking on the scene. The accuracy of the GlassGaze system is about 0.5 degrees. The GlassGaze hardware and

[1] https://github.com/dmardanbeigi/GlassGaze

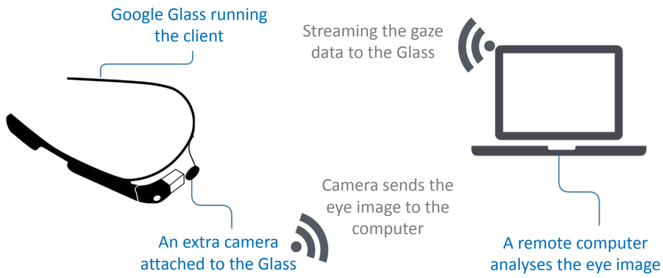


Figure 6.1: *System architecture of the GlassGaze [78].*

software platform is used for exploring eye-based interaction with WPAs explained in the next sections of this chapter.

6.2.1.2 EyeDroid: an open-source mobile gaze tracking system on Android

One of the main reasons that even the state of the art eyewear computers do not support eye-based interaction is the fact that the image processing required for gaze tracking is extremely complex and power demanding. Unfortunately, this computational demand is very far from what can be accomplished on existing eyewear devices such as Google Glass. Due to this limitation, we assigned the image processing task to a remote server in the GlassGaze application like other mobile gaze tracking systems (e.g. [106, 110, 157, 121]). However, this approach decreases the mobility of the user to a limited area in the range of wireless network coverage.

There are some commercial products from companies such as EyeTribe[2], Tobii[3], and Umoove[4] which support eye tracking on mobile and wearable devices. They usually provide a cellphone-size device for image processing. But the commercial mobile gaze trackers are usually so expensive and hard to afford. That is the reason why some of the recent studies have tried to use cheap and small processors such as Raspberry Pi [56] and micro-controllers [126] for eye tracking. Ferhat et.at [56] have presented a cheap eye tracking solution running on a Raspberry Pi device. They based their work on the open-source Opengazer [136]. The average gaze estimation error of their system is about 1.4° for an image size of 640 × 480 pixels with the frame rate of 3Hz. Although their system was running on a small device, it was only tested on a stationary setup for gaze tracking on a computer screen. The iShadow eye tracker by Mayberry et.al [126] focuses on head-mounted

[2] *https://theeyetribe.com/*
[3] *http://www.tobii.com/*
[4] *http://www.umoove.me/*

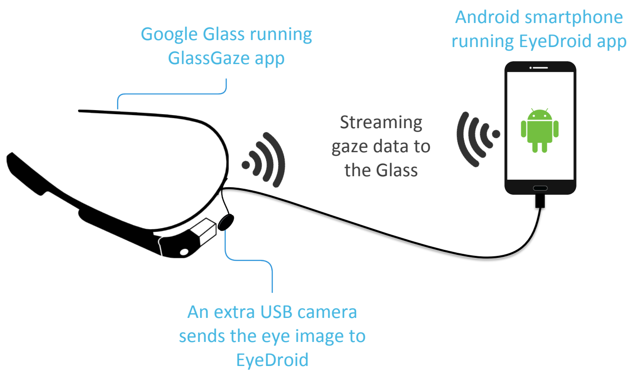


Figure 6.2: *A schematic view of the EyeDroid system architecture.*

gaze tracking. They have presented a fully mobile eye tracking solution using a low-power ARM Cortex M3 micro-controller. The iShadow system uses a very efficient image processing approach that can run on a small micro-processor. They achieved real-time gaze tracking in an image captured by a front-view camera with an error of about 3 degrees.

Since in eyewear computers (e.g. Google Glass) the size of the display is very small (less than 15 degrees), the accuracy of the eye tracker needs to be relatively high (compared to the stationary standard displays) in order to enable a graceful interaction. To achieve a higher accuracy in eye tracking (about 1 degree), we rely on the processing capacity of commonly used Android smartphones for processing eye images. This means in our GlassGaze system, we replaced the remote server [121] with an Android smart phone (see Figure 6.2). We implemented a we built and used Java Lightweight Processing Framework (JLPF)[5] to optimize the image processing load on the phone. The evaluation of the system indicated an accuracy of 1.06 degrees and a battery lifetime of approximate 4.5 hours. Details of the system implementation and evaluation is explained in the Part II-Paper 5.

6.2.2 Accuracy of eye trackers & usability challenges

Just like previous mass-market user interface paradigms used in smartphones and PCs, interaction with eyewear devices relies heavily on the visual modality. The typical use of gaze in graphical user interface is a pointing mechanism to control the cursor position on the screen [19, 100]. Gaze pointing

[5] *https : //github.com/centosGit/JLPF*

has also been explored for interaction with head-mounted displays [1, 101]; however, due to the inaccuracy of existing gaze tracking approaches and the subconscious jittery motions of eye [216], using eye-pointing is limited to the pointing towards large targets on the screen [19]. Aside from the target size limitation, eye-pointing has in several studies been found to be an inconvenient way of pointing [78, 19]. In fact, overloading the visual channel with a motor control task can be the main reason for eye-pointing to be recognized as an inconvenient pointing technique [216].

Since eyewear computers sit on the user's head and in front of his/her eyes, head and eye movements can be an used as input modalities for eyewear computers. To explore the advantages and limitations of using eye and head movements for interaction with WPAs, I investigated head and eye movements for pointing towards a graphical user interface on a HMD. The performance of users in head and eye pointing has been compared with mouse pointing as a baseline condition. The results of our experiment showed that the eye pointing is significantly faster than head and mouse pointing; however, our participants felt that the head pointing is more accurate and convenient. The details of the study can be found in Part II-Paper 6.

Based on findings from the above-mentioned study, we decided to combine head and eye movements for a target acquisition task. We extended the old idea of the MAGIC (Manual And Gaze Input Cascaded)-pointing [216] for eyewear computers. Our MAGIC pointing method utilizes eye movements implicitly for moving the cursor as close as possible to the target where the cursor is controlled by head movements for fine-grained adjustment. We conducted an experiment to compare the proposed MAGIC pointing approach with head pointing. We found that the proposed MAGIC approach benefits from both the speed of eye-pointing and the accuracy of head pointing. In addition, the MAGIC method decreases the amplitude of head-movements and thus ergonomic problems of head pointing towards long distances. More details about the experiment is explained in the Part II-Paper 7.

6.2.3 Calibration problem

Due to the differences between individual eye geometries existing eye trackers need to be calibrated for each user in order to enable precise gaze tracking that is often required in gaze-based interaction. Furthermore, different factors such as relative movements of the eye and the eye tracker, ambient light conditions, calibration quality affect the accuracy of gaze tracking during long term (a few hours) operation. In an eyewear computer the calibration problem can be even more serious because in eyewear computers there are two different spaces for eye-based interaction: 1) the display space, 2) the real-world space. This means the eye tracker needs to be calibrated for both spaces. That is the reason why calibration-free interaction techniques can be a big advantage specially for eyewear computers. To address the calibration challenge, we invented EyeGrip which is a calibration-free interaction technique for seamless interaction with scrolling contents in computer screens.

§ 6.3 EyeGrip: a calibration-free method for seamless interaction

6.3.1 Interaction with moving visual content

The fact that our brain processes images significantly faster than text [33] might be one of the reasons of why we are often more engaged with images than textual information in the Internet and why viewing pictures is among the most popular functions in the many computing devices such as mobile phones and even eyewear computers. For instance, in Google Glass due to the relatively small size of the screen, the user needs to frequently scroll among different *cards* even to select a menu item. This means when an application includes a lot of cards, it takes a lot of time to find the intended card among others.

6.3.2 EyeGrip for interaction with scrolling content

EyeGrip facilitates seamless interaction with eyewear computers and helps the user intuitively stop a sequence of moving visual contents displayed on the computer screen. The key idea behind EyeGrip is the fact that during a visual search task while we look at the scrolling visual contents, our eyes perform a combination of saccadic and smooth pursuit eye movements which is Optokinetic Nystagmus (OKN) eye movements. During OKN eye movements, it is very likely that more interesting contents particularly attract our visual attention. By tracking the eye movements in the same direction of the moving visual field, we see a sawtooth-like signal shape (Figure 6.3). By monitoring deviations in the sawtooth-like Optokinetic Nystagmus (OKN) eye movements, the system knows where on the sequence of the contents the user is visually interested in. Based on this information, the system can react immediately by e.g. stop scrolling and bringing back that interesting content in front of the user's eye.

6.3.3 Study characteristics of the EyeGrip method

We assessed the viability of EyeGrip in a scenario where the user looks at a sequence of images that is moving horizontally in the display while his/her eye movements are tracked by an eye tracker. We conducted an experiment that shows the performance of the proposed approach. We also investigated

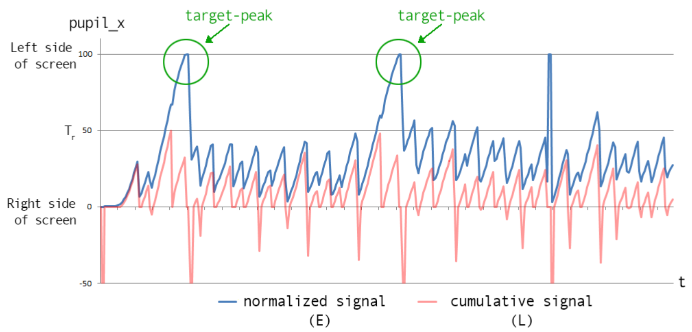


Figure 6.3: *The blue color signal (E) is generated from horizontal eye movements in a visual search task among uni-directional moving objects that move from right to the left side of the screen. The red color signal (L) is the signal generated in the classification process. The target-peak is detected when L exceeds the T_r limit.*

the influence of speed and maximum number of visible images per frame on the accuracy of EyeGrip. Based on the experiment results, we proposed guidelines for designing EyeGrip-based interfaces. Our study showed that by selecting appropriate speed and maximum number of visible images per frame, the proposed method can be used in a fast scrolling task where the system accurately (87%) detects the moving images that are visually appealing to the user, stops the scrolling, and brings the item(s) of interest back to the screen. More details about the study can be found in Part II-Paper 8.

6.3.4 EyeGrip for interaction with eyewear computers

We also empirically evaluated the usability of the EyeGrip for seamless interaction with eyewear computers through implementation and evaluation of two different applications for eyewear devices: 1) a menu-scroll-viewer system and 2) a Facebook [6] newsfeed reader [77]. The results of our usability study showed that the EyeGrip technique performs as good as keyboard which has long been a well-known and easy way of providing input to computers. Even in some usability aspects such as speed and intuitiveness, EyeGrip outperforms the manual method. Moreover, the accuracy of the EyeGrip method for menu item selection was even higher than manual method. However in the Facebook study, the participants did not find EyeGrip as accurate as the manual method due to the involuntary nature of the interaction. The details about the study can be found in Part II-Paper 9.

[6] www.facebook.com

To the best of our knowledge there has been no studies investigating the use of OKN eye movements in visual search tasks and interaction with dynamic user interfaces.

§ 6.4 Discussion

6.4.1 The WPA for clinicians

My exploration on utilizing eye movements for interaction with WPAs, revealed the great potential of eye gaze as a touch-less input modality for WPAs. However, in addition to the classical challenges of eye-based interaction such as accuracy, calibration drift, and Midas touch problem, integrating eye trackers with WPAs is associated with technological complexities. The most important technological challenges are building an unobtrusive hardware platform with minimum coverage of the user's field of view and optimizing eye tracking algorithms for wearable computers with limited processing resources. Our GlassGaze hardware and software platform together with EyeDriod eye tracking system adds eye-tracking capability to the Google Glass as a commercially available eyewear computer. This modified version of the Google Glass can be a better platfrom for the WPA that I developed for orthopedic surgeons because it provides a completely new touch-less modality that can be used to provide input to the WPA. For example, the EyeGrip method can be a viable alternative to touch gestures and voice commands to select a menu item or even find medical images faster without touching the touch-pad.

6.4.2 The WPA concept

Since OKN is a natural reaction of the eyes to the moving visual content in the field of view, EyeGrip opens room for more intuitive methods of eye-based interaction with WPAs. The EyeGrip technique exploits a completely new HCI path (3-4-9-15) in our body-and-mind-centric model of the WPA (Figure 2.3, page 28). In this new path, the WPA displays the scrolling content at a certain speed in front of the user's eyes (3). The unconscious part of the user's cognition perceives the visual scrolling contents (4). OKN reflections happen unconsciously (9) when the user is looking at the scrolling content. By monitoring the OKN eye movements, the WPA detects the object of interest and stops scrolling (15). This very simplified explanation of the EyeGrip method based on our body-and-mind-centric model of the WPA indicates the potential of the EyeGrip technique for removing the conscious cognitive processing during the interaction which can potentially demand less cognitive resources for interaction.

- Chapter 7 -

Interruption Management & Wearable Personal Assistants

In this chapter, first I explain the challenges of interruption handling in wearable devices. Then a brief review of the previous approaches on interruption handling in wearable computers is presented. In the next section, I introduce the concept of self-mitigated interruption that is our novel approach to interruption handling in WPAs. Finally in the discussion section, I explain the concept of self-mitigated interruption based on our human-body-and-mind-centric model of the WPA (Figure 2.3, page 28), and I discuss how our findings addresses some of the important challenges of designing WPAs for clinicians. This chapter answers to RQ3 (page 5).

§ 7.1 Interruption handling in mobile & wearable computers

Interruption is a well-studied topic in the HCI community, and several studies defined the term "interruption". In this thesis, I adhere to Boehm-Davis and Remington's definition who define an interruption to be *"the suspension of one stream of work prior to completion, with the intent of returning to and completing the original stream of work"* (p. 1125) [25]. While interruptions caused by external events can be regarded as natural and unavoidable in the real world as part of everyday life, the research community has recognized a lot of negative side effects that interruptions can have such as reducing performance [2], decreasing the ability to drive safely [199], and increasing accidents by healthcare providers [63, 207]. Mark et al. found that workers compensate for interruptions by increasing work rate and maintaining work quality but experiencing an increase in stress, frustration, time pressure and increase in effort [124].

7.1.1 Challenges of interruption handling for WPAs

In stationary systems, interruptions can decrease users' performance while in a mobile scenario an improper interruption can cause a safety critical issue (e.g. a car accident). The following items clarify the complexities of interruption handling for the envisioned WPAs.

WPAs are hard to ignore: Wearable personal assistants are envisioned as always-on systems to support users in everyday activities. This means WPAs are supposed to be continuously present in the users *perception space* [150] (e.g. a HMD that covers part of the user's field of view). Therefore, in WPAs system notifications can distract users from the task at hand easier than other computing devices such as smartphones. In fact when the existing smartphones are in silent mode and placed out of the user's perception space (e.g. inside the user's pocket), they hardly can interrupt their users.

WPAs need to interrupt within a shorter time: Since the envisioned WPAs are to be used in parallel with everyday tasks, it is crucial that the user attends to the notifications that might have an important piece of information for the task at hand *in time* [211]. This increases the challenge of interruption handling for WPAs since they have simply a shorter time period to find the right moment for interrupting users compared to stationary computers. This is due the fact that in mobile scenarios users switch more frequently between different tasks compared to stationary settings [211]. Usually determining a good time to interrupt requires a complex assessment of context and message content [68]. A complementary approach to minimize the negative effect of such unavoidable interruptions is to communicate the message through an appropriate modality that has less overlap with the perceptual channels occupied by the ongoing task [181, 168].

WPAs proactive or annoying? WPAs are envisioned to proactively provide relevant information to the current situation. This means such proactive WPAs sometimes will initiate the interaction and interrupt users more often than existing mobile devices. Each time the WPA proactively provides information, it needs to compete for the user's attention and possibly interrupt the ongoing task. Even today wearable and mobile computers are running context-aware services such as location or activity-based reminders. But as we know, such systems do not make flawless decisions in determining when and what information should be presented. Due to the same reason in the future, the number of unwanted interruptions increases as the WPAs will potentially be able to sense more context data.

WPAs in collaboration: One of the challenges in some work domains such as hospitals where people need to be mobile and at the same time collaborate with each other is to find a particular person. Usually in such occasions, people use mobile devices or pagers (WPAs in the future) to call each other. But these devices are inherently interruptive, and there is

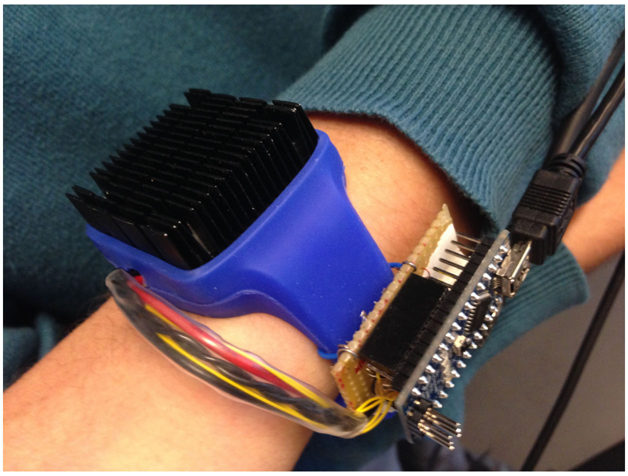


Figure 7.1: *Wearable thermal haptic notifier system [27].*

contradiction between being available for other colleagues and performing their task fluently [18, 81]. Since WPAs are hard to ignore, using WPAs in such work domains can increase unwanted interruptions.

7.1.2 Opportunities for interruption handling by WPAs

Apart from challenges of interruption handling in wearable computers, there are some unique opportunities for wearable computers through which some of the challenges of interruption handling can be addressed.

Sensing context through wearable sensors: In contrast to desktop computers and smartphones, wearable computers can benefit from body-worn sensors (e.g. inertial motion sensors) for fine-grained context recognition (e.g. activity recognition [93]) to predict users' interruptibility and avoid unwanted interruptions (e.g. [99, 34, 192]).

Wearable output devices for more graceful interruption: Wearable computers are physically closer to the human body more than any other computing device; therefore, they can be a hardware platform for haptic output devices. Since the haptic displays are able to provide subtle notifications, they can be used as an output modality for graceful interruptions [145]. We developed and evaluated a wrist-worn thermo-haptic device (Figure 7.1 - Paper 10) for graceful interruption that enables the system to display thermal stimulus (both heat and cool) with an adjustable intensity. Temperature differences are produced by the Peltier principle by passing electrical current through a Thermoelectric cooler. More details about our wrist-worn device can be found in Part II - Paper 10.

§ 7.2 Related work

Previous approaches to interruption handling in mobile devices have mainly focused on managing phone calls (e.g. [57, 186]) and notifications (e.g. [151, 166, 68]) through analyzing user context and predicting interruptibility of the user when receiving a phone call or notification. By predicting interruptibility the system decides to postpone the interruption to another time or interrupt the user. Interruptibility estimation based on the user's context have also been explored for handling interruptions in wearable computers [94]. However, designing intelligent systems for managing interruptibility is not an easy task due to the complexities in types and objectives of interruptions, the diversity of contextual elements that need to analyzed [202], changing user behaviour over time that decreases the accuracy of the context recognition (the concept drift problem)[185], and challenges of detecting importance of message based on e.g. communication history between the two parties [61].

Another approach to minimize the negative effect of interruption in wearable computers is to design for parallel involvement of the user in performing real-world tasks and interacting with computers instead of only focusing on the context recognition [211]. McFarlane [127] introduced four primary methods for coordinating interruption when user involves in a dual-task situation: *immediate, negotiated, mediated,* and *scheduled*. Drugge et al. [49] investigated the effect of interrupting users of wearable computers through scheduled and negotiated interruption methods where the users had to play a game on a desktop computer as the primary task. The interruption task was a simple color and shape matching task displayed on the HMD. According to their findings the scheduled method has less negative effect on performance of the users in both tasks. A similar study [211] with a more realistic primary task (hotwire game [212]) reported the negotiated method as the worst approach for interrupting users of wearable computers.

Apart from above approaches for managing interruptions, there are other strategies in system design that indirectly addresses the interruption problem. For example, one strategy to reduce the negative effects of interruptions in mobile scenarios is to minimize the interaction duration: "Microinteractions are interactions with a device that take less than four seconds to initiate and complete."[9]. Microinteractions can potentially be beneficial by letting users interact with their mobile devices in parallel with ongoing primary tasks. This can expand the set of tasks that users of wearable computers can perform on-the-go [213]. The main idea behind Microinteractions is to minimize the time that user needs to spend for an interruption generated by a mobile computing device (e.g a WPA). Minimizing the time duration of an interruption means a shorter pause in the main task at hand, and the user can resume the interrupted task more easily.

Another example for such indirect approaches in reducing negative effects of interruption is to minimize the need for providing explicit input to the system by sensing context (e.g. human activities) and feeding *implicit input to the system* instead of involving users in the loop [173, 172]. This approach can reduce number of interruptions that might happen due to the need for providing explicit input to the system.

A similar idea can be applied to the system output. If the system output is implied implicitly to the user, there is no need for interrupting the user for displaying explicit output. *Subliminal cueing* [8] is one of the possible ways of presenting implicit output to a user. The methods and effects of providing subliminal data on human's cognition and behavior is more studied in Psychophysics. However, recently HCI community started exploring subliminal cueing in user interface design (e.g. [8, 153, 160]).

In this thesis, as an alternative to the previous approaches in interruption handling, we introduced the *self-mitigated interruption* concept defined based on a symbiotic interplay between the existing interruption management infrastructure in our brains on the one hand, and the system on the other hand. The self-mitigated interruption concept is explained briefly in the next section. More details about the self-mitigated interruption method is presented in Part II - Paper 10.

§ 7.3 Self-mitigated interruption

The main idea behind the mechanism which we call *self-mitigated interruption* (see Figure 7.2) is to let the existing supervisory attentional system [177] play the main part in deciding whether a notification is to interrupt the current primary task or not, instead of primarily let the system rely on sensing and interpreting the context and then rather brutally call for attention.

In *self-mitigated interruption* we assume that a system knows about a given message's importance and urgency and can use variable intensity for a stimuli to reduce the number of unwanted interruptions based on an understanding of how a person's sensitivity to the stimuli changes as a function of cognitive load. Using the terminology of the interruption stage management model [103], the low intensity stimuli does not exceed the *detection threshold*. Notably, however, this detection threshold is different for different levels of cognitive load: higher cognitive load means a higher detection threshold.

There are many ways in which interruption management based on self-mitigated interruption could be implemented in a wearable personal device such as a WPA. We investigated using the thermohaptic modality for perceptually and cognitively graceful integration of notifications originating from digital services with ongoing tasks that the human agent is performing.

Of course, the self-mitigated interruption approach is modality agnostic.

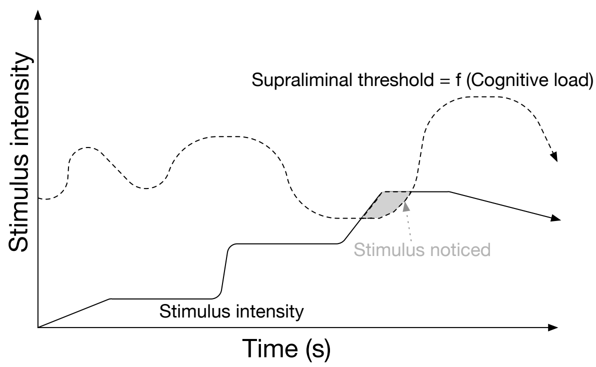


Figure 7.2: *Principle of self-mitigated interruption: The intensity of the stimulus (represented by the solid line) and the supraliminal threshold (represented by the dashed line), which is dependent on level of cognitive load, determines whether the stimulus gets consciously noticed. By increasing the stimulus intensity level, the likelihood of the stimulus being noticed increases [27].*

Our interest in the thermal modality is motivated by the fact that it is a modality rarely explicitly used in everyday activities and thus a potentially very useful information channel for interruption by not intruding directly into modalities that might already “be in use” [208].

As we explained in Paper 10, we conducted the first experiment to investigate the feasibility of using thermohaptic stimuli for sel-mitigated interruption. We built a wearable thermohaptic device that is able to successfully display thermal stimulus with different intensity for both cooling and heating. To manipulate the intensity of the stimuli, the speed and amount of energy can be adjusted. We also repeated the experiment for a condition where the user was asked to play a computer game in parallel with responding to the thermal stimuli as a high cognitive load condition. However, due to the high number of independent variables (2 (conditions: low/high cognitive load) $\times$ 2 (modes: cooling/heating) $\times$ 3 (heating rates) $\times$ 3 (temperature changes).) and few participants (15), we could not detect statistically significant difference between two conditions. If we would have selected only two levels of intensity (high/low) for the thermal stimuli based on our experiment, we could have seen statistically significant effect of cognitive load on sensitivity of participants to the thermal stimuli since it is more likely to find statistically significant difference with 2 $\times$ 2 experiment compared to the above-mentioned experiment design. Also the cognitive load level could have been adjusted to low and high using a standard cognitive load task such as n-back task.

§ 7.4 Discussion

7.4.1 The WPA concept

Any interruption management mechanism that is implemented in a mobile or wearable system will be effective only to the degree it is aware of, and can control, the sources of interruptions affecting the person carrying/wearing the device. Just like today's mobile phones are used as central portal to sources of information on the Internet to provide various services, we envision Wearable Personal Assistants (WPAs) to act as a central bottleneck through which the majority of digital services need to push any potentially interrupting notifications. In fact, WPAs should be able to offer certain explicit control to their users with respect to which digital services that should be allowed to make use of what notification modality and under which circumstance. Using the terminology of our body-and-mind centric model of WPAs, the *self-mitigated interruption* targets the unconscious cognitive processing in the human brain. When the WPA displays a stimuli with a certain intensity, body sensors send a signal to unconscious cognition processing (path 3-4). If the brain is cognitively loaded a weak stimuli is filtered out by the unconscious cognition and not sent to conscious cognition. But if the intensity of the stimuli is higher than the sensitivity threshold due to high intensity of the stimuli or low cognitive load the stimuli is sent to the conscious cognition and the user consciously perceives the notification. We can represent the paths when a stimuli hits the user's attention or gets filtered out by the unconscious cognitive processing as follows where threshold = f (cognitive load):

$$Path = \begin{cases} 3-4-6 & \text{if } intensity > threshold \\ 3-4 & \text{if } intensity < threshold \end{cases} \tag{7.1}$$

7.4.2 The WPA for clinicians

Unwanted interruptions by mobile phones, pagers, and other communication devices are well-recognized problems in healthcare work domain [17]. On the one hand due to the distribution of medical devices and infrastructures in the hospital, clinicians need to move between departments, offices, and wards all the time. On the other hand, healthcare work is highly collaborative. Therefore, clinicians need to use mobile phones, pagers and other communication devices to find each other in the hospital. But these devices are highly interruptive. In my ethnography at hospital, the mobile phone of a surgeon called when he was treating an emergency trauma case and he needed to answer the call and return to his medical task again. Previous work in hospitals have used context-aware systems (e.g [18, 54]) to reduce interruptions. But even in the context-aware systems, the system shares the context of one party with another to allow the interruptee person decides about the interruption which means the interrupted person has no choice to get interrupted or not. The *self-mitigated interruption* can be a complementary approach which takes the cognitive state of the interrupted person into consideration.

- Chapter 8 -

Conclusions & Future Work

§ 8.1 Conclusions

The idea of using wearable digital assistants to support mobile users on the move has long been discussed in the HCI community. However, the vision of using wearable assistants in everyday life has not come true yet due to several technical limitations (e.g. low battery life and limited computational resources) and interaction related challenges of using computing devices in parallel with real world tasks (e.g. interruptions). Recent advances in miniaturizing electronics and hardware has resulted in solving some of the technical challenges of building relatively unobtrusive wearable computers (e.g. Google Glass). This was the main driver for my first research questions.

RQ1: *Given the unobtrusive form factor of state of the art smart glasses such as Google Glass and possibility of hands-free interaction, how much closer can the modern smart glasses bring the visionary wearable assistants to reality?*

To answer this question, I designed and implemented a wearable personal assistant (WPA) for orthopedic surgeons in a hospital. The hospital setting was selected due to the high mobility of clinicians (specially surgeons) that motivates the use of a WPA to have easier access to the medical information in different situations. The WPA was evaluated through a clinical simulation study in a hospital where real orthopedic surgeons used the WPA in three different scenarios: 1) touch-less interaction in operation theatre, 2) tele-presence, and 3) mobile access to the electronic patient records (EPR) during ward rounds. Apart from some technical limitations of Google Glass (addressed also in previous studies [28, 209, 204]) such as low battery life, overheating, and low resolution of the HMD specially for medical diagnosis, the results of my user study revealed the great potential of the WPA to support remote collaboration and touch-less interaction in surgical settings. However, using the WPA in the ward round scenario raised some challenges in human to human interactions between the surgeons and the patients. This shows that assuming the above-mentioned technical challenges are solved in the future, the modern smart glasses such as Google Glass can be extremely useful in particular scenarios (e.g. touch-less interaction and tele-presence in surgery room) where social interaction is not crucial.

Designing the WPA for surgeons was challenging since there was no stable definition and defined funtionalities for wearable assistants. Also due to the novelty of wearable computers for ordinary people, involving users in the initial steps of the design process was not useful. This raised my second research question.

RQ2: *How can we design a wearable personal assistant that accommodates for interaction on the move and parallel interaction with real-world tasks?*

Due to the above-mentioned reasons in order to design the WPA for surgeons, I went through a concept-driven interaction design research where I defined the concept of *wearable personal assistant (WPA)* informed by previous definitions of wearable assistants and inspired by the *Egocentric Interaction* paradigm that introduces a human-centric approach to model interaction. Previous efforts in wearable computing have mainly aimed at bringing computers physically closer to the human body while in my research, I have focused on bringing the human mind aspect into the design space by introducing and deploying a human-body-and-mind-centric approach to the design of WPAs (Figure 2.3, page 28). According to our human-body-and-mind centric approach, three types of assistance can be provided by the envisioned WPAs: 1) perception, 2) cognition, and 3) action assistance. I developed a conceptual design framework (Figure 3.2, page 39) that includes both human (body and mind) aspects and characteristics of the healthcare work domain into the design space. I used the framework to design and implement the WPA prototype for orthopedic surgeons on the Google Glass platform.

According to my findings from evaluation of the WPA in the hospital and previous work in the pervasive healthcare area [17], unwanted interruption is one of the main challenges of using mobile computing devices in hospitals. The problems raises from unwanted interruptions by computing devices are well-recognized (e.g. increasing stress, accidents, frustration, and time pressure are some). The interruption problems can be even more annoying in WPAs since they are envisioned to be with users all the time and easily draw users' attention to provide them useful information proactively. This mean not only does the number of interruptions increase for the WPA users, but also the interruptions are harder to ignore. Therefore, my third research question was about interruption management for WPAs.

RQ3: *How can we establish a symbiotic interplay between the existing interruption management infrastructure in our brain and wearable personal assistant, approaching graceful interruption management?*

The dominant approach to mitigate the interruption problems is using context to predict interruptibility of the users; however, the context recognition approach is proven to be really hard. As an alternative approach, we introduced the *self-mitigated interruption* concept that is essentially a symbiosis of artificial external stimuli tuned to existing human attention management mechanisms. In *self-mitigated interruption* assuming that the system knows about a given message's importance and urgency, it can use variable intensity for a stimuli to reduce the number of unwanted interruptions based on an understanding of how a person's sensitivity to the stimuli changes as a function of cognitive load. We performed a pilot study laying the ground for using a wrist-worn thermohaptic actuator for self-mitigating interruption.

My observations in the hospital showed that interaction with computers is a big challenge for clinicians in situations where the clinicians cannot touch computing devices due to the sterility restrictions. While modern smart glasses such as Google Glass accept voice commands and head movements as two touch-less input modalities, my clinical simulation study showed that there is a need for more touch-less input modalities (e.g. the tele-presence scenario 3.6.2). My fourth research question was about other touch-less input modalities such as body gestures and eye movements.

RQ4: *How can body gestures and eye movements be used as touch-less input modalities for interaction with wearable personal assistants? More specifically, what are the technical and usability challenges of touch-less interaction through body gestures and eye movements?*

1. **Exploring hand & foot gestures:** To investigate the utility of using hand and foot gestures for interaction with medical images in the surgery room, we designed and implemented a system that helps orthopedic surgeons interact with 2D X-ray and 3D CT scans displayed on two different displays through hand and foot gestures [84]. We used a single wearable sensor including accelerometer and gyroscope to detect 6 different hand gestures. To address the clutch mechanism problem (the difficulty for the system to know whether the user intends to address the system or if the actions are directed towards the real world and should be ignored), we investigated the utility of foot gestures for starting and finishing the interaction, and switching the interaction mode from one system to another one. To detect the foot gestures we used capacitive sensors embedded into the floor. Our system is able to detect 12 hand and foot gestures with an acceptable accuracy for interaction with medical images. Hand-gesture-based interaction leaves hands of the user free for performing real-world tasks. This makes hand-gesture-based interaction a viable input modality for WPAs when the users' hands are dirty, needs to be kept sterile (e.g. during a surgery), or the user needs to hold a tool in hands; however,

the user's hands are still involved temporarily in the interaction. Foot-based interaction can be a good complement for hand gestures when the user is standing still and not walking (e.g. during a surgery). Due to the slow speed of the discrete gesture-based commands, gesture-based input is not appropriate for providing a lot of commands in a short period of time (e.g. adjusting the position of the X-ray on the screen during orthopedic surgeries). The result of our clinical study on the WPA [82] showed that the head movements can be useful for providing continuous input.

2. **Eye-based interaction with WPAs:** Eye-based interaction with eyewear computers based on embedded eye trackers could in theory be an intuitive unobtrusive method for communication between the user and WPA. However, despite the great potentials of using our eyes for interaction, eye-based interaction techniques are still not widely used due to the obtrusiveness of existing eye trackers (hardware and software), calibration problems, inaccuracy of the eye trackers, and challenges involved in using a perceptual organ, the eye, as an input modality. To address the hardware and software problem of eye tacking in eyewear computers, we made an unobtrusive camera-based eye-tracker for Google Glass (GlassGaze) that offloads images processing and calculating eye coordination to our eye tracker on the Android platform (EyeDroid). To address the usability problem of explicit interaction with eyes, we proposed a MAGIC pointing approach for eyewear computers. Our MAGIC pointing uses gaze as an implicit input in a pointing task where the departure point is set implicitly based on the gaze coordination, and the head movements is used for fine-grained adjusting the pointer. Our study showed that MAGIC pointing is faster than head pointing for far and small targets. Finally, to address the calibration problem, inaccuracy, and usability challenge of using gaze as explicit input, we developed a calibration-free eye-movement-based interaction technique called EyeGrip. EyeGrip analyses optokinetic nystagmus natural eye movements to detect object of interest among a series of unidirectional moving objects. Our experiment showed that EyeGrip can accurately (87%) detect the image that draws the user's attention among a series of scrolling images on the screen. The utility of EyeGrip was investigated as a method for achieving seamless interaction with scrolling contents on eyewear computers. We studied two different applications on eyewear computers : 1) a menu scroll viewer and 2) a Facebook newsfeed. The results of our study shows that the EyeGrip technique performs as good as a more common manual selection method (keyboard). Moreover, the accuracy of the EyeGrip method for menu item selection was higher while in the Facebook study participants found the keyboard more

accurate.

Even if technological advances in developing unobtrusive and high performance Head-Mounted Displays (HMDs) have made them the currently most important output device for eyewear computers, and even if this has opened new horizons for applications of wearable computers in everyday life, fundamental challenges still remain. Eye fatigue, small field of view, swimming effects, limited resolution, and multiple focus planes, are some of the most famous problems associated with HMDs [102, 162]. My last research question investigates an alternative visual output method for eyewear computers.

RQ5: *Given the known user challenges associated with Head-Mounted Displays (HMDs), is motor-controlled laser pointer technology a viable alternative?*

Wearable laser pointer: To answer this question, I developed a wearable motor-controlled laser pointer and quantitatively determined if wearable laser pointers are viable alternatives to HMDs for indicating where in the physical environment the user should direct her/his attention. The potential benefit of the laser pointer would be reduced eye fatigue due to the fact that the refocusing challenges associated with HMDs would be completely eliminated. 10 participants were asked to perform a short tele-guided pick-and drop task using both approaches. The statistical analysis indicates that user performance in the laser pointer condition is higher than the HMD approach (P = .064, $\alpha = 0.1$). While all 10 participants found the task easy in both conditions, 8 of 10 participants found the laser pointer system more convenient.

§ 8.2 Future Work

1. **Develop next version of the WPA for clinicians:** One future direction for further research is to design and implement the second version of the WPA for clinicians informed by findings from the empirical study, and equipped with novel unobtrusive interaction techniques such as EyeGrip and MAGIC pointing. In that case, the GlassGaze application together with the EyeDriod gaze tracker on an Android device could facilitate the system development.

2. **Provide subliminal cues to the user in a 3D environment using the wearable laser pointer:** Another future study is to modify the wearable laser pointer to provide subliminal visual stimuli. One way of providing such subliminal stimuli is to display the laser beam during such a short time interval that the stimuli remains below the visual threshold that grabs conscious attention from the given individual in the given context. This system can be used as an apparatus

for further experiments to investigate the effect of subliminal cueing in 3D environment on human performance in real-world tasks as a component in wearable persuasive systems.

3. **Self-mitigated interruption:** Our initial experiment on using a thermohaptic wristband for graceful interruptions can be used for further experiments on the self-mitigated interruption concept where only two levels of intensity (high/low) are selected for the thermal stimuli based on our experiment. Also the cognitive load level could be adjusted to low and high using a standard cognitive load task such as the n-back task. Furthermore, the self-mitigated interruption concept can be explored for other modalities such as audio and vision.

4. **Investigate the feasibility of using an EOG sensing approach for EyeGrip:** The first implementation of our EyeGrip interaction technique uses a camera-based eye tracking approach which can be cumbersome due to the complexities and limitations of camera-based eye tracking. However, EyeGrip does not need any gaze data, and other unobtrusive technologies for tracking eye movements (e.g. electrooculography (EOG)-based methods) can be used for implementing EyeGrip. Another future work is to investigate other eye tracking methods for implementing EyeGrip.

Concluding remarks: I believe that while the technological advances in building unobtrusive eyewear computers will eventually make WPAs popular in some application areas such as healthcare, widespread general-purpose use (such as in the case of smartphones) is not likely to happen anytime soon. However, with the Wearable Personal Assistant concept, and the body-and-mind centric design approach lined out in this thesis, I believe to have pointed out some open research avenues which in a longer time perspective could lead to wearable computers that are useful in a wider range of everyday applications.

Bibliography

[1] Data glasses controlled by eye movements, online: http://www.fraunhofer.de, retrived sep, 2014. 79

[2] Piotr D Adamczyk and Brian P Bailey. If not now, when?: The effects of interruption at different moments within task execution. In *Proc. of the CHI '4*, pages 271–278. ACM, 2004. 4, 32, 83

[3] Kurt Adamer, David Bannach, Tobias Klug, Paul Lukowicz, Marco Luca Sbodio, Mimi Tresman, Andreas Zinnen, and Thomas Ziegert. Developing a wearable assistant for hospital ward rounds: An experience report. In *The Internet of Things*, pages 289–307. Springer, 2008. 34, 40, 70

[4] Kamran Ahmed, Muhammed Jawad, May Abboudi, Andrea Gavazzi, Ara Darzi, Thanos Athanasiou, Justin Vale, Mohammad Shamim Khan, and Prokar Dasgupta. Effectiveness of procedural simulation in urology: a systematic review. *The Journal of urology*, 186(1):26–34, 2011. 49

[5] Antti Ajanki, Mark Billinghurst, Hannes Gamper, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, et al. An augmented reality interface to contextual information. *Virtual reality*, 15(2-3):161–173, 2011. 74

[6] Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. Putting your best foot forward: Investigating real-world mappings for foot-based gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1229–1238, New York, NY, USA, 2012. ACM. 69

[7] Frank Althoff, Rudi Lindl, Leonhard Walchshausl, and S Hoch. Robust multimodal hand-and head gesture recognition for controlling automotive infotainment systems. *VDI BERICHTE*, 1919:187, 2005. 70

[8] Gabor Aranyi, Sid Kouider, Alan Lindsay, Hielke Prins, Imtiaj Ahmed, Giulio Jacucci, Paolo Negri, Luciano Gamberini, David Pizzi, and Marc Cavazza. Subliminal cueing of selection behavior in a virtual environment. *Presence: Teleoperators and Virtual Environments*, 23(1):33–50, 2014. 31, 32, 87

[9] Daniel Lee Ashbrook. Enabling mobile microinteractions. 2010. 23, 32, 52, 86

[10] Thomas Augsten, Konstantin Kaefer, René Meusel, Caroline Fetzer, Dorian Kanitz, Thomas Stoff, Torsten Becker, Christian Holz, and Patrick Baudisch. Multitoe: High-precision interaction with back-projected floors based on high-resolution multi-touch input. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 209–218, New York, NY, USA, 2010. ACM. 67, 69

[11] Chris Baber. Wearable computers: a human factors review. *International Journal of Human-Computer Interaction*, 13(2):123–145, 2001. 56

[12] Brian P Bailey, Joseph A Konstan, and John V Carlis. Measuring the effects of interruptions on task performance in the user interface. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 2, pages 757–762. IEEE, 2000. 40

[13] Brian P Bailey, Joseph A Konstan, and John V Carlis. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Proceedings of INTERACT*, volume 1, pages 593–601, 2001. 40

[14] Gilles Bailly, Jörg Müller, Michael Rohs, Daniel Wigdor, and Sven Kratz. Shoesense: A new perspective on gestural interaction and wearable applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1239–1248, New York, NY, USA, 2012. ACM. 67, 68

[15] Saskia Bakker. Design for peripheral interaction, phd thesis, 2013. 4, 32

[16] Jakob E. Bardram. Activity-based computing: Support for mobility and collaboration in ubiquitous computing. *Personal Ubiquitous Comput.*, 9(5):312–322, September 2005. 21

[17] Jakob E. Bardram and Claus Bossen. Mobility work: The spatial dimension of collaboration at a hospital. *Computer Supported Cooperative Work (CSCW)*, 14(2):131–160, 2005. 7, 33, 39, 40, 42, 44, 71, 89, 92

[18] Jakob E. Bardram and Thomas R. Hansen. The aware architecture: Supporting context-mediated social awareness in mobile cooperation. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 192–201, New York, NY, USA, 2004. ACM. 32, 44, 85, 89

[19] Richard Bates and Howell O Istance. Why are eye mice unpopular? a detailed comparison of head and eye controlled assistive technology pointing devices. *UAIS*, 2(3):280–290, 2003. 70, 75, 78, 79

[20] Moniruzzaman Bhuiyan and Rich Picking. Gesture-controlled user interfaces, what have we done and what's next. In *Proceedings of the Fifth Collaborative Research Symposium on Security, E-Learning, Internet and Networking (SEIN 2009), Darmstadt, Germany*, pages 25–29, 2009. 65, 68

[21] Ali Bigdelou, Loren Schwarz, and Nassir Navab. An adaptive solution for intra-operative gesture-based human-machine interaction. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, pages 75–84. ACM, 2012. 67, 68

[22] Mark Billinghurst and Thad Starner. Wearable devices: new ways to manage information. *Computer*, 32(1):57–64, 1999. 55

[23] WO Blackwood, TR Anderson, CT Bennett, JR Corson, MR Endsley, PA Hancock, J Hochberg, JE Hoffman, and RV Kruk. Tactical displays for soldiers: Human factors considerations. *Panel on Human Factors in the Design of Tactical Display Systems for the Individual Soldier*, 1997. 56

[24] Giancarlo Bo, Andrea Lorenzon, Nicolas Chevassus, and Valerie Blondel. Wearable computing and mobile workers: The aeronautic maintenance showcase in the wearit@work project. In *Applied Wearable Computing (IFAWC), 2006 3rd International Forum on*, pages 1–12, March 2006. 1

[25] Deborah A Boehm-Davis and Roger Remington. Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accident Analysis & Prevention*, 41(5):1124–1129, 2009. 83

[26] Richard A. Bolt. "put-that-there": Voice and gesture at the graphics interface. *SIGGRAPH Comput. Graph.*, 14(3):262–270, July 1980. 65, 67

[27] Frank Bolton, Shahram Jalaliniya, and Thomas Pederson. A wrist-worn thermohaptic device for graceful interruption. *Interac-*

tion Design & Architecture Journal, 26(Special Issue on Peripheral Interaction):39–54, 2015. 85, 88

[28] A Brent, O Lasun, K Phani, et al. Emerging technology in surgical education: Combining real-time augmented reality and wearable computing devices. *Orthopedics*, 37(11):751–757, 2014. 91

[29] Andreas Bulling and Hans Gellersen. Toward mobile eye-based human-computer interaction. *Pervasive Computing, IEEE*, 9(4):8–12, 2010. 73, 74, 75

[30] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. *Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments*. ETH, Eidgenössische Technische Hochschule Zürich, Wearable Computing Laboratory, 2009. 74

[31] Andreas Bulling, Daniel Roggen, and Gerhard Troster. What's in the eyes for context-awareness? *IEEE Pervasive Computing*, (2):48–57, 2010. 73

[32] Andreas Bulling, Jamie.A. Ward, Hans Gellersen, and Gerhard Troster. Eye movement analysis for activity recognition using electrooculography. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(4):741–753, April 2011. 73

[33] Lynell Burmark. Visual presentations that prompt, flash & transform here are some great ways to have more visually interesting class sessions. *Media and methods*, 40:4–5, 2004. 80

[34] Daniel Chen and Roel Vertegaal. Using mental load for managing interruptions in physiologically attentive user interfaces. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1513–1516, New York, NY, USA, 2004. ACM. 85

[35] Jingyuan Cheng, David Bannach, Kurt Adamer, Thomas Bernreiter, and Paul Lukowicz. A wearable, conductive textile based user interface for hospital ward rounds document access. In *Smart Sensing and Context*, pages 182–191. Springer, 2008. 40

[36] YL Chi, SK Ong, ML Yuan, and AYC Nee. Wearable interface for the physical disabled. In *Proceedings of the 1st international convention on Rehabilitation engineering & assistive technology: in conjunction with 1st Tan Tock Seng Hospital Neurorehabilitation Meeting*, pages 28–32. ACM, 2007. 70

[37] Adrian F. Clark. What do we want from a wearable user interface. In *Proceedings of Workshop on Software Engineering for Wearable and Pervasive Computing*, page 3, 2000. 3, 4, 8, 21

[38] Manfred E Clynes. Cyborgs and space. *Astronautics*, 26:74–75, 1960. 22

[39] Andrea Colaço, Ahmed Kirmani, Hye Soo Yang, Nan-Wei Gong, Chris Schmandt, and Vivek K. Goyal. Mime: Compact, low power 3d gesture sensing for interaction with head mounted displays. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 227–236, New York, NY, USA, 2013. ACM. 32, 65, 67, 68, 69

[40] Andrew Crossan, Stephen Brewster, and Alexander Ng. Foot tapping for mobile interaction. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*, BCS '10, pages 418–422, Swinton, UK, UK, 2010. British Computer Society. 69

[41] Mary Czerwinski, Edward Cutrell, and Eric Horvitz. Instant messaging and interruption: Influence of task type on performance. In *OZCHI 2000 conference proceedings*, volume 356, page 361, 2000. 40

[42] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. A diary study of task switching and interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 175–182, New York, NY, USA, 2004. ACM. 20

[43] Robert Danforth, Andrew Duchowski, Robert Geist, and Elizabeth McAliley. A platform for gaze-contingent virtual environments. In *Smart Graphics (Papers from the 2000 AAAI Spring Symposium, Technical Report SS-00-04)*, pages 66–70, 2000. 74

[44] Yuri Danilov and Mitchell Tyler. Brainport: an alternative input to the brain. *Journal of integrative neuroscience*, 4(04):537–550, 2005. 24

[45] Richard W. DeVaul, Alex "Sandy" Pentland, and Vicka R. Corey. The memory glasses: Subliminal vs. overt memory support with imperfect information. In *Proceedings of ISWC'03*, ISWC '03, pages 146–146, Washington, DC, USA, 2003. IEEE Computer Society. 8

[46] Paul Dourish, W. Keith Edwards, Anthony LaMarca, and Michael Salisbury. Presto: An experimental architecture for fluid interactive document spaces. *ACM Trans. Comput.-Hum. Interact.*, 6(2):133–161, June 1999. 20

[47] Heiko Drewes and Albrecht Schmidt. Interacting with the computer using gaze gestures. In *Proceedings of the 11th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II*, INTERACT'07, pages 475–488, 2007. 75

[48] Mikael Drugge, Josef Hallberg, Peter Parnes, and Kåre Synnes. Wearable systems in nursing home care: prototyping experience. *Pervasive Computing, IEEE*, 5(1):86–91, Jan 2006. 34

[49] Mikael Drugge1, Marcus Nilsson, Urban Liljedahl, Kåre Synnes, and Peter Parnes. Methods for interrupting a wearable computer user. In *Wearable Computers, 2004. ISWC 2004. Eighth International Symposium on*, volume 1, pages 150–157, Oct 2004. 86

[50] Andrew Duchowski. *Eye tracking methodology: Theory and practice*, volume 373. Springer Science & Business Media, 2007. 74

[51] Lars C Ebert, Gary Hatch, Garyfalia Ampanozi, Michael J Thali, and Steffen Ross. You can't touch this touch-free navigation through radiological images. *Surgical innovation*, 19(3):301–307, 2012. 67

[52] D Gareth Evans, Roger Drew, and Paul Blenkhorn. Controlling mouse pointer position using an infrared head-operated joystick. *Rehabilitation Engineering, IEEE Transactions on*, 8(1):107–117, 2000. 70

[53] David Gareth Evans, Stephen Pettitt, and Paul L Blenkhorn. A head operated “joystick”. In *Proceedings of the 5th International Conference on Computers Helping People with Special Needs. Part I*, ICCHP '96, pages 85–91, Munich, Germany, Germany, 1996. R. Oldenbourg Verlag GmbH. 70

[54] Jesus Favela, Ana I Martinez-Garcia, et al. Context-aware mobile communication in hospitals. *Computer*, (9):38–46, 2003. 89

[55] Jesus Favela, Marcela Rodriguez, Alfredo Preciado, and Victor M. Gonzalez. Integrating context-aware public displays into a mobile hospital information system. *Information Technology in Biomedicine, IEEE Transactions on*, 8(3):279–286, Sept 2004. 39

[56] Onur Ferhat, Fernando Vilarino, and FJ Sánchez. A cheap portable eye-tracker solution for common setups. *Journal of Eye Movement Research*, 7(3):2, 2014. 77

[57] Robert Fisher and Reid Simmons. Smartphone interruptibility using density-weighted uncertainty sampling with reinforcement learning. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, volume 1, pages 436–441, Dec 2011. 4, 86

[58] Eric Freeman and David Gelernter. Lifestreams: A storage model for personal data. *SIGMOD Rec.*, 25(1):80–86, March 1996. 20

[59] E.T. Gendlin. *Experiencing and the Creation of Meaning: A Philosophical and Psychological Approach to the Subjective*. Northwestern University studies in phenomenology & existential philosophy. Northwestern University Press, 1962. 30

[60] Chauncey Graetzel, T Fong, Sebastien Grange, and Charles Baur. A non-contact mouse for surgeon-computer interaction. *Technology and Health Care*, 12(LSRO2-ARTICLE-2004-001), 2004. 67

[61] Sukeshini A Grandhi and Quentin Jones. Conceptualizing interpersonal interruption management: A theoretical framework and research program. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on System Science*, pages 1–10. IEEE, 2009. 32, 86

[62] JO Gray, Pei Jia, Huosheng H Hu, Tao Lu, and Kui Yuan. Head gesture recognition for hands-free control of an intelligent wheelchair. *Industrial Robot: An International Journal*, 34(1):60–68, 2007. 70

[63] Tobias Grundgeiger and Penelope Sanderson. Interruptions in healthcare: theoretical views. *International journal of medical informatics*, 78(5):293–307, 2009. 83

[64] Teng Han, Jason Alexander, Abhijit Karnik, Pourang Irani, and Sriram Subramanian. Kick: Investigating the use of kick gestures for mobile interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 29–32, New York, NY, USA, 2011. ACM. 69

[65] Holger Harms, Oliver Amft, Rene Winkler, Johannes Schumm, Kusserow Martin, and Gerhard Troester. Ethos: Miniature orientation sensor for wearable human motion analysis. In *Sensors, 2010 IEEE*, pages 1037–1042, Nov 2010. 10

[66] Chris Harrison, Hrvoje Benko, and Andrew D Wilson. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 441–450. ACM, 2011. 57, 58, 67

[67] Doris Hausen. Peripheral interaction - exploring the design space, phd thesis, 2014. 4, 32

[68] Joyce Ho and Stephen S. Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 909–918, New York, NY, USA, 2005. ACM. 4, 84, 86

[69] Nick Hunn. The market for smart wearable technology a consumer centric approach, 2015. 1

[70] Aulikki Hyrskykari, Päivi Majaranta, and Kari-Jouko Räihä. From gaze control to attentive interfaces. In *Proceedings of HCII*, volume 2, 2005. 75

[71] Dynastream Innovations. Ant message protocol and usage, rev 5.1, 2014. 10

[72] Toshiki Iso and Kenichi Yamazaki. Gait analyzer based on a cell phone with a single three-axis accelerometer. In *Proceedings of the 8th Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '06, pages 141–144, New York, NY, USA, 2006. ACM. 69

[73] Richard J. Adams, Aaron B. Olowin, Blake Hannaford, and O. Scott Sands. Tactile data entry for extravehicular activity. In *World Haptics Conference (WHC), 2011 IEEE*, pages 305–310, June 2011. 70

[74] Julie A Jacko. *Visual Displays in Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications.* CRC press, 2012. 56

[75] Robert J. K. Jacob. What you look at is what you get: Eye movement-based interaction techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 11–18, New York, NY, USA, 1990. ACM. 73, 75

[76] Shahram Jalaliniya and Diako Mardanbegi. Eyegrip: Detecting targets in a series of uni-directional moving objects using optokinetic nystagmus eye movements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5801–5811, New York, NY, USA, 2016. ACM. 30

[77] Shahram Jalaliniya and Diako Mardanbegi. Seamless interaction with scrolling contents on eyewear computers using optokinetic nystagmus eye movements. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*, ETRA '16, pages 295–298, New York, NY, USA, 2016. ACM. 81

[78] Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. Magic pointing for eyewear computers. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ISWC '15, pages 155–158, New York, NY, USA, 2015. ACM. 70, 75, 77, 79

[79] Shahram Jalaliniya, Diako Mardanbegi, Thomas Pederson, and Dan Witzner. Head and eye movement as pointing modalities for eyewear computers. In *Wearable and Implantable Body Sensor Networks Workshops (BSN Workshops), 2014 11th International Conference on*, pages 50–53, June 2014. 70, 75

[80] Shahram Jalaliniya, Diako Mardanbegi, Ioannis Sintos, and Daniel Garcia Garcia. Eyedroid: An open source mobile gaze tracker on android for eyewear computers. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, pages 873–879, New York, NY, USA, 2015. ACM. 10

[81] Shahram Jalaliniya and Thmoas Pederson. Designing wearable personal assistants for surgeons: An egocentric approach. *Pervasive Computing, IEEE*, 14(3):22–31, July 2015. 27, 37, 39, 43, 48, 85

[82] Shahram Jalaliniya and Thomas Pederson. *Qualitative Study of Surgeons Using a Wearable Personal Assistant in Surgeries and Ward Rounds*, pages 208–219. Springer International Publishing, Cham, 2017. 49, 50, 51, 52, 53, 54, 94

[83] Shahram Jalaliniya, Thomas Pederson, and Steven Houben. Wearable laser pointer versus head-mounted display for tele-guidance applications? In *Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program*, pages 193–200. ACM, 2014. 59, 61, 62

[84] Shahram Jalaliniya, Jeremiah Smith, Miguel Sousa, Lars Büthe, and Thomas Pederson. Touch-less interaction with medical images using hand & foot gestures. In *Adjunct Proceedings of UBICOMP'13*, pages 1265–1274. ACM, 2013. 71, 93

[85] Monique WM Jaspers. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *International journal of medical informatics*, 78(5):340–353, 2009. 49

[86] Brett Jones, Rajinder Sodhi, David Forsyth, Brian Bailey, and Giuliano Maciocci. Around device interaction for multiscale navigation. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 83–92. ACM, 2012. 65

[87] Holger Junker, Oliver Amft, Paul Lukowicz, and Gerhard Tröster. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition*, 41(6):2010–2024, 2008. 68

[88] Daniel Kahneman. *Attention and effort.* Citeseer, 1973. 40

[89] Ashish Kapoor and Rosalind W Picard. A real-time head nod and shake detector. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–5. ACM, 2001. 70

[90] Victor Kaptelinin. Creating computer-based work environments: An empirical study of macintosh users. In *Proceedings of the 1996 ACM SIGCPR/SIGMIS Conference on Computer Personnel Research*, SIGCPR '96, pages 360–366, New York, NY, USA, 1996. ACM. 20

[91] David R Karger, Karun Bakshi, David Huynh, Dennis Quan, and Vineet Sinha. Haystack: A customizable general-purpose information management tool for end users of semistructured data. In *Proc. of the CIDR Conf*, 2005. 20

[92] Toshikazu Karitsuka and Kosuke Sato. A wearable mixed reality with an on-board projector. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 321. IEEE Computer Society, 2003. 57

[93] Nicky Kern and Bernt Schiele. Context-aware notification for wearable computing. In *null*, page 223. IEEE, 2003. 85

[94] Nicky Kern, Bernt Schiele, and Albrecht Schmidt. Multi-sensor activity context detection for wearable computing. In *Ambient Intelligence*, pages 220–232. Springer, 2003. 86

[95] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 167–176, New York, NY, USA, 2012. ACM. 65, 67, 68

[96] HyungJun Kim and DaeHyun Ryu. Computer control by tracking head movements for the disabled. In Klaus Miesenberger, Joachim Klaus, WolfgangL. Zagler, and ArthurI. Karshmer, editors, *Computers Helping People with Special Needs*, volume 4061 of *Lecture Notes in Computer Science*, pages 709–715. Springer Berlin Heidelberg, 2006. 70

[97] Yoon Sang Kim, Byung Seok Soh, and Sang-Goog Lee. A new wearable input device: Scurry. *Industrial Electronics, IEEE Transactions on*, 52(6):1490–1499, Dec 2005. 67, 68

[98] David B Koons, Carlton J Sparrell, and Kristinn Rr Thorisson. Integrating simultaneous input from speech, gaze, and hand gestures. *MIT Press: Menlo Park, CA*, pages 257–276, 1993. 67

[99] Andreas Krause, Asim Smailagic, and Daniel P. Siewiorek. Context-aware mobile computing: learning context- dependent personal preferences from a wearable sensor array. *Mobile Computing, IEEE Transactions on*, 5(2):113–127, Feb 2006. 85

[100] Manu Kumar, Andreas Paepcke, and Terry Winograd. Eyepoint: practical pointing and selection using gaze and keyboard. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430. ACM, 2007. 78

[101] Andrew Toshiaki Kurauchi, Carlos Hitoshi Morimoto, Diako Mardanbegi, and Dan Witzner Hansen. Towards wearable gaze supported augmented cognition. 2013. 79

[102] Robert S Laramee and Colin Ware. Rivalry and interference with a head-mounted display. *ACM Transactions on Computer-Human Interaction*, 9(3):238–251, 2002. 5, 56, 95

[103] Kara A. Latorella. Investigating interruptions: implications for flight-deck performance [phd thesis]. *Buffalo, NY: State University of New York. Acessado em*, 10. 87

[104] Christl Lauterbach, Axel Steinhage, and Axel Techmer. Large-area wireless sensor system based on smart textiles. In *International Multi-Conference on Systems, Sygnals & Devices*, 2012. 72

[105] Jae-Young Lee, Hyung-Min Park, Seok-Han Lee, Soon-Ho Shin, Tae-Eun Kim, and Jong-Soo Choi. Design and implementation of an augmented reality system using gaze interaction. *Multimedia Tools and Applications*, 68(2):265–280, 2014. 74

[106] Dongheng Li, Jason Babcock, and Derrick J. Parkhurst. openeyes: A low-cost head-mounted eye-tracking solution. In *Proceedings of the 2006 Symposium on Eye Tracking Research &Amp; Applications*, ETRA '06, pages 95–100, New York, NY, USA, 2006. ACM. 77

[107] Joseph Carl Robnett Licklider. Man-computer symbiosis. *Human Factors in Electronics, IRE Transactions on*, (1):4–11, 1960. 22

[108] Andrew R Lingley, Muhammad Ali, Y Liao, R Mirjalili, M Klonner, M Sopanen, S Suihkonen, T Shen, BP Otis, H Lipsanen, et al. A single-pixel wireless contact lens display. *Journal of Micromechanics and Microengineering*, 21(12):125014, 2011. 59

[109] David Liu, Simon A. Jenkins, and Penelope M. Sanderson. Clinical implementation of a head-mounted display of patient vital signs. In *Wearable Computers, 2009. ISWC '09. International Symposium on*, pages 47–54, Sept 2009. 34

[110] Kristian Lukander, Sharman Jagadeesan, Huageng Chi, and Kiti Müller. Omg!: A new robust, wearable and affordable open source mobile gaze tracker. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, pages 408–411, New York, NY, USA, 2013. ACM. 77

[111] Faul Lukowicz, Andreas Timm-Giel, Michael Lawo, and Otthein Herzog. Wearit@ work: Toward real-world industrial wearable computing. *Pervasive Computing, IEEE*, 6(4):8–13, 2007. 1

[112] Paul Lukowicz, Tnde Kirstein, and Gerhard Troster. Wearable systems for health care applications. *Methods of Information in Medicine-Methodik der Information in der Medizin*, 43(3):232–238, 2004. 1, 3, 24

[113] Paul Lukowicz, Andreas Poxrucker, Jens Weppner, B. Bischke, Jochen Kuhn, and Michael Hirth. Glass-physics: Using google glass to support high school physics experiments. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ISWC '15, pages 151–154, New York, NY, USA, 2015. ACM. 1

[114] Pattie Maes et al. Agents that reduce work and information overload. *Communications of the ACM*, 37(7):30–40, 1994. 21

[115] Pattie Maes and Robyn Kozierok. Learning interface agents. In *AAAI*, volume 93, pages 459–465, 1993. 21

[116] Paul P Maglio, Teenie Matlock, Christopher S Campbell, Shumin Zhai, and Barton A Smith. Gaze and speech in attentive user interfaces. In *Advances in Multimodal Interfaces—ICMI 2000*, pages 1–7. Springer, 2000. 75

[117] Päivi Majaranta and Kari-Jouko Räihä. Twenty years of eye typing: Systems and design issues. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, ETRA '02, pages 15–22, New York, NY, USA, 2002. ACM. 75

[118] Steve Mann. Introduction: On the bandwagon or beyond wearable computing? *Personal Technologies*, 1(4):203–207, 1997. 1, 4, 8, 21, 24

[119] Steve Mann. Wearable computing: A first step toward personal imaging. *Computer*, 30(2):25–32, 1997. 55

[120] Steve Mann. Continuous lifelong capture of personal experience with eyetap. In *Proceedings of the the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, CARPE'04, pages 1–21, New York, NY, USA, 2004. ACM. 31

[121] Diako Mardanbegi. Haytham gaze tracker. `http://itu.dk/research/eye/`, June 2014. 10, 76, 77, 78

[122] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. Eye-based head gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 139–146, 2012. 70, 75

[123] Diako Mardanbegi and Shahram Jalaliniya. Glassgaze. `https://github.com/dmardanbeigi/GlassGaze`, 2014. [Online; accessed 30-Dec-2015]. 10, 76

[124] Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work: more speed and stress. In *Proc. of the CHI '8*, pages 107–110. ACM, 2008. 83

[125] Joe Marshall and Paul Tennent. Mobile interaction does not exist. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 2069–2078, New York, NY, USA, 2013. ACM. 3, 23, 34

[126] Addison Mayberry, Pan Hu, Benjamin Marlin, Christopher Salthouse, and Deepak Ganesan. ishadow: Design of a wearable, real-time mobile gaze tracker. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '14, pages 82–94, New York, NY, USA, 2014. ACM. 77

[127] Daniel McFarlane. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Hum.-Comput. Interact.*, 17(1):63–139, March 2002. 86

[128] Daniel C McFarlane and Steven M Wilder. Interactive dirt: increasing mobile work performance with a wearable projector-camera system. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 205–214. ACM, 2009. 58

[129] James E Melzer. *Head-mounted displays: designing for the user.* McGraw-Hill Professional, 1997. 56

[130] Martin Missfeldt. How google glass works, 2015. 47

[131] Pranav Mistry, Pattie Maes, and Liyan Chang. Wuw-wear ur world: a wearable gestural interface. In *CHI'09 extended abstracts on Human factors in computing systems*, pages 4111–4116. ACM, 2009. 57, 58, 67

[132] Thomas P Moran and Shumin Zhai. Beyond the desktop metaphor in seven dimensions. *Beyond the Desktop Metaphor—Designing Integrated Digital Work Environments*, pages 335–354, 2007. 4, 20

[133] Louis-Philippe Morency and Trevor Darrell. Head gesture recognition in intelligent interfaces: The role of context in improving recognition. In *Proceedings of the 11th International Conference on Intelligent User Interfaces*, IUI '06, pages 32–38, New York, NY, USA, 2006. ACM. 70

[134] Desmond Morris. *Bodytalk: the meaning of human gestures*. Crown Publishers, 1994. 69

[135] Michael J Muller and Sarah Kuhn. Participatory design. *Communications of the ACM*, 36(6):24–28, 1993. 4

[136] Emli-Mari Nel, DJC MacKay, Piotr Zieliński, Ollie Williams, and Roberto Cipolla. Opengazer: open-source gaze tracker for ordinary webcams. 77

[137] Susanna Nilsson, Torbjorn Gustafsson, and Per Carleberg. Hands free interaction with virtual information in a real environment: Eye gaze as an interaction tool in an augmented reality system. *PsychNology Journal*, 7(2):175–196, 2009. 74

[138] Donald A Norman and Stephen W Draper. User centered system design. *New Perspectives on Human-Computer Interaction, L. Erlbaum Associates Inc., Hillsdale, NJ*, 1986. 25

[139] Scott D. Novich and David M. Eagleman. [d79] a vibrotactile sensory substitution device for the deaf and profoundly hearing impaired. In *Haptics Symposium (HAPTICS), 2014 IEEE*, pages 1–1, Feb 2014. 24

[140] Scott D Novich and David M Eagleman. Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. *Experimental brain research*, 233(10):2777–2788, 2015. 24

[141] Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, and Tom Carrell. Touchless interaction in surgery. *Commun. ACM*, 57(1):70–77, January 2014. 67

[142] Toshikazu Ohshima, Hiroyuki Yamamoto, and Hideyuki Tamura. Gaze-directed adaptive rendering for interacting with virtual space. In *Proceedings of the 1996 Virtual Reality Annual International Symposium (VRAIS 96)*, VRAIS '96, pages 103–, Washington, DC, USA, 1996. IEEE Computer Society. 74

[143] Volker Paelke, Christian Reimann, and Dirk Stichling. Kick-up menus. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1552–1552, New York, NY, USA, 2004. ACM. 69

[144] Toni Pakkanen and Roope Raisamo. Appropriateness of foot interaction for non-accurate spatial tasks. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1123–1126. ACM, 2004. 69

[145] Matthew K.X.J. Pan, Joanna McGrenere, Elizabeth A. Croft, and Karon E. MacLean. Exploring the role of haptic feedback in enabling implicit hci-based bookmarking. *Haptics, IEEE Transactions on*, 7(1):24–36, Jan 2014. 85

[146] Joseph Paradiso, Craig Abler, Kai-yuh Hsiao, and Matthew Reynolds. The magic carpet: Physical sensing for immersive environments. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '97, pages 277–278, New York, NY, USA, 1997. ACM. 67, 69

[147] Joseph A Paradiso, Kai Yuh Hsiao, and Ari Benbasat. Interfacing to the foot: apparatus and applications. In *CHI'00 extended abstracts on Human factors in computing systems*, pages 175–176. ACM, 2000. 69

[148] Glenn Pearson and Mark Weiser. Exploratory evaluations of two versions of a foot-operated cursor-positioning device in a target-selection task. *ACM SIGCHI Bulletin*, 19(3):70–75, 1988. 69

[149] Thomas Pederson, Lars-Erik Janlert, and Dipak Surie. Towards a model for egocentric interaction with physical and virtual objects. In *Proceedings of NordiCHI'10*, pages 755–758. ACM, 2010. 24, 25

[150] Thomas Pederson, Lars-Erik Janlert, and Dipak Surie. A situative space model for mobile mixed-reality computing. *IEEE pervasive computing*, 10(4):73–83, 2011. 7, 8, 26, 84

[151] Veljko Pejovic and Mirco Musolesi. Interruptme: Designing intelligent prompting mechanisms for pervasive applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pages 897–908, New York, NY, USA, 2014. ACM. 4, 86

[152] John Kangchun Perng, Brian Fisher, Seth Hollar, and Kristofer SJ Pister. Acceleration sensing glove. In *iswc*, page 178. IEEE, 1999. 68

[153] Bastian Pfleging, Niels Henze, Albrecht Schmidt, Dominique Rau, and Bastian Reitschuster. Influence of subliminal cueing on visual search tasks. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 1269–1274. ACM, 2013. 87

[154] Catherine Plaisant and Ben Shneiderman. Organization overviews and role management: inspiration for future desktop environments. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 1995., Proceedings of the Fourth Workshop on*, pages 14–22, Apr 1995. 20

[155] Michael I Posner, Mary J Nissen, and Raymond M Klein. Visual dominance: an information-processing account of its origins and significance. *Psychological review*, 83(2):157, 1976. 55

[156] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. *Computer Graphics Forum*, 17(3):41–52, 1998. 25

[157] Ville Rantanen, Toni Vanhala, Outi Tuisku, Pekka-Henrik Niemenlehto, Jarmo Verho, Veikko Surakka, Martti Juhola, and Jukka Lekkala. A wearable, wireless gaze tracker with integrated selection command source for human x2010;computer interaction. *Information Technology in Biomedicine, IEEE Transactions on*, 15(5):795–801, Sept 2011. 77

[158] Madhu C. Reddy, Paul Gorman, and Jakob Bardram. Special issue on supporting collaboration in healthcare settings: The role of informatics. *International Journal of Medical Informatics*, 80(8):541–543, 2011. 40

[159] Jun Rekimoto. Gesturewrist and gesturepad: unobtrusive wearable interaction devices. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, pages 21–27, 2001. 32, 65, 67, 68

[160] Andreas Riener, Pierre Chalfoun, and Claude Frasson. The potential of subliminal information displays to change driver behavior. *Presence*, 23(1):51–70, 2014. 87

[161] George Robertson, Greg Smith, Brian Meyers, Patrick Baudisch, Mary Czerwinski, Eric Horvitz, Daniel Robbins, and Desney Tan. Explorations in task management on the desktop. *Beyond the Desktop Metaphor, ed. Kaptelinin V. and Czerwinski, M*, pages 101–138, 2007. 21

[162] Jannick P Rolland and Henry Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3):287–309, 2000. 5, 56, 95

[163] A.F. Rovers and H.A. van Essen. Guidelines for haptic interpersonal communication applications: an exploration of foot interaction styles. *Virtual Reality*, 9(2-3):177–191, 2006. 69

[164] Ravikrishna Ruddarraju, Antonio Haro, Kris Nagel, Quan T Tran, Irfan A Essa, Gregory Abowd, and Elizabeth D Mynatt. Perceptual user interfaces using vision-based eye tracking. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 227–233. ACM, 2003. 75

[165] Guilherme Cesar Soares Ruppert, Leonardo Oliveira Reis, Paulo Henrique Junqueira Amorim, Thiago Franco de Moraes, and Jorge Vicente Lopes da Silva. Touchless gesture user interface for interactive image visualization in urological surgery. *World journal of urology*, 30(5):687–691, 2012. 67

[166] Alireza Sahami Shirazi, Niels Henze, Tilman Dingler, Martin Pielot, Dominik Weber, and Albrecht Schmidt. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 3055–3064, New York, NY, USA, 2014. ACM. 4, 86

[167] Nobuchika Sakata, Takeshi Kurata, and Hideaki Kuzuoka. Visual assist with a laser pointer and wearable display for remote collaboration. 2006. 59

[168] Nadine Sarter. Multimodal support for interruption management: Models, empirical findings, and design recommendations. *Proceedings of the IEEE*, 101(9):2105–2112, 2013. 84

[169] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 483–492. ACM, 2012. 67

[170] Albrecht Schmidt. Implicit human computer interaction through context. *Personal Technologies*, 4(2-3):191–199, 2000. 29

[171] Albrecht Schmidt. Interactive context-aware systems interacting with ambient intelligence. *Ambient intelligence*, 159, 2005. 74

[172] Albrecht Schmidt, H-W Gellersen, and Christian Merz. Enabling implicit human computer interaction: a wearable rfid-tag reader. In

Wearable Computers, The Fourth International Symposium on, pages 193–194. IEEE, 2000. 32, 87

[173] Albrecht Schmidt, Hans W Gellersen, and Michael Beigl. A wearable context-awareness component. In *iswc*, page 176. IEEE, 1999. 32, 87

[174] Johannes Schöning, Florian Daiber, Antonio Krüger, and Michael Rohs. Using hands and feet to navigate and manipulate spatial data. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 4663–4668, New York, NY, USA, 2009. ACM. 69

[175] Björn Schwerdtfeger, Daniel Pustka, Andreas Hofhauser, and Gudrun Klinker. Using laser projectors for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 134–137. ACM, 2008. 59

[176] Jeremy Scott, David Dearman, Koji Yatani, and Khai N. Truong. Sensing foot gestures from the pocket. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 199–208, New York, NY, USA, 2010. ACM. 69

[177] Tim Shallice and Paul Burgess. Supervisory control of action and thought selection. In *Attention: Selection, Awareness, and Control : a Tribute to Donald Broadbent*, pages 171–187. Oxford University Press, 1993. 87

[178] Jeffrey S Shell, Roel Vertegaal, and Alexander W Skaburskis. Eyepliances: attention-seeking devices that respond to visual attention. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 770–771. ACM, 2003. 75

[179] Takaaki Shiratori, Hyun Soo Park, Leonid Sigal, Yaser Sheikh, and Jessica K. Hodgins. Motion capture from body-mounted cameras. In *ACM SIGGRAPH 2011 Papers*, SIGGRAPH '11, pages 31:1–31:10, New York, NY, USA, 2011. ACM. 67

[180] Linda E. Sibert and Robert J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, pages 281–288, New York, NY, USA, 2000. ACM. 75

[181] Daniel Siewiorek, Asim Smailagic, Junichi Furukawa, Andreas Krause, Neema Moraveji, Kathryn Reiger, Jeremy Shaffer, and Fei Lung Wong. Sensay: A context-aware mobile phone. In *null*, page 248. IEEE, 2003. 84

[182] Mikael B. Skov and Rune Th. Hoegh. Supporting information access in a hospital ward by a context-aware mobile electronic patient record. *Personal Ubiquitous Comput.*, 10(4):205–214, March 2006. 39

[183] Asim Smailagic and Daniel P. Siewiorek. The cmu mobile computers: a new generation of computer systems. In *Compcon Spring '94, Digest of Papers.*, pages 467–473, Feb 1994. 1

[184] David Canfield Smith, Charles Irby, Ralph Kimball, and Eric Harslem. The star user interface: An overview. In *Proceedings of the June 7-10, 1982, National Computer Conference*, AFIPS '82, pages 515–528, New York, NY, USA, 1982. ACM. 19, 20

[185] Jeremiah Smith and Naranker Dulay. Ringlearn: Long-term mitigation of disruptive smartphone interruptions. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014 IEEE International Conference on*, pages 27–35. IEEE, 2014. 32, 86

[186] Jeremiah Smith, Anna Lavygina, Jiefei Ma, Alessandra Russo, and Naranker Dulay. Learning to recognise disruptive smartphone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 121–124. ACM, 2014. 4, 86

[187] Stefan Soutschek, Jochen Penne, Joachim Hornegger, and Johannes Kornhuber. 3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–6. IEEE, 2008. 67

[188] Oleg Špakov, Poika Isokoski, and Päivi Majaranta. Look and lean: accurate head-assisted eye pointing. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pages 35–42. ACM, 2014. 70

[189] Vince Stanford. Wearable computing goes live in industry. *Pervasive Computing, IEEE*, 1(4):14–19, 2002. 1

[190] Thad Starner. The challenges of wearable computing: Part 1 & 2. *IEEE Micro*, 21(4):44–52, 2001. 1, 22, 23, 24, 28

[191] Thad Starner. Project glass: An extension of the self. *Pervasive Computing, IEEE*, 12(2):14–16, April 2013. 22, 34

[192] Thad Starner, Bernt Schiele, and Alex Pentland. Visual contextual awareness in wearable computing. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*, pages 50–57, Oct 1998. 85

[193] Thad Eugene Starner. *Wearable computing and contextual awareness.* PhD thesis, Citeseer, 1999. 4, 8, 21

[194] Thomas Stiefmeier, Daniel Roggen, Georg Ogris, Paul Lukowicz, and Gerhard Tröster. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing*, (2):42–50, 2008. 1

[195] Erik Stolterman and Mikael Wiberg. Concept-driven interaction design research. *Human–Computer Interaction*, 25(2):95–118, 2010. 6

[196] Rory Stuart. *Design of virtual environments.* Barricade Books, Incorporated, 2001. 56

[197] Ewelina Światek-najwer, Romuald Bedziński, Paweł Krowicki, Krzysztof Krysztoforski, Peter Keppler, and Josef Kozak. Improving surgical precision–application of navigation system in orthopedic surgery. *Acta of Bioengineering and Biomechanics*, 10(4):55–62, 2008. 36

[198] Vildan Tanriverdi and Robert JK Jacob. Interacting with eye movements in virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 265–272. ACM, 2000. 74

[199] Robert J Teather and I Scott MacKenzie. Effects of user distraction due to secondary calling and texting tasks. In *Proc. of the International Conference on Multimedia and Human Computer Interaction*, pages 115.1–115.8. International ASET, Inc, 2013. 83

[200] James N. Templeman, Patricia S. Denbrook, and Linda E. Sibert. Virtual locomotion: Walking in place through virtual environments. *Presence*, 8(6):598–617, Dec 1999. 69

[201] Bruce H Thomas, Gerald Quirchmayr, and Wayne Piekarski. Through-walls communication for medical emergency services. *International Journal of Human-Computer Interaction*, 16(3):477–496, 2003. 34, 45

[202] Liam D Turner, Stuart M Allen, and Roger M Whitaker. Interruptibility prediction for ubiquitous systems: conventions and new directions from a growing field. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 801–812. ACM, 2015. 32, 86

[203] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 439–448. ACM, 2013. 75

[204] Wolfgang Vorraber, Siegfried Voessner, Gerhard Stark, Dietmar Neubacher, Steven DeMello, and Aaron Bair. Medical applications of near-eye display devices: An exploratory study. *International Journal of Surgery*, 12(12):1266 – 1272, 2014. 91

[205] Juan Wachs, Helman Stern, Yael Edan, Michael Gillam, Craig Feied, Mark Smith, and Jon Handler. Gestix: a doctor-computer sterile gesture interface for dynamic environments. In *Soft Computing in Industrial Applications*, pages 30–39. Springer, 2007. 67

[206] Mark Weiser and John Seely Brown. Designing calm technology. *POWERGRID JOURNAL*, 1, 1996. 30

[207] Johanna I Westbrook, Amanda Woods, Marilyn I Rob, William Dunsmuir, and Richard O Day. Association of interruptions with an increased risk and severity of medication administration errors. *Archives of Internal medicine*, 170(8):683, 2010. 83

[208] Christopher D. Wickens. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2):159–177, 2002. 88

[209] Antoine Widmer, Roger Schaer, Dimitrios Markonis, and Henning Muller. Facilitating medical information search using google glass connected to a content-based medical image retrieval system. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 4507–4510, Aug 2014. 91

[210] P.S. Windyga and D.M. Wink. A wearable computing based system for the prevention of medical errors committed by registered nurses in the intensive care unit. In *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, volume 3, pages 1940–1941 vol.3, Oct 2002. 1, 34

[211] Hendrik Witt. User interfaces for wearable computers. 2007. 56, 84, 86

[212] Hendrik Witt and Mikael Drugge. Hotwire: An apparatus for simulating primary tasks in wearable computing. In *CHI'06 Extended Abstracts on Human Factors in Computing Systems*, pages 1535–1540. ACM, 2006. 86

[213] Katrin Wolf, Anja Naumann, Michael Rohs, and Jörg Müller. A taxonomy of microinteractions: Defining microgestures based on ergonomic and scenario-dependent requirements. In *Human-Computer Interaction–INTERACT 2011*, pages 559–575. Springer, 2011. 86

[214] Jiahui Wu, Gang Pan, Daqing Zhang, Guande Qi, and Shijian Li. Gesture recognition with a 3-d accelerometer. In *Ubiquitous intelligence and computing*, pages 25–38. Springer, 2009. 67, 68

[215] Tetsuya Yamamoto, Masahiko Tsukamoto, and Tomoki Yoshihisa. Foot-step input method for operating information devices while jogging. In *Applications and the Internet, 2008. SAINT 2008. International Symposium on*, pages 173–176. IEEE, 2008. 69

[216] Shumin Zhai, Carlos Morimoto, and Steven Ihde. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 246–253, New York, NY, USA, 1999. ACM. 70, 75, 79

[217] Thomas G. Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. A hand gesture interface device. In *ACM SIGCHI Bulletin*, volume 18, pages 189–192. ACM, 1987. 65, 68

[218] Thomas G. Zimmerman, Joshua R. Smith, Joseph A. Paradiso, David Allport, and Neil Gershenfeld. Applying electric field sensing to human-computer interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, pages 280–287, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co. 67, 69

[219] Andreas Zinnen, Bernt Schiele, and Thomas Ziegert. Browsing patient records during ward rounds with a body worn gyroscope. In *Wearable Computers, 2007 11th IEEE International Symposium on*, pages 107–108. IEEE, 2007. 67, 68

Part II

PAPERS

Paper 1: Designing WPAs for Surgeons: An Egocentric Approach

Title of Paper

Designing Wearable Personal Assistants for Surgeons: An Egocentric Approach

Authors:

Shahram Jalaliniya, Thomas Pederson

Published:

IEEE Pervasive Computing, vol.14, no. 3, pp. 22-31, July-Sept. 2015

Abstract

The design of general-purpose wearable computers demands particular care for how human perception, cognition, and action work and work together. The authors propose a human body-and-mind centric (egocentric as opposed to device-centric) design framework and present initial findings from deploying it in the design of a wearable personal assistant (WPA) for orthopedic surgeons. The result is a Google Glass-based prototype system aimed at facilitating touchless interaction with x-ray images, browsing of electronic patient records (EPR) when on the move, and synchronized ad hoc remote collaboration. This article is part of a special issue on digitally enhanced reality.

Designing Wearable Personal Assistants for Surgeons:
An Egocentric Approach

A wearable personal assistant prototype based on Google Glass facilitates touchless interaction with x-ray images, allows clinicians to browse electronic patient records (EPRs) while on the move, and supports synchronized ad hoc remote collaboration.

Shahram Jalaliniya and Thomas Pederson
IT University of Copenhagen

Increasingly powerful wearable computers suggest that a tight integration between human and computer is achievable. However, to reach a deep integration that offers timely cognitive support, we need to better understand how humans perceive, think, and act in the world. An ideal wearable intelligent assistant "augments memory, intellect, creativity, communication, and physical senses and abilities."[1] We prefer the term "wearable *personal* assistant" (WPA) to emphasize the tight integration between a single mind, body, and computer.

Human body-and-mind-centric design approaches can complement existing technology-driven efforts (that is, efforts based on available state-of-the-art hardware) in addressing many of the challenges of human-computer systems because, ultimately, the power of these systems depends on the level of integration.[1] Although we can adapt and modify the artificial cognitive architecture (the "computer" system), we cannot change the body and brain of human agents. As system designers, we can only ensure that our WPAs talk to our relatively static biological setup in the way evolution designed it to interpret and act in the world. Our focus here is on human perception, cognition, and action.

How we think we interpret the world around us in everyday life is not how we (our brains) actually do it. In the last few decades, research in cognitive science, perception psychology, and neuropsychology has resulted in some remarkable findings, some of which are still debated:

- About 95 percent of measurable brain activity is unconscious.[2]
- The 5 percent of human conscious cognitive processing (attention) is volatile and easily interrupted by internal unconscious processes or external stimuli.
- Human attention does not multitask.[3]
- By the time external stimuli grasps our attention (if it does), it has already undergone significant filtering and transformation by unconscious processes.[4]
- Human routine actions are often initiated and controlled by unconscious cognitive processes triggered by direct external stimuli, leaving conscious processes out of the loop.

As developers of interactive systems, we should care about these findings, because the

Published by the IEEE CS ■ 1536-1268/15/$31.00 © 2015 IEEE

systems we aim to design are closer to the user's body and mind than classical personal computing devices have ever been. Inspired by these findings, we propose an information flow model that considers the perception, cognition, and action of human agents while (unlike in more classical HCI models) dealing with conscious and unconscious cognition separately. This lets us take a more holistic view of the role of WPAs, making it evident that the explicit interaction between WPAs and users occurs in an information-rich context, in which our brains process much more than we traditionally model as system designers. It also lets us start speculating about functionalities that WPAs could offer that interface directly with the unconscious part of our cognitive processes, something that is undoubtedly still hard to implement in practice, even if successful attempts have indeed been made.[5]

Although this article's main focus is our egocentric approach to the design of WPAs inspired by modern cognitive science, we also discuss our experiences deploying the framework in the hospital domain and our initial WPA prototype for orthopedic surgeons based on the Google Glass platform.

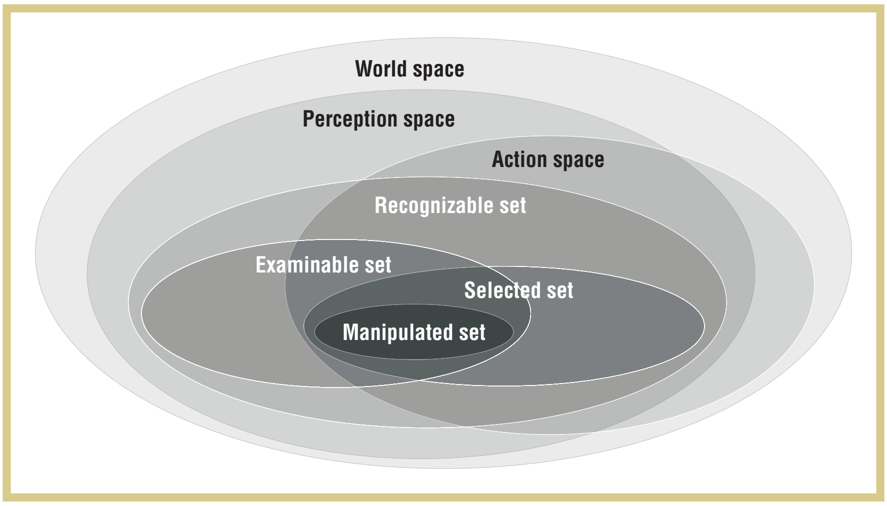


Figure 1. The situative space model (SSM).[9] We developed the SSM to capture what a specific human agent can perceive and not perceive, reach and not reach, at any given moment in time. In particular, the perception space is the space around the agent that can be perceived at each moment. The action space is the space around the agent that is currently accessible to the agent's physical actions.

Egocentric Interaction

Both system designers and users increasingly face a new HCI paradigm that redefines the relationship between the human, computer system, and world: an *egocentric interaction paradigm*.[6] This paradigm extends and modifies the classical user-centered approach in HCI[7] on several points:

- *Situatedness* acknowledges the primacy of the agents' current bodily situation at each point in time in guiding and constraining agents' behavior.
- *Attention to the complete local environment* emphasizes the need to consider the entire environment, not just a single targeted artifact or system.
- The *proximity principle* assumes that proximity plays a fundamental role in determining what can be done, what events signify, and what agents are up to.
- *Changeability* of the environment and of the agents' relationship with the environment takes into account agents' more or less constant body movements, including the head, hands, and sensing organs.
- The *physical-virtual equity principle* pays equal attention to interaction with both virtual (digital) objects (classical HCI) and physical objects (classical ergonomics).

The term "egocentric" signals that it is the body and mind of a specific individual that (literally) acts as the center of reference, so all modeling is anchored to this individual's body and mind in this interaction paradigm. The term is analogously used in psychology and virtual reality to denote the conceptual and spatial frames of reference that humans by necessity rely on when thinking and acting in the world and when collaborating with others.[8]

Action and Perception Instead of Input and Output

In the egocentric interaction paradigm, the modeled individual must be viewed as an agent that can move about in a mixed-reality environment (an environment consisting of both directly accessible everyday "real" entities and virtual/digital objects accessed through mediating digital devices), not as a user performing a dialogue with a computer. Adopting the physical-virtual equity principle, we suggest substituting the concepts of (device) input and output with (human agent) action and perception.

The New "Desktop"

To facilitate the design of egocentric interaction systems such as the WPAs we focus on here, we developed a situative space model (SSM) to capture what a specific human agent can perceive and not perceive, reach and not reach, at any given moment in time (see Figure 1). This model is for the emerging egocentric interaction paradigm what the virtual desktop is for the PC/WIMP (window, icon, menu, pointing device) interaction paradigm: more or less everything of interest to a

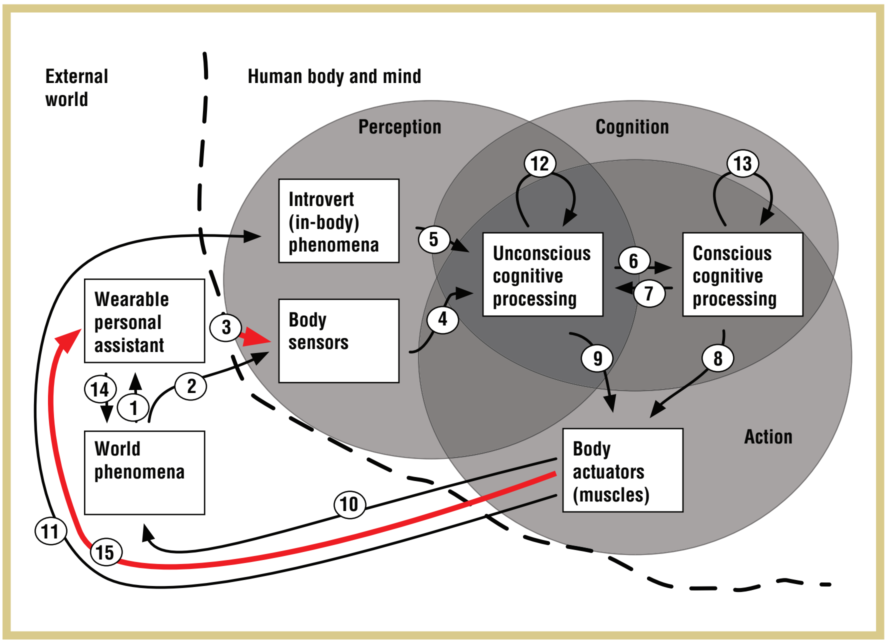


Figure 2. A body-and-mind-centric model of how wearable personal assistants (WPAs) fit into the flow of perception, cognition, and action of human agents. Classical HCI models are typically concerned only with pathway 15-3 (the bold red arrows) and how it relates to conscious cognitive processing. Yet by and large, our perception of the world (pathway 2-4) and our perception of our body state (arrow 5) is beyond our conscious control.

specific human agent is assumed to, and supposed to, happen here. We describe only the perception and action spaces here and point the reader elsewhere for more details.[9]

The *perception space* is the space around the agent that can be perceived at each moment. Like all spaces and sets of the SSM, it is agent-centered, varying continuously with the agent's movements. Different senses have differently shaped perception spaces, with different operating requirements, range, and spatial and directional resolution with regard to the perceived sources of the sense data. Compare vision and hearing, for example.

The *action space* is the space around the agent that is currently accessible to the agent's physical actions. Objects within this space can be directly acted upon. The outer range limit is less dependent on object type than that of the perception space and is basically determined by the agent's physical reach.

Perception-Cognition-Action Loop

Figure 2 shows a simplified model of information flows occurring as the result of a human agent acting in the world. The purpose is not to provide a completely accurate account but a good-enough model for designing future interactive systems.

Perception

By and large, our perception of the world (pathway 2-4 in Figure 2) and our perception of our body state (arrow 5) is beyond our conscious control. However, conscious cognitive processes influence unconscious processes (arrow 7), as in the case when we deliberately address our attention to a certain speaker in a crowd and automatically (thanks to unconscious processing), to some degree, single out the voice we want to hear. We can also consciously and indirectly affect unconscious processing by orienting our body sensors (such as vision) toward phenomena of interest (pathway 8-10-2-4).

Cognition

Human cognition is divided into unconscious and conscious processing (arrows 12 and 13 in Figure 2, respectively). The human agent receives input from sensors capturing in-body phenomena (such as proprioceptive information about limb positions and for maintaining homeostasis) and from sensors capturing information from the external world. No external world or in-body phenomena is subject to conscious cognitive processing before it has been unconsciously processed (pathways 2-4-6 and 5-6, respectively).

Conscious processing is slower than unconscious processing. For instance, muscular reactions to immediate threats are initiated unconsciously (pathway 4-9) long before conscious processes are engaged. We protect our faces with our hands instinctively from approaching projectiles such as hockey pucks even when we are consciously aware of the fully protective shields of transparent material in front of us.

Action

Human action is initiated and controlled by a mix of conscious and unconscious cognitive processes. An example of an activity mostly driven by an unconscious perception-cognition-action loop is walking along a well-known road with no exposure to obstacles (pathways 2-4-9-10 and 11). An example of an activity that uses a combination of conscious and unconscious cognition is attempting to thread a needle, which demands focused visual and tactile conscious attention (pathway 2-4-6-8-10) in parallel with unconscious detailed control of hand, finger, and arm muscles (pathway 5-9-11).

The Wearable Personal Assistant in the Loop

By including unconscious cognitive processing and all perceivable world phenomena (including everyday objects such as coffee cups and footballs), the model in Figure 2 provides a more complete perspective of the context in which a WPA operates than classical HCI models, which are typically concerned only with pathway 15-3 (red arrows in the figure) and how it relates to conscious cognitive processing. It becomes evident that any information generated by a WPA (arrow 3) is just one source of information among many others that hit the unconscious and conscious parts of our brains, which together try to make sense of it all. Heuristics for serving that information in a timely manner, arriving at successful "attention management," is probably best based on knowing what else is hitting the senses in parallel. The context-aware systems community has investigated this for years but often using a system- or device-centered approach. We believe that a human-centric approach toward determining what matters in a given situation (for example, using the SSM in Figure 1) will reveal interesting complementary information.

Note that conceptually, the hardware user interface of a WPA (such as the head-mounted display [HMD], microphone, loudspeaker, and Google Glass touchpad) receives input from the human agent and provides output that the human agent can sense in the shape of world phenomena (arrows 10 and 2). In Figure 2, these information flows are re-represented as separate information flows (arrows 15 and 3) to facilitate the following discussion.

Implicit Input and Output

Although seldom clearly defined in the HCI literature, the distinction between explicit and implicit input and output is useful for discussing some important properties of WPAs[10]:

- *Explicit input* is action intentionally and consciously directed toward the WPA. For example, a human agent navigates the GUI presented on the Google Glass HMD by swiping the touch area (pathway 8-15).
- *Implicit input* is action performed by a human agent without the conscious intention of communicating with the WPA. For example, a human agent acts in the real world (moves about, manipulates objects, and interacts with other human agents), which is partially sensed by the WPA (pathways 8-10-1 and 9-10-1).
- *Explicit output* occurs when a WPA addresses the conscious mind. The WPA creates a change in the agent's perception space (Figure 1) that the human agent cannot avoid consciously perceiving (pathway 3-4-6), thereby inviting the human agent to act (arrows 10, 11, and 15).
- *Implicit output* occurs when a WPA addresses the unconscious mind. The WPA generates a phenomena in the agent's perception space (Figure 1) that reaches the unconscious part of cognition (pathway 3-4) but not the conscious part (pathway 6)—for example, through ambient displays such as the "dangling string"[11]).

By placing actuators and sensors on or very close to the body and keeping them there for large parts of the day, we would argue that the WPA can potentially sense and affect several of the information flows shown in Figure 2 with more precision than more traditional interactive systems (such as PCs and smartphones). This leads to the intriguing idea of future WPAs being able to facilitate the transition from "felt sense" tacit knowledge, generated as the human agent experiences the world, to knowledge that the agent can consciously reflect on and articulate,[12] augmenting human cognition at the core. Space limitations, and our wish to discuss possibilities that are more directly applicable in the near future, make us end this section by mentioning some more concrete potential mechanisms for using implicit input and output in the context of WPAs (all currently explored by the pervasive computing community but not necessarily in the context of designing WPAs).

Situation identification is one such mechanism. By implicitly monitoring body state through pathway 10-1 (for example, body posture, galvanic skin response, and heartbeat), and correlating it with the state of the nearby world (arrow 1), the WPA has a reasonable platform for determining the human agent's current situation.

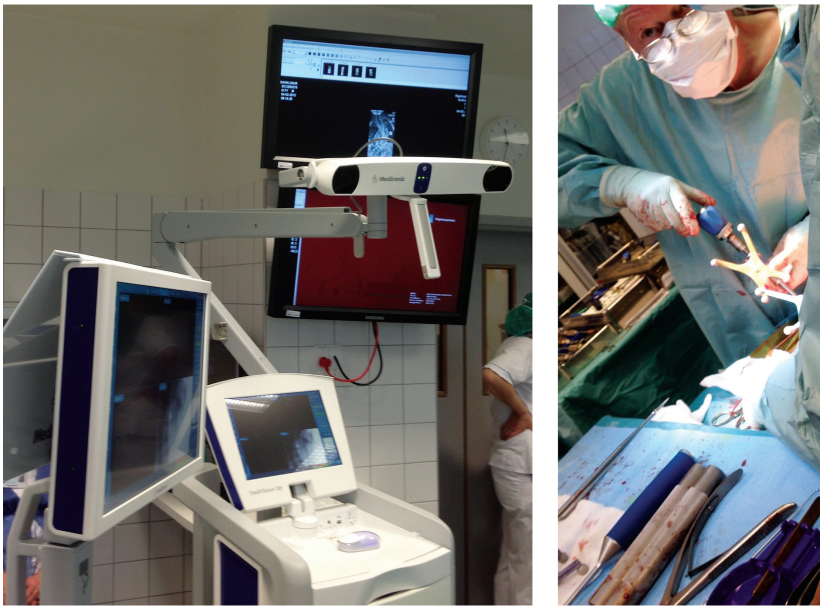
Another mechanism is *subliminal cueing*. Certain phenomena measured best close to the body (for example, eye movements, facial expressions, and electromyography [EMG]) can provide important insights into ongoing conscious and unconscious cognitive processing. They can therefore help determine the intensity level and type of stimuli that could be used for subliminal cueing,[13] and for the WPA to subliminally direct the human agent's gaze in a certain direction (pathway 3-4-9).

Finally, work has also been done in *mediated reality*. If the WPA is sufficiently integrated into the visual perception flow, beyond Google Glass see-through and partially covering monocular HMDs and toward Steve Mann's EyeTap vision,[14] the WPA could act both as a filter and highlighter. In this way, it could directly alter the perception of the surrounding world (arrow 2) so as to facilitate tasks. Naturally, security and ethics become important topics if development leads us in this direction.

A WPA for Orthopedic Surgeons

Healthcare is a highly collaborative work domain, and even single-user hospital systems, such as electronic patient record (EPR) systems, are used for coordination and collaboration. Moreover, because of the spatial distribution of departments, wards, meeting rooms, and offices, clinicians move around a lot. Their mobile yet collaborative work style forces them to use pagers, mobile phones, and other devices to communicate, which can interrupt and interfere with ongoing activities.[15]

However, previous attempts to develop WPAs for clinicians have faced several



(a) (b)

Figure 3. Design framework for a WPA prototype for surgeon assistance: (a) the navigation system and several large screens to monitor medical images in the operating room, and (b) surgeons look at the screens while operating on the patient.

technical and human-related challenges.[16] Emerging unobtrusive eyewear computers such as Google Glass raise hope for solving some of the technical issues. Furthermore, we believe that our egocentric design approach demystifies some of the human-related complexities by defining the concept of a WPA based on human needs for different kinds of assistance. We are currently developing a WPA for orthopedic surgeons to support them throughout their workday based on our egocentric design approach.

To understand the healthcare work domain, we conducted an ethnographic study in Rigshospital in Copenhagen. As part of the ethnographic study, we shadowed an on-call orthopedic surgeon during a workday. Our initial observations showed that surgeons have among the highest mobility in the hospital. Moreover, we observed several orthopedic surgeons in different types of orthopedic surgeries.

A typical surgeon's workday starts with a daily or weekly meeting with other surgeons. During these meetings, surgeons discuss general topics such as administrative issues and special patient cases. After regular meetings, all surgeons meet in radiology conference rooms to review patients' latest x-rays, MRIs, and CT scans; the medical images are presented on several large screens. During the radiology meetings, surgeons discuss important cases and make notes in special notebooks. Afterward, surgeons booked for different types of orthopedic surgeries go to the operating rooms (or theaters). Operating rooms are prepared based on the type of surgery planned because each type requires particular medical infrastructures. For example, some complex orthopedic surgeries require a navigation system to monitor a 3D model of the surgical site and the position of the operation instruments. Several screens in the operating room allow surgeons to monitor patient records, x-rays, MRIs, and CT scans (Figure 3a).

Because computers and their peripherals are difficult to sterilize during a surgery, an assistant or nurse operates the mouse and keyboard for surgeons. We observed a complex orthopedic surgery (scoliosis surgery) in which the surgical team used the navigation system to increase accuracy of the operation. In surgeries using the navigation system, a CT scan of the patient is displayed on a large screen, while the navigation system tracks the positions of surgical tools in relation to the patient's coordinate systems. Thus, the surgeon needs to look at the screen and at the same time use the surgical tools to operate (see Figure 3b). In such situations, the surgeon needs to frequently switch visual focus between the surgical site and the screen. During some complex surgeries, the surgeon might call an experienced colleague to help. In such cases, the experienced surgeon provides guidance either over the phone or in person.

In trauma cases, the acute department calls the on-call orthopedic surgeon, and the surgeon must go to the acute department immediately. The acute department sends a short text message to the surgeon giving a brief history of the patient. The surgeon reads the message before arriving at the acute department. In trauma cases, a group of clinicians from different departments work together to save the patient from life-threatening injuries.

Visiting patients in the ward is also part of the daily routine. Surgeons make their ward rounds together with a nurse, moving from patient to patient. The surgeon makes a diagnosis and prescribes treatment facilitated by the nurse, who has general knowledge of the patients and can provide an overview of their current conditions. Before ward rounds, the surgeon needs to review the patient records and recent medical images of the patients on a computer in his or her office. After ward rounds, the surgeon reports results from the rounds using a Dictaphone. Later, administrative personnel transcribe the recorded voice for the EPR system.

Because of clinicians' high mobility, they usually bump into each other in corridors, wards, and other parts of the hospital. In these ad hoc collaborative situations, they might talk about a particular patient or medical task. They might then move to one physician's office to look at a patient's medical information. In some cases, they go to the ward to visit a particular patient and discuss appropriate treatments.

WPAs in Hospitals: Design Framework

Based on the egocentric design approach and the ethnographic study, we developed a WPA design framework for supporting clinicians—in particular, surgeons (see Figure 4). The main characteristics of hospital work are the core elements of the framework, represented by the five rows in Figure 4, while the columns correspond to the three kinds of assistance defined by deploying our egocentric design approach. The boxes within present our initial 12 ideas for our healthcare WPA, which we briefly discuss here.

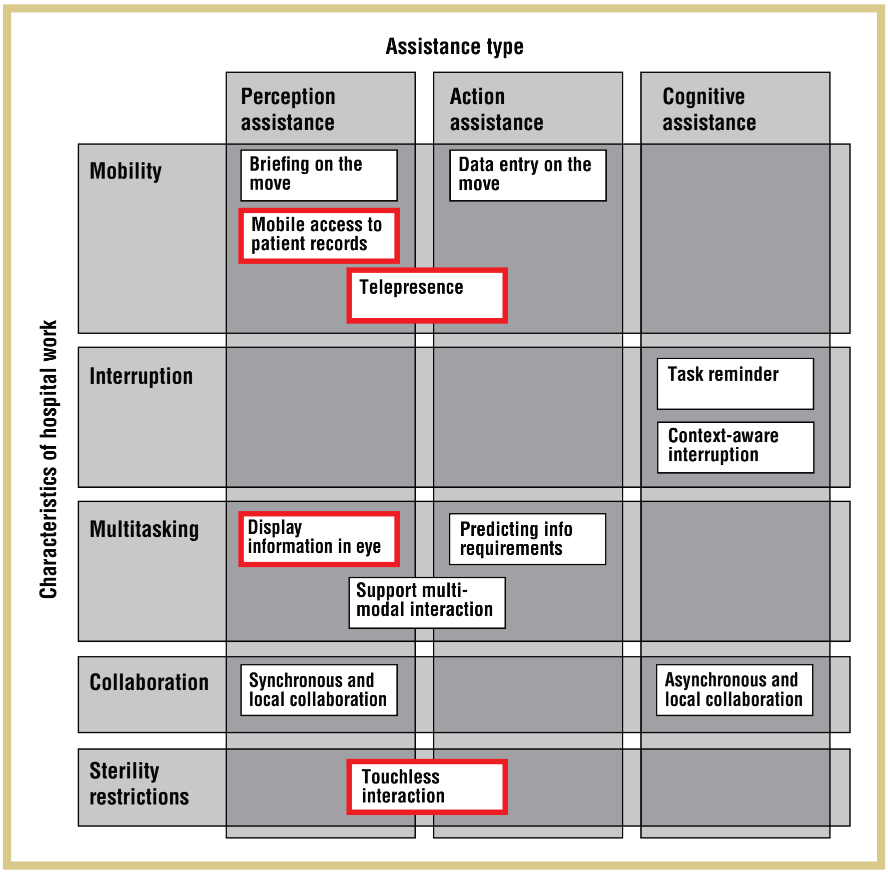


Figure 4. A design framework for WPAs in hospital settings. The rows present the main characteristics of hospital work, while the columns correspond to the three kinds of assistance defined by deploying our egocentric design approach. The boxes within present our initial 12 ideas for our healthcare WPA, while those in bold red were implemented in an initial Google Glass prototype based on feedback from surgeons.

Perception Assistance

Perception assistance includes both augmenting the world by adding needed information to the perception space and simplifying the world by filtering out potentially distracting or irrelevant phenomena from the perception space.

Briefing on the move. To utilize time on the move, the WPA provides information relevant to the situation that the surgeon will soon be entering. For example, surgeons called to the acute department get important health status information about the emergency patient from the WPA.

Mobile access to patient records. The main reason for mobility in hospitals is the need to be in different physical places, to get in contact with a particular person, or to access knowledge and shared resources. A WPA connected wirelessly to the EPR system facilitates information and knowledge sharing and can potentially be further simplified by automatic retrieval based on location (history) of the clinicians (for example, data for the patient nearest the clinician's current location), or bookmarked x-ray images for a surgeon attending a radiology conference.

Telepresence. Although digital resources can be easily shared through mobile devices, sharing physical resources, people, and tacit knowledge is still a challenge. WPAs could help share tacit knowledge among clinicians by offering ad hoc telepresence sessions between a remote specialist and local clinicians.

Display information in eye. During surgery, the surgical team needs to monitor important information on a display. For example, during some orthopedic surgeries, it is necessary to take periodic x-rays of the patient (fluoroscopy). Fluoroscopic surgeries force the surgeon to frequently switch focus between the surgical site and the screen. The WPA can display this information directly to the surgeon's eyes, letting the surgeon maintain focus on the patient, reducing surgery time and avoiding complications from x-ray exposure.

Support for multimodal interaction. The WPA facilitates parallel activities by providing appropriate input and output modalities according to action and perception restrictions of the situation in which clinicians perform medical tasks, as captured by the SSM (Figure 1). For instance, when the user is performing a visual task, the WPA could switch automatically or manually to aural modality for displaying helpful information.

We selected three scenarios for further implementation: mobile access to patient records, telepresence, and touchless interaction and display information in the eye.

Synchronous and local collaboration. In synchronous and local collaborative scenarios, WPAs can increase awareness among team members. For example, during a surgery, the attention point of a surgeon on the surgical site, tracked by a gaze tracker, could be shared with the surgical team.

Cognitive Assistance

Many ideas for perception assistance ultimately target cognitive assistance. Hence, there is some overlap in the assistance mechanisms targeting perception and cognition. What we list here are support mechanisms that aim to affect cognitive processes more directly.

Context-aware interruption. As a continuously running device that stays with clinicians at all times, a WPA could determine current context and determine whether or not its wearer is interruptible in a given situation.

Task reminder. To mitigate the risk of a clinician forgetting a task that was interrupted, the WPA could remind the user about the (state of) the interrupted tasks after interruption.

Asynchronous and local collaboration. Many medical tasks are coordinated through asynchronous collaboration. For example, clinicians update the time schedules of personnel on whiteboards, enter descriptions of medical tasks performed on patients during a work shift into computer systems, and so on. Ubiquitous WPAs would allow for binding virtual objects to physical objects, locations, or situations.

Action Assistance

The main focus of action assistance is to make the surrounding environment easier to manipulate by providing information relevant to the task at hand through appropriate modalities.

Predicting information requirements. The WPA would be able to recognize clinicians' activities and allow them to access relevant information quickly, without sacrificing their connection to the patient or the procedure at hand.

Data entry on the move. In a clinical setting, everything needs to be properly recorded, for legal reasons and to ensure continuity of care. For instance, after visiting a patient in the ward, clinicians should record diagnosis results and decisions. The WPA could support data entry on the move by automatically recording the visited patient's ID and the date, time, and author (WPA user).

Touchless interaction. As noted earlier, in most operating rooms, several computers and large displays monitor different medical information before and during surgery, and because of sterility restrictions, input devices are handled by an assistant or a nurse instructed by the surgeon. This can sometimes cause misunderstandings and delays. A WPA could act as an interface between stationary computers and the surgeon through touchless modalities such as speech, body gestures, and gaze.

A WPA Prototype for Selected Scenarios

We interviewed three orthopedic surgeons to evaluate the utility of the 12 ideas in practice. First, we explained the ideas to the surgeons and asked them to rank the usefulness of each idea using a 5-point Likert scale. Next, we discussed the ideas in more detail with open questions about the situation in which the proposed idea could play a role, how frequently these situations occur, what information is needed in each situation, and what specific restrictions and requirements (such as sterility) should be observed. Based on the results of the interview study, we selected three main scenarios for further implementation: mobile access to patient records, telepresence, and touchless interaction and display information in the eye.

Because of Google Glass's unobtrusiveness, its support of various input channels, such as voice commands, head motion, and touchpad. Consequently, we developed the WPA prototype on the Google Glass platform. The first prototype supports three selected scenarios.

Mobile Access to Patient Records

According to the interviewed surgeons, mobile access to patient records through the WPA provides valuable support in situations such as making ward rounds, working in the operating room, performing ad hoc collaborations. However, each situation requires the WPA to provide different interaction modalities. For example, the WPA should support touchless interaction in the operating room because of the sterility restrictions, whereas during ward rounds, surgeons can use a Google Glass touchpad to provide input to the device. In fact, support for multimodal interaction is a crucial characteristic for the WPA to provide mobile access to the patient records. But different strategies could be applied to support multimodalities; the interaction modalities can be switched automatically or manually in different situations.

To compare the effect of automatic and manual switching between modalities on user performance, we conducted a lab experiment in which we asked participants to complete a physical task (a hotwire game, as illustrated in Figure 5a) while simultaneously answering simple math questions either displayed on the HMD or played through the headset.[18] Participants answered the visual questions through head gestures, and the aural questions using audio modality (speech). We measured participants' performance based on the speed and error rate in the hotwire task. The results of the experiment show that performance was higher when participants answered the questions using the audio modality than when they used visual and gesture-based modalities, probably because the auditory modality interferes less with the motorically and visually demanding hotwire task. Furthermore, the participants' performance was higher when the modalities were switched automatically than when they were switched manually, but manual switching was preferred because of the higher controllability.

Based on findings from this experiment, we implemented redundant input channels for the WPA and let users choose appropriate modalities manually in different situations. The main steps for mobile access of the patient records on the Google Glass application are as follows.

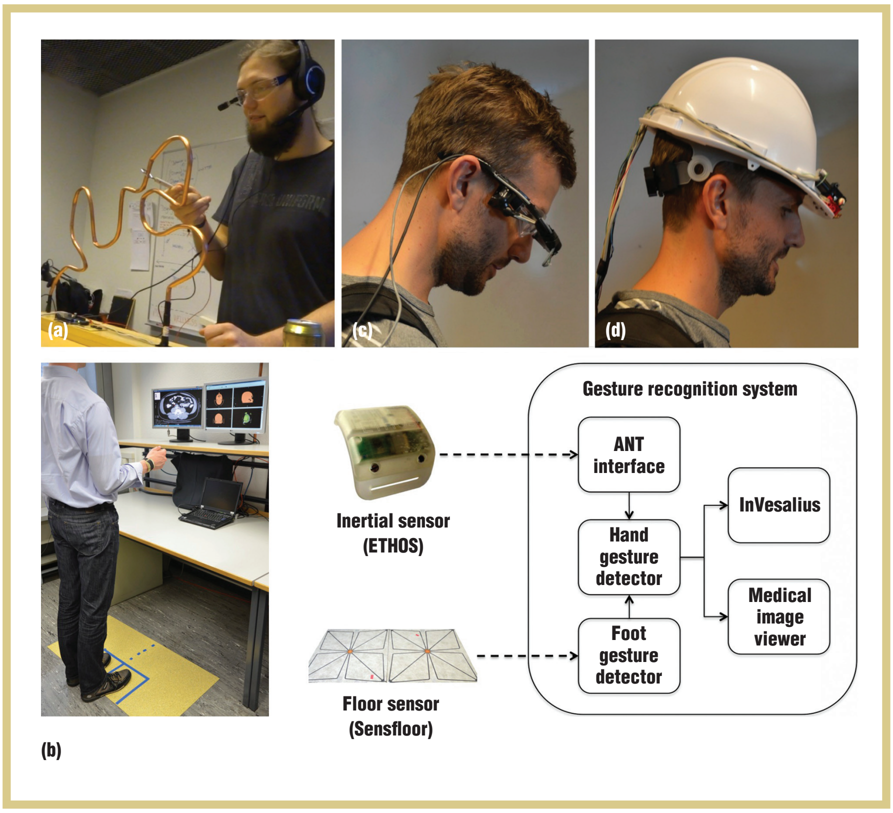


Figure 5. Early experiments informing the design of the WPA prototype: (a) a participant completing the Hotwire task[18]; (b) a participant touchlessly interacting with medical images on the screens,[19] (c) a head-mounted display for teleguidance applications[17]; and (d) a wearable laser pointer.

Find the records. To find and retrieve a patient's health records, the system provides two main channels: voice and QR code. Users can filter the patient list by saying the patient's name or social security number. They can also find the patient's records by reading the patient's QR code using the front-view camera of Google Glass. The latter method is faster and more accurate when the QR code is available.

Switch between textual data and medical images. Patient records are distributed in several pages (cards) on the Google Glass. Because of the display size in Google Glass, the textual part of the patient records is shortened. However, medical images (x-rays, CT scans, and so on) are the most important part for orthopedic surgeons. Users can switch between cards using voice commands (such as "Ok Glass," "next," or "previous") or performing swipe gestures (swipe left and swipe right) on the Google Glass touchpad.

Interact with medical images. Users can zoom in or out, rotate, and navigate through a zoomed-in view of the medical images using voice commands or touch gestures on the touchpad (tap for zooming in, long press for zooming out, and swipe up for rotation). To navigate through an enlarged image in real-time, the Google Glass head tracker sensor is used, allowing the user to quickly scan over the complete image by moving his or her head (Figure 6).

Touchless Interaction

We studied the utility of touchless interaction techniques in the operating room in a lab experiment (Figure 5b).[19] Based on the experiment's promising results, we designed a touchless interaction module for the WPA that enables surgeons to provide touchless input to other computers in the operating room using voice commands and head gestures.

The surgeon can interact with two systems: a 3D medical imaging system

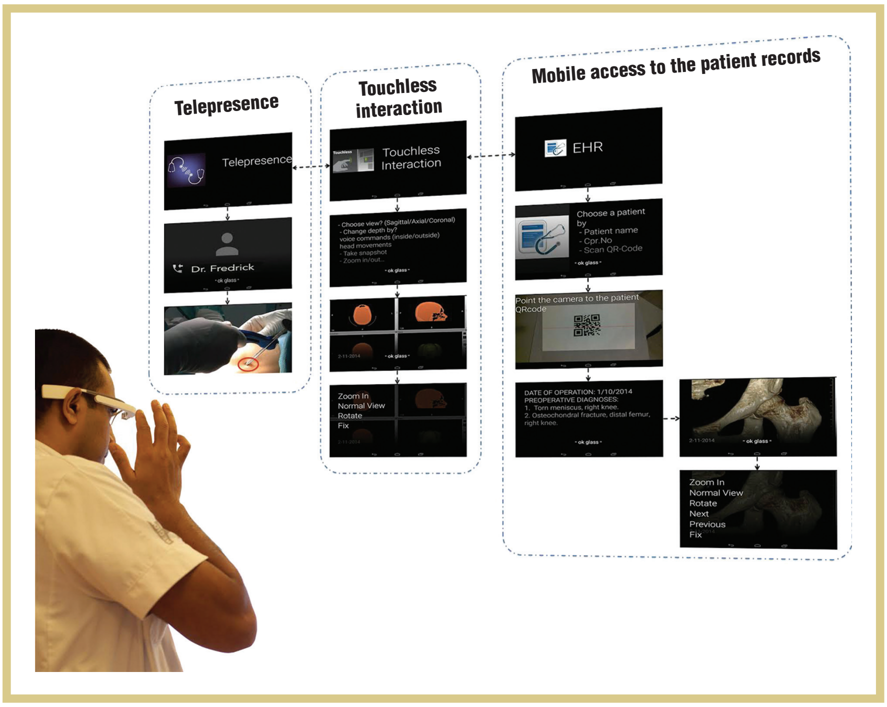


Figure 6. Screenshots of the main cards (pages) of three applications. Users can switch between cards using voice commands (such as "Ok Glass," "next," or "previous") or performing swipe gestures on the Google Glass touchpad. To navigate through an enlarged image in real-time, the Google Glass head tracker sensor is used, allowing the user to quickly scan over the complete image by moving his or her head.

(InVesalius) or a 2D image viewer to review x-rays and other 2D images. The surgeon can switch between 2D images, zoom in or out, and navigate through an enlarged image on the stationery screen via voice commands and head gestures through the WPA. To interact with the 3D imaging system, the user should first choose the desired view (Axial, Sagittal, or Coronal). The surgeon can adjust the view of different slices of the 3D model continuously by vertical movements of the head. To send commands using head movements, we divided the Google Glass screen into three areas distributed vertically. The user's vertical head movement is mapped to the position of a pointer on the screen. When the pointer enters the top area, the depth view of the 3D model increases, whereas crossing the lower border with the pointer decreases the depth. The same method is used to navigate through an enlarged x-ray on the screen, in which the screen is divided into five areas: up, down, left, right, and middle. By moving the pointer from the middle to the four other areas on the Google Glass screen, the enlarged image moves in the same direction on the big screen.

In addition to supporting touchless interaction, this module allows users to take a snapshot from the stationery screens and display it in surgeon's Google Glass HMD (see Figure 6).

Telepresence

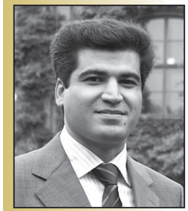
The three surgeons we interviewed noted that, during complex surgeries, the surgeon might need help from an expert colleague. In such situations, the surgeon asks the expert colleague to personally help in the operating room or to provide guidance through a phone call. To enhance the effectiveness of collaborations over the phone, the WPA's telepresence module is designed to share a still image taken by the local surgeon with a remote expert. The remote expert can use a mobile application to see the shared image and provide guidance through vocal communication and also by adding sketches on the still image. The graphical content provided by the remote expert is superimposed on the local surgeon's HMD in real time.

Our comparative study on a remote pointing scenario using a laser pointer and HMD technologies (Figures 5c and 5d) revealed the challenge of stabilizing content provided by the remote expert on the local side when there is a live video stream between local and remote sides.[17] Thus, we used still image and vocal communication in the telepresence module. In addition, because of the technical limitations of Google Glass, the experienced quality of the live video stream is lower than still images, which makes still images the better choice. The feedback from two surgeons who tried the telepresence system also showed that in orthopedic surgeries, sharing still images is more useful than live streams. However, sharing live video can of course be useful in other situations.[20]

At the time of this writing, we just finished an empirical evaluation of the system at the ITX hospital simulation facilities in Copenhagen involving two orthopedic surgeons who tried the WPA in two scenarios: surgery and ward rounds. Preliminary results confirm the WPA's utility and practicality in realistic settings. However, the surgeons and patients reflected on some social challenges in the ward rounds scenario. Our future plan is to improve interaction aspects of the WPA to address the social challenges and other minor issues.

From a theoretical perspective, we hope to help extend current context-awareness research to better explore more powerful future personal assistance systems. Our proposed egocentric design approach, which includes unconscious cognitive processing as part of the system design, is well-grounded in modern cognitive science but poses huge challenges for us as engineers. We believe that the pervasive computing community will take on this challenge with increasing success as body-worn technology and the modeling of human cognition continue to improve.

the AUTHORS



Shahram Jalaliniya is a PhD fellow at the IT University of Copenhagen, where he is a member of the Pervasive Interaction Technology (PIT) Lab. His research interests include wearable computing, HCI, pervasive computing, and multimodal interaction. Jalaliniya has a master's degrees in information systems from Lund University and in software and technology from the IT University of Copenhagen. He is a member of IEEE. Contact him at jsha@itu.dk.

Thomas Pederson is an associate professor at the IT University of Copenhagen, where he is a member of the Pervasive Interaction Technology (PIT) Lab and the Interaction Design Group. His research interests include HCI, context-aware systems, and pervasive/ubiquitous computing. Pederson has a PhD in computing science from Umeå University, Sweden. He is a member of the ACM. Contact him at tped@itu.dk.

ACKNOWLEDGMENTS

We thank Ulrik Kaehler Olesen of the Department of Traumatology and Orthopedic Surgery at the Rigshospitalet, Copenhagen, as well as our fellow Pervasive Interaction Technology (PIT) Lab members for their valuable input at various stages of the design of the WPA prototype. This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

REFERENCES

1. T. Starner, "The Challenges of Wearable Computing: Part 1," *IEEE Micro*, vol. 21, no. 4, 2001, pp. 44–52.
2. G. Lakoff and M. Johnson, *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, Basic Books, 1999.
3. C. Rosen, "The Myth of Multitasking," *The New Atlantis*, vol. 20, Spring, 2008, pp. 105–110.
4. S.J. Brooks et al., "Exposure to Subliminal Arousing Stimuli Induces Robust Activation in the Amygdala, Hippocampus, Anterior Cingulate, Insular Cortex and Primary Visual Cortex: A Systematic Meta-Analysis of FMRI Studies," *Neuroimage*, vol. 59, no. 3, 2012, pp. 2962–2973.
5. R.W. DeVaul, A.S. Pentland, and V.R. Corey, "The Memory Glasses: Subliminal vs. Overt Memory Support with Imperfect Information," *Proc. IEEE Symp. Wearable Computers* (ISWC), 2003, pp. 146–146.
6. T. Pederson, L.-E. Janlert, and D. Surie, "Towards a Model for Egocentric Interaction with Physical and Virtual Objects," *Proc. 6th Nordic Conf. Human-Computer Interaction: Extending Boundaries* (NordiCHI), 2010, pp. 755–758.
7. D.A. Norman and S.W. Draper, *User Centered System Design, New Perspectives on Human-Computer Interaction*, L. Erlbaum Associates, 1986.
8. I. Poupyrev et al., "Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques," *Computer Graphics Forum*, vol. 17, no. 3, 1998, pp. 41–52.
9. T. Pederson, L.-E. Janlert, and D. Surie, "A Situative Space Model for Mobile Mixed-Reality Computing," *IEEE Pervasive Computing*, vol. 10, no. 4, 2011, pp. 73–83.
10. A. Schmidt, "Implicit Human Computer Interaction through Context," *Personal Technologies*, vol. 4, nos. 2–3, 2000, pp. 191–199.
11. M. Weiser and J.S. Brown, Designing Calm Technology, *PowerGrid J.*, vol. 1, no. 1, 1996, pp. 75–85.
12. E. Gendlin, *Experiencing and the Creation of Meaning: A Philosophical and Psychological Approach to the Subjective*, Northwestern Univ. Press, 1962.
13. M. Vidal, D.H. Nguyen, and K. Lyons, "Looking At or Through? Using Eye Tracking to Infer Attention Location for Wearable Transparent Displays," *Proc. ACM Int'l Symp. Wearable Computers* (ISWC), 2014, pp. 87–90.
14. S. Mann, "Continuous Lifelong Capture of Personal Experience with EyeTap," *Proc. 1st ACM Workshop Continuous Archival and Retrieval of Personal Experiences*, (CARPE), 2004, pp. 1–21.
15. J.E. Bardram and C. Bossen, "Mobility Work: The Spatial Dimension of Collaboration at a Hospital," *Computer Supported Cooperative Work* (CSCW), vol. 14, no. 2, 2005, pp. 131–160.
16. K. Adamer et al., "Developing a Wearable Assistant for Hospital Ward Rounds: An Experience Report," *The Internet of Things*, Springer, 2008, pp. 289–307.
17. S. Jalaliniya, T. Pederson, and S. Houben, "Wearable Laser Pointer Versus Head-Mounted Display for Tele-Guidance Applications?" *Adjunct Proc. ACM Int'l Symp. Wearable Computers* (ISWC), 2014, pp. 193–200; doi: 10.1145/2641248.2641354.
18. S. Jalaliniya, "Towards Dynamically Switching Multimodal User Interfaces," MS thesis, IT Univ. of Copenhagen, 2014.
19. S. Jalaliniya et al., "Touch-less Interaction with Medical Images Using Hand & Foot Gestures," *Adjunct Proc. ACM Conf. Pervasive and Ubiquitous Computing* (Ubicomp), 2013, pp. 1265–1274; doi: 10.1145/2494091.2497332.
20. B.A. Ponce et al., "Telementoring: Use of Augmented Reality in Orthopedic Education, *J. Bone & Joint Surgery*, vol. 96, no. 10, 2014, p. e84; doi:10.2106/JBJS.M.00928.

cn Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.

Paper 2: A Wearable Personal Assistant for Orthopedic Surgeons

Title of Paper

Qualitative Study of Surgeons Using a Wearable Personal Assistant in Surgeries and Ward Rounds

Authors:

Shahram Jalaliniya, Thomas Pederson

Published:

Proceeding of eHealth 360: International Summit on eHealth, Budapest, Hungary, June 14-16, Springer International Publishing, 208-219, 2017

Abstract

In this paper, we report on the utility of a wearable personal assistant (WPA) for orthopedic surgeons in hospitals. A prototype of the WPA was developed on the Google Glass platform for supporting surgeons in three different scenarios: (1) touch-less interaction with medical images in surgery room, (2) tele-presence colleague consultation during surgeries, and (3) mobile access to the Electronic Patient Records (EPR) during ward rounds. We evaluated the system in a simulation facility of a hospital with two real orthopedic surgeons. The results of our study showed that while the WPA can be a viable solution for touch-less interaction with medical images and remote collaborations during surgeries, using the WPA in the ward rounds can have a negative impact on social interaction between surgeons and patients.

Qualitative Study of Surgeons Using a Wearable Personal Assistant in Surgeries and Ward Rounds

Shahram Jalaliniya( ) and Thomas Pederson

IT University of Copenhagen, Copenhagen, Denmark
{jsha,tped}@itu.dk

Abstract. In this paper, we report on the utility of a wearable personal assistant (WPA) for orthopedic surgeons in hospitals. A prototype of the WPA was developed on the Google Glass platform for supporting surgeons in three different scenarios: (1) touch-less interaction with medical images in surgery room, (2) tele-presence colleague consultation during surgeries, and (3) mobile access to the Electronic Patient Records (EPR) during ward rounds. We evaluated the system in a simulation facility of a hospital with two real orthopedic surgeons. The results of our study showed that while the WPA can be a viable solution for touch-less interaction with medical images and remote collaborations during surgeries, using the WPA in the ward rounds can have a negative impact on social interaction between surgeons and patients.

Keywords: Wearable Personal Assistant · Google Glass · Hospital work

1 Introduction

Mobility is one of the main characteristics of work in hospitals. Due to the spatial distribution of departments, wards, and offices in clinical settings, clinicians need to move between different departments all the time. Aside from the considerable time that clinicians waste on moving in hospitals, having access to the right information in different situations is a big challenge. The majority of previous work on providing remote access to the patient information have used mobile devices (e.g. PDAs and smartphones). However, most mobile devices do not support interaction on the move, which means the users need to stop, pick up their device, and direct their attention away from the task at hand [1]. This way of interaction often requires the user's full attention and occupies at least one hand which most of the time interferes with the task at hand. Furthermore, interaction with the dominant touchscreen-based mobile devices does not comply with sterility restrictions in hospitals. Emerging new generation of eyewear computers e.g. Google Glass that provide various hands-free input modalities (e.g. head motion and voice commands), raises the question as to whether this new platform can address some of the challenges of interaction on the move.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2017
K. Giokas et al. (Eds.): eHealth 2016, LNICST 181, pp. 208–219, 2017.
DOI: 10.1007/978-3-319-49655-9_28

What are the potential advantages and limitations of using such devices in hospitals? To answer these questions, we implemented and evaluated a wearable personal assistant (WPA) for orthopedic surgeons based on a previous study on design of wearable personal assistants for surgeons [2]. Our WPA supports three specific tasks throughout a workday of surgeons: (1) touch-less interaction with medical images, (2) tele-presence during surgeries, and (3) mobile access to the Electronic Patient Records (EPR) during ward rounds.

2 Related Work

2.1 Early Wearable Assistants for Clinicians

The first generation of wearable computers for hospital work domain [3–7] comprised a head mounted display (HMD), a microphone and earphone for vocal interaction, a compact processing unit connected to a wireless network, and other peripherals such as wrist-mounted keyboards, trackball mice, and etc. RNPSS [3] was one of the first wearable systems for clinicians. The main goal of this system was to decrease the medical errors of nurses. A similar project [6] was done to support nurses in home care tasks. Supporting physicians in ward rounds was another application for the early wearable assistants [4]. The ward round system supported hand gesture interaction using inertial sensors [4] and conductive textile sensor [8]. These initial prototypes of wearable assistant for clinicians increased hopes for using wearable computers in practice, but due to the technical, social, and usability challenges [9] those system never took off.

2.2 Using Google Glass in Healthcare

In [10], an expert surgeon provided guidance to a local surgeon over distance. The guidance was provided through vocal communication and the image of the remote surgeon's hand was superimposed on the live view of the surgical site on the Google Glass HMD. This study showed some problems with battery life, audio and image quality, and difference between camera view and the surgeon view. In another study [11], Google Glass was used to retrieve similar medical cases by sending a picture and relevant keywords to a remote server. In this paper, similar technical issues were reported such as limited battery life, unstable WIFI connection, lack of auto-focus functionality, which decreases the quality of the pictures. Muensterer et al. [12] showed the utility of the Google Glass for hands-free photo and video recordings, hands-free calls, looking up billing codes, and searching for unfamiliar medical terms in a hospital. The feasibility of using Google Glass for monitoring patient's vital signs in the surgeon's eye was investigated by Vorraber et al. [13]. Their study showed that using Google Glass decreases head and neck movements of the surgeon and increases the surgeon's focus on the operation. They reported over-heating problems of the Google Glass in addition to the other technical issues. While previous work has focused on the technical feasibility of using Google Glass in healthcare scenarios, our focus here is on human-computer interaction challenges emerging from using the device as a wearable assistant in hospitals.

In the work presented in this paper we investigate the ecological validity of the WPA design explained in more detail elsewhere [2] by asking real orthopedic surgeons use the WPA in a clinical simulation.

3 Method

Since deploying the WPA in a real clinical setting needs legal approval, we evaluated the WPA in a clinical simulation facility. Such simulations is common and have been proven efficient in the medical work domain [14]. Our simulation facility includes different hospital departments from patient wards to surgery rooms. We set up the facility for the above-mentioned three scenarios. The touch-less interaction and tele-presence scenarios were played out in the surgery room (see Fig. 1b), and for the mobile access to the EPR scenario we set up a patient room with two beds (see Fig. 1a).

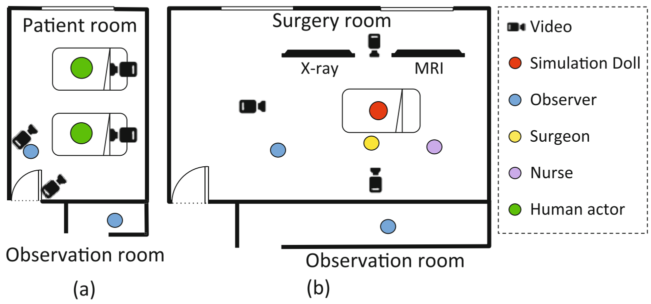


Fig. 1. (a) The simulation setup for the ward round scenario. The room is equipped with hidden cameras, microphones, and an observation room behind a one-way mirror. (b) The simulation surgery room for touch-less interaction and tele-presence scenarios is equipped with surgical equipments, a simulation doll connected to a monitor displaying simulated vital signs, and two large screens for displaying X-rays and Magnetic Resonance Images (MRIs).

3.1 Participants

During a full day simulation, two orthopedic surgeons, a senior nurse, and two human actors (to play the role of patients) participated in the study. Since surgeons are extremely busy and hard to recruit for such studies we could manage to find only two surgeons. This is a big limitation for finding statistical significance; therefore, we only rely on qualitative findings from interviews and observations. The entire simulation was recorded using video cameras, note taking, photographing, and observations behind a one-way mirror. After welcoming the participants, a brief introduction was delivered on the purpose of the study and the scenarios. Both surgeons performed all three scenarios. Before starting each scenario, the surgeons were briefly trained on how to use the WPA.

Each training session took about 30 min. After each scenario, the surgeons were asked to complete a structured questionnaire polling their experiences completing the task and using the system. The result of questionnaires is represented in Fig. 3. Immediately after the questionnaire the surgeons were interviewed to get deeper insights into their experience of using the WPA.

3.2 Scenario-Based Evaluation

We took a scenario-based approach in evaluation of the WPA. The scenarios were defined based on a previous study [2]. Scenarios included: (1) Touch-less interaction, (2) Tele-presence, and (3) Mobile access to the EPR in ward rounds. These three scenarios are part of a bigger scenario which starts with a patient getting an orthopedic surgery. Before the surgery, the surgeon needs to review the medical images of the patient. The WPA helps the surgeon find relevant medical images and adjust the view through touch-less modalities. During the surgery, the surgeon needs another experienced surgeon's opinion about the surgery. The WPA helps the local surgeon to have a tele-presence session with the remote colleague. After surgery, the patient is moved to the ward, and the surgeon visits the patient in the ward. The WPA enables the surgeon to see the patient electronic records on the go and review the new medical images after the surgery.

3.3 Preparing Data for the Study

Since all three scenarios are related to each other, for this study we needed real medical cases. We selected two cases with the help of our medical partner. We anonymized the data and assigned unreal names to the selected cases. Two human actors from university staff played the role of the patients during the ward round scenario. We also used real pictures of the surgical site taken during real surgeries. The pictures were printed and attached to the simulation doll to create a more realistic view (see Fig. 2-a).

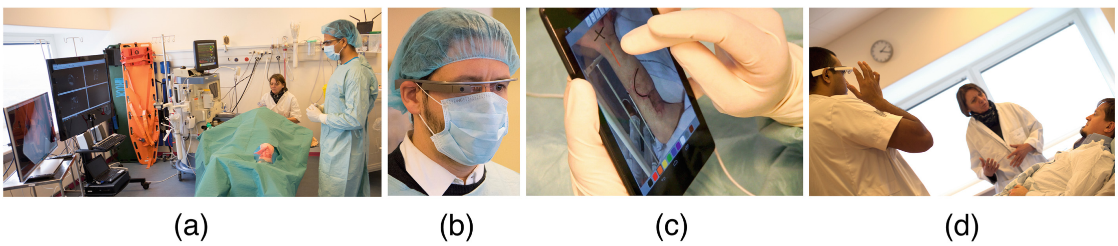


Fig. 2. (a) A surgeon uses the WPA for touch-less interaction with X-rays and MRIs. (b) The local surgeon sees the visual guidance on the HMD in real-time. (c) A remote surgeon uses a tablet device to provide guidance to the local surgeon. (d) A surgeon uses the WPA to browse EPR and X-rays in the ward round scenario.

4 Scenario 1: Touchless Interaction

In the surgery room, the surgical team including a surgeon and a nurse, are about to start the surgery. Before starting the surgery, the surgeon looks at X-rays and MRIs. But his/her hands are sterile and s/he cannot touch the mouse or keyboard. Therefore, the surgeon uses the WPA for browsing X-rays and MRIs on two different screens in the operation room through voice commands and head movements. The surgeon might need to zoom in, rotate, or navigate through the medical images until s/he finds a good view. The surgeon can also take a snapshot of the screens and see the content on the HMD.

4.1 Apparatus

We used Google Glass to implement the WPA since Google Glass provides at least two touch-less input modalities: voice commands and head movements. Moreover, the unobtrusive form factor of the Google Glass and covering small part of the users field of view makes the Google Glass the best available option for applications where having a good view over the real-world is crucial. We developed a simple image browser for displaying the X-rays in the surgery room. To visualize MRIs and X-ray scans, we modified Invesalius software that is an open-source medical imaging system[1].

All three systems were connected to a dedicated local WIFI network. We used UDP protocol for communication between Google Glass and other two medical systems. The WPA app on the Glass accepts both voice commands and head movements for interaction. Voice modality is used for discrete commands such as activating/deactivating the interaction, switching between X-rays, zooming in/out X-rays, changing the views in MRIs between (sagittal, coronal, and axial). While head motion is used for continuous commands such as adjusting the position of the X-rays on the screen. In the latter case, we used the user's head similar to a mouse where the vertical and horizontal head movements are translated into the vertical and horizontal movements of the pointer displayed on the HMD. We defined some command areas in the GUI of the Google Glass. By moving and keeping the pointer in each area, the WPA sends an appropriate command to the X-ray and MRI systems. As soon as the pointer exits from the selected area the WPA stops sending commands. Table 1 shows the modalities used for sending commands to the WPA.

4.2 Procedure

After briefing the participants and setting up the surgery room, the surgeons started the scenario one after another. First the nurse gave a brief explanation about the patient to the surgeon. Then the surgeon activated the stationary X-ray system through the WPA to find an X-ray and adjust the scale and position of it on the large display. To find a good view the surgeon used either voice

[1] http://svn.softwarepublico.gov.br/trac/invesalius/.

Table 1. Input modalities for each module of the WPA

System module	Commands to the WPA	Voice	Head	Touch
Touchless interaction	Wake up the Glass		×	
	(De)activate the X-ray/MRI system	×		
	Switching X-rays (next/previous)	×		
	Positioning X-rays on the screen		×	
	Changing MRI views	×		
	Change the depth of the MRI views		×	
	Take snapshot of X-rays/MRI views	×		
Tele-precense	Wake up the Glass		×	
	(De)activate the tele-precense system	×		
	Take a picture	×		
	Select a picture for sharing	×		
	Call a clinician	×		
	End call	×		
EPR	Wake up the Glass		×	×
	(De)activate the EPR system	×		×
	Select a patient record	×		×
	Switch X-rays	×		×
	Zoom in/out X-rays	×		×
	Rotate X-rays	×		×
	Navigate through X-rays		×	
	Browse EPR pages	×		×

commands or head movements as shown in Table 1. After finding the appropriate view, the surgeon took a snapshot of the stationary X-ray which made it come up on the HMD. This snapshot helps the surgeon to examine the X-ray image during the surgery without having to change the head orientation towards the large display. Each surgeon repeated the scenario for both patients. Since the second patient had also some MRIs, in the second surgery, the surgeons used the WPA for interaction with both X-rays and MRI systems. To interact with the MRI system, the surgeon needed to activate three different views (sagittal, coronal, axial) through voice commands and adjust an appropriate depth view.

4.3 Results

Interview: We asked surgeons about the pros and cons of the WPA for touchless interaction compared to the current indirect interaction (asking a nurse to control a computer mouse as proxies for surgeons). Participant 1 (P1) indicated the higher speed of interaction using the WPA; however, he believes that it might take more time for older surgeons to learn how to use the WPA. P2 thinks the direct interaction through the WPA can be a big advantage and saves time of surgeons in the surgery room because sometimes it is very hard to explain to a nurse the view that the surgeon is looking for. However, interaction with X-rays by head movements is not easy since the user needs to look through the HMD to see the pointer and at the same time look at the X-rays or MRIs on the large screens which demands frequently switching between the HMD and the large screens.

We also asked whether they prefer voice commands or head movements for interaction with X-rays and MRIs. P1 thinks the voice commands are more convenient for interaction with X-rays where the user usually needs to provide a few commands while in the MRI case the head movements can be more beneficial since finding the right depth view among a lot of slices can be frustrating by voice commands. P2 prefers voice commands since interaction through head movements was challenging for him due to the need for switching frequently between the HMD and the large screens.

The last question was about the snapshot function. Both P1 and P2 indicated that the snapshot functionality can be extremely useful when the surgeon needs a reference X-ray or MRI to monitor the state of the surgical site during the surgery. In such cases, the surgeon needs to frequently turn his/her head towards the screen. To have a snapshot of such reference images in the HMD, saves surgeons' time and energy for the surgery.

Observations: Both surgeons quickly learned how to use the voice commands for interaction through the WPA; however, P1 felt more comfortable with head-based interaction compared to P2. When P2 wanted to adjust the position of the X-rays in the screen by head movements, he lost the control of the system because he had problems with looking at both the HMD (to control the pointer) and the large screen (to see the X-rays) at the same time. The same problem happened when P2 wanted to adjust the MRI depth view.

5 Scenario 2: Tele-Presence

After adjusting the medical images on the screen (in the previous scenario) during the surgery, the surgeon encounters a complex situation and needs help from an expert colleague. The surgeon uses the WPA to start a tele-presence session with the remote colleague. The local surgeon takes a picture of the surgical site and calls the remote surgeon using the Glass. The remote surgeon answers the call. Then the local surgeon explains the situation and shares the taken picture with the remote surgeon. The remote surgeon provides some voice guidance while at the same time marking the shared photo on his tablet (Fig. 2c). The local surgeon sees the content provided by the remote surgeon on the Glass and also hears the voice of the remote surgeon in real-time (Fig. 2b) in real-time.

5.1 Apparatus

We developed a tele-presence app on the Google Glass for the local surgeon while for the remote surgeon, we developed an Android application on an Asus

Nexus 7 tablet. The audio communication is done over WIFI connection using UDP protocol. Due to the limitations in processing resources of the Google Glass and to avoid registration challenges in an augmented reality user interface, the Glass application shares a still picture (instead of video) of the local side, and the remote person is able to draw sketches on top of the shared image using the Android application on the tablet. The sketches are superimposed over the shared image in real-time on the Google Glass HMD of the local user.

5.2 Procedure

In the tele-presence simulation, we ran the scenario twice, and during each time one of the surgeons played the role of a remote expert and the other surgeon played the role a local surgeon. In the second run, the surgeons swapped their role and the surgery case was also changed from patient 1 to the patient 2. Before starting each run, we attached the printed image of the surgical site on the simulation doll. The remote surgeon sat on a chair in the hallway outside of the surgery room. After activating the Google Glass by head nudge gesture, the local surgeon opens the tele-presence application by voice command and takes a picture of the surgical site. Then the local surgeon calls the remote colleague by saying his/her name from a list on the HMD. The remote surgeon receives and accepts the call. As soon as the call is accepted the audio communication is possible and the taken picture is displayed on both sides. The local surgeon explains the situation and asks for the remote surgeon's opinion. The remote surgeon provides vocal and visual guidance by marking the shared image of the surgical site using different colors on his tablet device, markings that show up immediately in the Google Glass display carried by the local surgeon.

5.3 Results

Interview: We asked surgeons what other content they would like to share in a tele-presence session. P1 believes sharing still images of the surgical site (like our implementation) is very useful for orthopedic surgeries while live videos can be useful in emergency cases. Also sharing medical images such as X-rays or MRIs can be valuable in cases where a junior surgeon needs an approval from a senior surgeon. Currently the senior surgeon needs to come personally to the surgery room and have a look at the X-rays or the junior surgeon sends the X-ray using a smartphone. P2 thinks the quality of the image on the HMD is not good enough for complex surgeries with a lot of soft tissues. He suggested to add a zoom-in functionality to overcome the limited resolution of the HMD.

Observations: The communication between the two surgeons was smooth. There was about half a second delay in the audio communication due to the WIFI-based communication. But the surgeons got used to it after a while. Also during the tele-presence scenario, when the local surgeon was talking to the remote surgeon, Google Glass detected the "Ok Glass" command by mistake and the surgeon needed to deactivate the voice command and continue the session.

6 Scenario 3: Mobile Access to Electronic Patient Records

It is one day after the surgeries. Patients are lying down in the bed in the ward. The surgeons should visit two patients who got surgery. The surgeons use the WPA to review the new X-rays and the latest state of the patients while walking to the ward together with a nurse. The surgeon searches for the patient records on the Glass by saying the patients name. After finding the patient records, the surgeon reads the updated EPR and looks at the recent X-rays and MRIs on the Glass. The surgeon zooms in/out, rotate, and navigate through the medical images. The nurse reports the latest state of the patient (last blood test, etc.) to the surgeon. The nurse answers the questions that the surgeon might ask during the ward round. The surgeon visits the patients and asks some questions about their pain, etc. Also the surgeon might need to use the EPR system for answering patients' questions. After visiting the patients, the surgeon prescribes the next treatments and the nurse writes down the prescriptions.

6.1 Apparatus

For this scenario, we only needed an EPR app on the Google Glass. Since in the ward round, the clinicians' hands do not necessarily need to be sterilized, the EPR app supports also touch-based interaction on the Google Glass side touchpad. Table 1 shows the ways surgeons can interact with the EPR app. We used different touch-gestures for interaction with text pages and medical images: swipe front/back for browsing EPR and X-rays, short tap for zoom in, long tap for zoom out, swipe up for 90 ° rotation, and swipe up to exit from an active card to the previous menu. Since it was not possible to connect the Google Glass to the EPR in the hospital, the patients data was hard-coded into the EPR app.

6.2 Procedure

In the ward round simulation, each surgeon performed the ward round scenario once where both patients (human actors) lying in the patient bed (Fig. 1-a) were visited. A nurse accompanied the surgeon during the ward round and provided necessary information. The surgeons used the WPA to see the recent EPR and X-rays while talking to the patients (see Fig. 2-d). They tried both voice commands and touch gestures to interact with the WPA. The patients also asked some questions about the result of the surgery.

6.3 Results

Interview: The surgeons were asked about the pros and cons of the EPR module during ward rounds. P1 mentioned that the most obvious advantage of using the EPR on the Glass is to reduce unnecessary moving between a stationary computer and the ward to check the EPR. However, P2 thinks the small screen in Google Glass makes it hard for the surgeon to read the EPR texts, while a stationary computer is more convenient for such intensive readings. P1 also mentioned that getting an overview of the EPR is much faster using a desktop computer since in Glass the text is distributed over several pages.

The other question was about the content that surgeons might need to have access to during a ward round in addition to the EPR and medical images. P2 mentioned that the main information the doctors need during a ward round is lab results that can also be provided on the Glass. However, due to the small size of the HMD in Google Glass, the lab results should be visualized in a way that the interesting results (important abnormal values) are highlighted, and the surgeon can get what s/he wants at a glance. P2 indicated that aside from the medical data, patients usually ask a lot of practical questions about e.g. when they can leave the hospital, when they have their next appointment, etc. The WPA should also provide such practical information to the surgeon.

We also asked about the modality they prefer to use during ward rounds. P1 mentioned that he prefers touch gestures since the voice commands interfere with communication with the patient. P2 said "*I also prefer touch gestures because the head movements look bizarre!*". All participants (two surgeons, a nurse, and two patients) were asked about the social acceptance of the Google Glass. P2 said: "*Some people might think wearing such a [smart] glasses is arrogant since you are not present with the patient*". The nurse mentioned that sometimes the surgeon was looking at the HMD but she thought the surgeon is looking at her. Moreover, the patients mentioned that they did not feel good when the surgeon was trying to interact with the Google Glass instead of talking to them.

Observations: During the ward round, P1 spent more time for interaction with the WPA compared to P2 and sometimes there was a long silence until the surgeon read the EPR on the Google Glass. The reason was that P2 was familiar with the medical cases used in the simulation while both cases were new for P1.

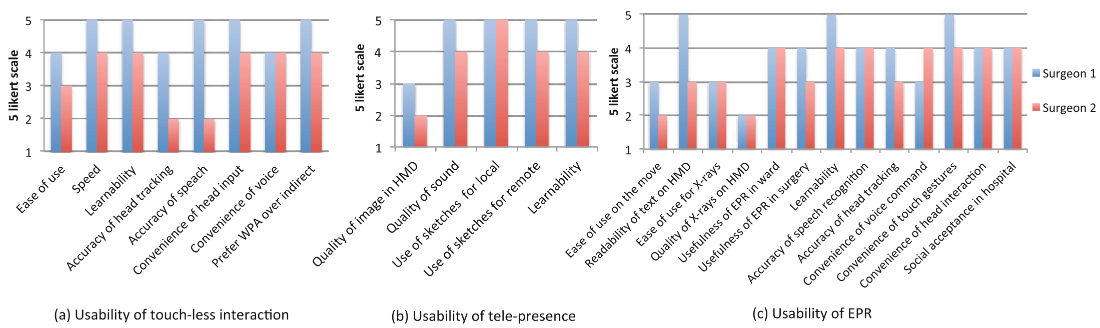


Fig. 3. (a) Usability of the touch-less interaction module of the WPA, (b) usability of the tele-presence module of the WPA, and (c) usability of the EPR module of the WPA.

7 Discussion and Conclusions

Our study indicates that using the WPA for touch-less interaction with medical images can save surgeons time and energy for the surgery. Moreover, by using the WPA for touch-less interaction, there is no need for a dedicated nurse to control the mouse for surgeons. However, there are some limitations in both voice commands and head movements for touch-less interaction. Using voice commands is a relatively reliable modality but due to the slow speed of the discrete voice commands, it is not an appropriate modality for providing a lot of commands within a short time. In contrast to the voice commands, the head movements can be useful for continuous interactions; however, due to the perceptual overlap between seeing the large screens and the pointer on the HMD, it is not easy to use the head movements as a mouse to control the pointer on the HMD. The lowest scores in Fig. 3-a are related to the accuracy of head tracking specially by P2. This reveals the challenge of using head movements for touch-less interaction.

Apart from the low quality of the image on the HMD indicated in both questionnaire (Fig. 3-b) and the complementary interview, the WPA was successfully used in the tele-presence scenario. As both surgeons mentioned, the tele-presence scenario was the best application for the WPA. However, in this scenario we observed the problem of overlapping between human to human conversation and voice commands to the system. This indicates a need for more touch-less input modalities (e.g. gaze) to avoid overlap between the input modality (voice commands) and surgeons' conversation. The most challenging scenario was the ward round which revealed the social problems of using Google Glass in parallel with human to human interactions. Apart from the social problems, the small HMD of the Glass turned out to be a limitation for intensive text readings which is in line with the concept of microinteractions [15] where interacting with the device should not exceed 4 s. To achieve such fast interactions, the WPA needs to prepare the information for the surgeons in a way that the surgeon can get what s/he needs at a glance.

The three scenarios in this paper are representatives of three types of interaction. (1) The touch-less interaction scenario defines the WPA as an *interface between the user and other computers*. In this type of scenarios, the human agent interacts with two different computers in parallel. (2) In the tele-presence scenario, the WPA is defined as an *interface between two human agents* which means the user interacts with another human agent through the WPA and there is no parallel interaction. (3) In the ward round scenario, the user interacts with another human agent and with the WPA *in parallel*. If we look at the results of the questionnaires and interviews, we can conclude that the WPA got the best scores in the tele-presence scenario where there was no parallel interaction, and the user interacts sequentially with the WPA and the other human agent. In the touch-less interaction scenario, the usability of the WPA is evaluated as average. In this scenario, the user interacts with two computers in parallel: the WPA and X-ray/MRI systems. The most challenging scenario is the ward round where the user needs to interact in parallel with the WPA and a human agent.

Acknowledgements. We thank Sanne Jensen at the ITX hospital simulation facility as well as MD Ulrik Kaehler Olesen, Dept. of Traumatology and Orthopedic Surgery, Rigshospitalet, Copenhagen. This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

References

1. Marshall, J., Tennent, P.: Mobile interaction does not exist. In: CHI EA 2013, pp. 2069–2078. ACM, New York (2013)
2. Jalaliniya, S., Pederson, T.: Designing wearable personal assistants for surgeons: an egocentric approach. IEEE Pervasive Comput. **14**(3), 22–31 (2015)
3. Windyga, P., Wink, D.: A wearable computing based system for the prevention of medical errors committed by registered nurses in the intensive care unit. In: 24th Annual Conference of Engineering in Medicine and Biology, vol. 3, pp. 1940–1941, October 2002
4. Adamer, K., Bannach, D., Klug, T., Lukowicz, P., Sbodio, M.L., Tresman, M., Zinnen, A., Ziegert, T.: Developing a wearable assistant for hospital ward rounds: an experience report. In: Floerkemeier, C., Langheinrich, M., Fleisch, E., Mattern, F., Sarma, S.E. (eds.) IOT 2008. LNCS, vol. 4952, pp. 289–307. Springer, Heidelberg (2008). doi:10.1007/978-3-540-78731-0_19
5. Liu, D., Jenkins, S., Sanderson, P.: Clinical implementation of a head-mounted display of patient vital signs. In: ISWC 2009, pp. 47–54, September 2009
6. Drugge, M., Hallberg, J., Parnes, P., Synnes, K.: Wearable systems in nursing home care: prototyping experience. IEEE Pervasive Comput. **5**(1), 86–91 (2006)
7. Thomas, B.H., Quirchmayr, G., Piekarski, W.: Through-walls communication for medical emergency services. Int. J. Hum. Comput. Interact. **16**(3), 477–496 (2003)
8. Cheng, J., Bannach, D., Adamer, K., Bernreiter, T., Lukowicz, P.: A wearable, conductive textile based user interface for hospital ward rounds document access. In: Roggen, D., Lombriser, C., Tröster, G., Kortuem, G., Havinga, P. (eds.) EuroSSC 2008. LNCS, vol. 5279, pp. 182–191. Springer, Heidelberg (2008). doi:10.1007/978-3-540-88793-5_14
9. Levin-Sagi, M., Pasher, E., Carlsson, V., Klug, T., Ziegert, T., Zinnen, A.: A comprehensive human factors analysis of wearable computers supporting a hospital ward round. In: IFAWC 2007, pp. 1–12, March 2007
10. Brent, A., Lasun, O., Phani, K., et al.: Emerging technology in surgical education: Combining real-time augmented reality and wearable computing devices. Orthopedics **37**(11), 751–757 (2014)
11. Widmer, A., Schaer, R., Markonis, D., Muller, H.: Facilitating medical information search using google glass connected to a content-based medical image retrieval system. In: Proceedings of EMBC 2014, pp. 4507–4510, August 2014
12. Muensterer, O.J., Lacher, M., Zoeller, C., Bronstein, M., Kbler, J.: Google glass in pediatric surgery: an exploratory study. Int. J. Surg. **12**(4), 281–289 (2014)
13. Vorraber, W., Voessner, S., Stark, G., Neubacher, D., DeMello, S., Bair, A.: Medical applications of near-eye display devices: an exploratory study. Int. J. Surg. **12**(12), 1266–1272 (2014)
14. Ahmed, K., Jawad, M., Abboudi, M., Gavazzi, A., Darzi, A., Athanasiou, T., Vale, J., Khan, M.S., Dasgupta, P.: Effectiveness of procedural simulation in urology: a systematic review. J. Urol. **186**(1), 26–34 (2011)
15. Ashbrook, D.L.: Enabling Mobile Microinteractions. Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA, USA, AAI3414437 (2010)

Paper 3: Wearable Laser Pointer vs Head-Mounted Display

Title of Paper

Wearable Laser Pointer Versus Head-Mounted Display for Tele-Guidance Applications?

Authors:

Shahram Jalaliniya, Thomas Pederson, Steven Houben

Published:

Proceedings of the ACM International Symposium on Wearable Computers: Adjunct Program (ISWC' 14 Adjunct), ACM, 2014.

Abstract

Wearable camera and display technology allow remote collaborators to guide activities performed by human agents located elsewhere. This kind of technology augments the range of human perception and actuation. In this paper we quantitatively determine if wearable laser pointers are viable alternatives to Head-Mounted Displays for indicating where in the physical environment the local agent should direct her/his attention. The potential benefit of the laser pointer would be reduced eye fatigue, due to the fact that the documented refocusing challenges associated with HMD use would be completely eliminated. 10 participants where asked to perform a short tele-guided pick-and drop task using both approaches. The quantitative analysis indicates that user performance in the laser pointer condition is higher than the HMD approach ($P = .064$, $\alpha = 0.1$). While all 10 participants found the task easy in both conditions, 8 of 10 participants found the laser pointer system more convenient.

Wearable Laser Pointer Versus Head-Mounted Display for Tele-Guidance Applications?

Shahram Jalaliniya
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S, Denmark
jsha@itu.dk

Thomas Pederson
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S, Denmark
tped@itu.dk

Steven Houben
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S, Denmark
shou@itu.dk

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ISWC'14 Adjunct, September 13 - 17, 2014, Seattle, WA, USA
Copyright 2014 ACM 978-1-4503-3048-0/14/09...$15.00.
http://dx.doi.org/10.1145/2641248.2641354

Abstract
Wearable camera and display technology allow remote collaborators to guide activities performed by human agents located elsewhere. This kind of technology augments the range of human perception and actuation. In this paper we quantitatively determine if wearable laser pointers are viable alternatives to Head-Mounted Displays for indicating where in the physical environment the local agent should direct her/his attention. The potential benefit of the laser pointer would be reduced eye fatigue, due to the fact that the documented refocusing challenges associated with HMD use would be completely eliminated. 10 participants where asked to perform a short tele-guided pick-and drop task using both approaches. The quantitative analysis indicates that user performance in the laser pointer condition is higher than the HMD approach (P = .064, $\alpha = 0.1$). While all 10 participants found the task easy in both conditions, 8 of 10 participants found the laser pointer system more convenient.

Author Keywords
Remote collaboration, tele-presence, tele-pointing, head-mounted display, laser pointer, wearable computers.

Introduction

Tele-presence technologies facilitate collaboration over distance by allowing domain experts to oversee and guide work processes in cases when they do not have the possibility to be physically co-located. Healthcare, mining, and maintenance are classical applications. In this paper we compare one of the most investigated approaches for presenting information to the person being guided (the HMD approach) with one much less explored: the use of wearable motor controlled laser pointers. Instead of presenting information on a semi-transparent display in front of the human agents eye(s), information is instead projected directly into the physical environment.

Head-Mounted Displays for telepointing applications
Wearable tele-guidance systems allow remote users to have a situational awareness of the current task environment also in mobile settings, while traditional stationary tele-conferencing systems tend to constrain activities to fix locations. A typical mobile setting includes, on the local side (the location where someone needs support), a head-mounted display (HMD), a head-mounted camera that captures the field of view of the wearer, and a small wearable processing unit connected wirelessly to a remote computer. This specification adequately describes state of the art HMD solutions offered by for instance Vuzix and Google. As HMDs become smaller and less obtrusive, they become interesting candidates for a growing set of mobile interactive applications including tele-presence and tele-pointing.

However, the new emerging HMDs still suffer from known limitations and challenges. Social acceptance, eye fatigue, and focusing problems are well documented (e.g. [8]). Laser pointers could be an interesting alternative for certain kinds of remote collaboration. While HMD tele-pointing solutions often rely on a video see-through Augmented Reality approach where the pointing cursor appears together with a video image of the local environment pictured on the HMD, laser pointer solutions show the remotely controlled pointing cursor directly in the real world environment. Thus, there is no need for the user to change focus depth or perform cognitive work to align the streamed image with the real world. However, the display of more complex content (beyond a point cursor) can be more challenging than when using the pixel matrix offered by HMDs.

Laser pointer Versus Head-Mounted Display?
Previous studies on remote collaboration systems have mainly focused on evaluating just one of these technologies in isolation or in combination [8]. We argue that the laser pointing approach alone could be an interesting alternative for tele-pointing applications. If performance on isolated single-person tasks such as the one investigated in this paper turns out to be comparable, laser pointer solutions could potentially outperform HMD-based solutions for a) very intense telepointing tasks where HMDs would cause fatigue, and b) for tasks where sharing of the remotely provided tele-guidance information with co-located peers is an advantage.

Related Work

Remote guidance technologies fall into three main categories: (1) stationary systems, (2) robot-mounted technologies, and (3) wearable solutions. In the stationary approach, a remote expert provides guidance to a local user by drawing or pointing to a specific object in the task space. This graphical information could be displayed on a monitor over the video streaming from the local side, or it could be overlaid on the physical objects by a stationary laser pointer [10]. In the robot-mounted systems, the combination of a camera and a laser pointer on a movable machine [12] or on a robot [7] allows a remote user to control laser pointer and point to any particular object. Wearable tele-guidance systems have typically been designed to support mobile users. A head-mounted camera carried by local users share their view of the real world and what they are doing with a remote collaborator. The remote instructor provides some graphical instructions, which could be visible for the local user through a HMD [1] or using a combination of HMD and laser pointer [8].

The types of remote guidance found in literature can be classified into four categories [5]: (1) cursor pointer (the local pointer follows the remote instructors mouse pointer); (2) laser pointer (the local laser pointer rests at a location determined by the remote instructor through a mouse click), (3) sketching [2] (the instructor draws figures, not just points), and (4) hand gestures (a representation of the instructors hands are shown to the local user). While previous studies have proved the superiority of the digital sketches over cursor pointer [3], and faster performance of hand gestures, no significant difference has been reported between user performance when receiving information projected directly onto physical objects vs. information displayed on an external monitor [6] such as the study presented in this paper. Finally, a combination of laser pointer and HMD has been proven to lead to a significant improvement in task completion time [8].

Another alternative to the laser pointer technology for Augmented Reality systems is using pocketsize Pico projectors [4], but the luminance of the state of the art projectors is less than laser pointers which limits the applications of portable projectors to indoor and low-light conditions. However the complexity of the content that can be projected by Pico projectors is much higher than laser pointers.

Stationary laser pointers have also been explored as an alternative to HMD for Augmented Reality applications [9] but our study is the first attempt to develop and evaluate a wearable laser pointer as an alternative to HMD for remote collaboration.

Research Question

Given the known challenges of HMDs such as eye fatigue, is motor-controlled laser pointer technology a viable alternative to HMDs for mobile remote guidance applications? We intend to answer this question by measuring user performance in both cases given the same tele-pointing task.

Experimental Design

To compare the task performance of users wearing both the head-mounted display (HMD) and wearable laser pointer, we conducted a comparative within subjects study. The study explores the response times of participants for a simple pick and drop task while being instructed by a remote instructor.

The experiment design was inspired by previous work in tele-guidance systems and special care was taken to reduce uncontrollable noise and to not bias the experiment in favor of any of the two conditions. For both conditions, no image/pointing stabilization system was used and only nearby objects were pointed at.

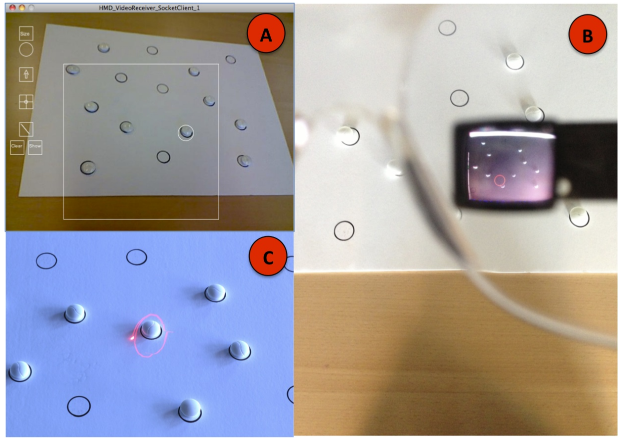


Figure 1: The user interface of the helper station (A), HMD-based system (B), and Laser pointer system (C).

Technical Setup

Both of the wearable remote guidance systems consist of two main components: a wearable system for the local user and the separate helper station which is controlled by the remote instructor. Both the user interface (UI) of the helper station (Figure 1A) as well as the remote instructor using it remained identical for both the laser and HMD condition throughout the whole experiment. The white square-shaped border in the UI (Figure 1A) indicates the area of the local environment to which the remote instructor can point remotely. Since the motor-controlled laser pointer did not cover the whole field of view of the camera, the same limited square-shaped pointing area was enforced also for the HMD condition. Although four different symbol presentations are supported by both systems (dot, circle, line, and polygon) we only made use of the circle symbol. The helper station communicated to the wearable systems through a WIFI network over the UDP protocol with very limited latency.

HMD-based system
In order to build a video-see-through HMD, we attached a webcam (1.3 MP) previously embedded in a laptop and a HMD (MicroOptical SV-9, 640×480 pixels) to an ordinary laptop (Macbook Pro 13 inches) residing in a backpack (Figure 2E).

Laser pointer system
The wearable laser pointer system consists of a similar laptop computer connected to a microcontroller to control a pair of galvanometers. The galvanometers have two mirrors to change direction of the laser point in X and Y dimensions. The galvanometers, laser pointer, and a laptop webcam (1.3 MP) was mounted on a helmet (Figure 2D). The maximum angle of the galvanometer is $30°$ which is slightly less than the maximum range of the camera $(40°)$. Therefore, we limited the pointable area to the white-bordered square shown in Figure 1A. In laser pointer systems, there is always a potential displacement between the intended (clicked) points on the screen and the actual laser-highlighted position in the real world. One mitigation strategy is to calibrate the system for different distances and use a depth sensor to adapt. Our approach was to place the camera very close to the laser pointer (<1cm) and calibrate the system for an average distance (2m) resulting in an accuracy of <5 pixels of error in the range of 1 to 5m.

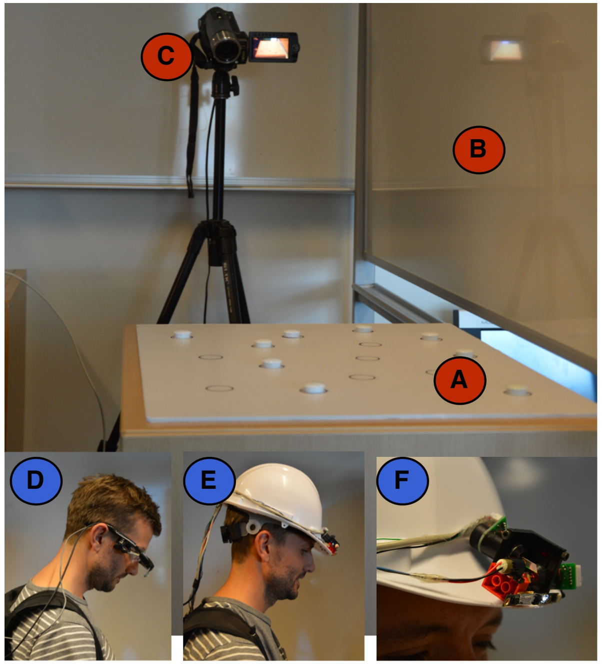


Figure 2: The experimental setup consisted of (A) a desktop with a number of magnets and indicators, (B) a separator to visually shield the remote instructor from the participant and (C) a high-resolution camera to capture the interaction between the participant and the board. The apparatus used for the experiment was a (D and F) custom-built remotely controlled laser pointer and (E) an off-the-shelf HMD.

Participants
10 participants (mean age=35, 1 female) were recruited to participate in the experiment. Participants were all highly skilled computer users ($\bar{X} = 5$, $\sigma = 0.7$). The setup consisted of a table with a white board containing a number of circular indicators and physical magnets (Figure 2A), a separation screen to visually separate the remote instructor from the participant in an effort to emulate remote guidance (Figure 2B). The entire experiment was captured in 1080 60p full HD video (Figure 2C) which was manually post-hoc annotated to measure the response times of the users.

Apparatus
In the laser pointer condition (Figure 2D), a custom-built remotely controlled laser pointer projected information directly into the real world environment. In the HMD condition, a head-mounted monocular display was used onto which the remote pointer information was displayed, blended with a video image of the environment in front of the participant. Both conditions included a wearable camera that allowed the remote instructor to see what the participants had in front of them.

In both conditions, the remote instructor could point to a specific magnet on the board (see Figure 2A) using a physical (in the laser pointer condition) or a digital (in the HMD condition) tele-pointer. The guidance system running on a computer at the remote end allowed the remote instructor to use four types of pointers, but only the circle was used in this experiment:

Procedure
The experiment started with a short introduction to the purpose of the experiment and the use of the apparatus. After participants were prepared for the experiment (for both conditions), they were asked to use the system until they felt comfortable. This usually took 1-2 minutes. Next, the participant was asked to complete the main task. The task consisted of picking up and dropping the magnet that was indicated by the remote instructor. The participant sees this indication either through the laser physically pointing to the board (in the laser pointer

condition) or through the video overlay in the HMD (in the HMD condition). Participants were requested to return to a fixed starting point after picking up or dropping each magnet, in order to reset the experiment in between each pick-and-drop operation. After the tasks were completed for both conditions, the user was asked to complete a short questionnaire with 5-point likert scale questions polling their experiences completing the task and using the system. The experimental setup was randomized to balance conditions.

Results

User Performance

We measured the completion time for single pick-and-drop operations for each participant. In order to calculate the time needed for a participant to grab or drop a magnet, we annotated the video of the experiment and extracted the completion time for each pick and drop operation in both conditions. Start and stop time for each operation was determined by the entrance/exit of the hand into the video frame captured by the camera shown in Figure 2C. Three of the ten participants at times used both their hands to move the magnets. Those data samples were removed. After removing outliers the sample size of the HMD condition was 138 while we had 137 pick-and-drop samples for laser pointer. For the HMD condition, the average time for a pick-and-drop operation was about 0.81 seconds with a standard deviation of 0.23. For the laser pointer condition, the average completion time was 0.77 seconds for each operation, with a standard deviation of 0.16.

The statistical t-test indicated that the pick-and-drop completion time in the laser pointer condition is significantly less than task completion time in the HMD condition ($P = .064$), confidence interval 90 percent.

Questionnaire

8 out 10 participants preferred using the laser pointer over the HMD, as they argued that using the HMD was significantly more tiring for their eyes (HMD $\bar{X} = 4$, $\sigma = 0.81$ see Table 1) than using the laser pointer (laser $\bar{X} = 1.5$, $\sigma = 0.52$). Completing the task was perceived as slightly easier using the laser pointer ($\bar{X} = 4.3$, $\sigma = 0.48$) than the HMD ($\bar{X} = 3.4$, $\sigma = 0.85$). Finally, participants argued that the visibility of the indicator was higher in the laser pointer condition ($\bar{X} = 4.5$, $\sigma = 0.70$) than the HMD (HMD $\bar{X} = 3.9$, $\sigma = 0.99$).

Table 1: The questionnaire results

Questions	min	$\bar{X}$	Max	σ
Completing task using HMD was easy	2	3.4	4	.85
Completing task using laser was easy	4	4.3	5	.48
Using laser pointer was, eye-tiring	1	1.5	2	.52
Using the HMD was eyes-tiring	3	4	5	.81
Indications on HMD were easy to see	2	3.9	5	.99
Indications by laser were easy to see	3	4.5	5	.70

Open Comments

Application ideas provided by participants included telemedicine, technical assistances for car repairment, guidance of art students to learn how to paint, or even remotely guided shopping.

Discussion and Conclusion

We have investigated the use of laser pointers as an alternative to HMDs for tele-guidance applications because 1) previous studies [11, 8] have reported on a number of challenges connected to HMDs such as focusing problems, eye fatigue and etc.; 2) no previous adequate comparative study could be found. The results of our experiment showed that laser pointer solutions can perform better than HMDs for simple tele-pointing tasks ($P = .064$; confidence interval 90percent). During our experiment, participants needed to switch only once between the digital image shown on the HMD to the surrounding physical world. For tasks with higher frequency of focus shifts, we expect a higher difference between two conditions; however, the complexity and amount of information that can be displayed by laser pointer is still much less than HMDs. Such an example would be the case of remote guidance during surgery (future investigation), in which the surgeon in the HMD condition would need to keep looking at the patients internal tissues, switching from digital view on the HMD to the real world and vice versa. Moreover, the visibility of the laser point depends on many factors such as lighting condition, distance, color, and texture of the projected surface which is a limitation for the laser pointing approach. More empirical studies are needed to determine the strengths and weaknesses of both pointing approaches given certain application contexts. For future work we intend to design a more complex experiment to further investigate the performance of the two approaches and also add performance accuracy to the set of measured parameters. Both systems will also receive a pointing stabilization component in order to become directly useful in real world tasks outside the lab.

Acknowledgements

This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

References

[1] Alem, L., Huang, W., and Tecchia, F. Supporting the changing roles of maintenance operators in mining: A human factors perspective. *Ergonomics Open Journal 4* (2011), 81–92.

[2] Chen, S., Chen, M., Kunz, A., Yantaç, A. E., Bergmark, M., Sundin, A., and Fjeld, M. Semarbeta: mobile sketch-gesture-video remote support for car drivers. In *Proceedings of the 4th Augmented Human International Conference*, ACM (2013), 69–76.

[3] Fussell, S. R., Setlock, L. D., Yang, J., Ou, J., Mauer, E., and Kramer, A. D. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction 19*, 3 (2004), 273–309.

[4] Harrison, C., Benko, H., and Wilson, A. D. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, ACM (2011), 441–450.

[5] Karim, R. A., Zakaria, N. F., Zulkifley, M. A., Mustafa, M. M., Sagap, I., and Latar, N. H. M. Telepointer technology in telemedicine: a review. *Biomedical engineering online 12*, 1 (2013), 21.

[6] Kirk, D., and Stanton Fraser, D. Comparing remote gesture technologies for supporting collaborative physical tasks. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2006), 1191–1200.

[7] Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, Y., and Heath, C. Dual ecologies of robot as communication media: thoughts on coordinating orientations and projectability. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2004), 183–190.

[8] Sakata12, N., Kurata12, T., and Kuzuoka, H. Visual assist with a laser pointer and wearable display for remote collaboration.

[9] Schwerdtfeger, B., Pustka, D., Hofhauser, A., and Klinker, G. Using laser projectors for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, ACM (2008), 134–137.

[10] Stevenson, D., Li, J., Smith, J., and Hutchins, M. A collaborative guidance case study. In *Proceedings of the ninth conference on Australasian user interface-Volume 76*, Australian Computer Society, Inc. (2008), 33–42.

[11] Wille, M., Wischniewski, S., Scholl, P. M., and Van Laerhoven, K. Comparing google glass with tablet-pc as guidance system for assembling tasks. In *Glass Eyewear Computers (GEC)*, IEEE Press, IEEE Press (Zurich, Switzerland, 06/2014 2014).

[12] Yamazaki, K., Yamazaki, A., Kuzuoka, H., Oyama, S., Kato, H., Suzuki, H., and Miki, H. Gesturelaser and gesturelaser car. In *ECSCW99*, Springer (1999), 239–258.

Paper 4: Touch-less Interaction with Medical Images

Title of Paper

Touch-less Interaction with Medical Images Using Hand & Foot Gestures

Authors:

Shahram Jalaliniya, Jeremiah Smith, Thomas Pederson, Miguel Sousa, Lars Büthe, Thomas Pederson

Published:

Proceedings of the ACM conference on Pervasive and ubiquitous computing adjunct publication (UbiComp '13 Adjunct), ACM, 2013.

Abstract

Sterility restrictions in surgical settings make touch-less interaction an interesting solution for surgeons to interact directly with digital images. The HCI community has already explored several methods for touch-less interaction including those based on camera-based gesture tracking and voice control. In this paper, we present a system for gesture-based interaction with medical images based on a single wristband sensor and capacitive floor sensors, allowing for hand and foot gesture input. The first limited evaluation of the system showed an acceptable level of accuracy for 12 different hand & foot gestures; also users found that our combined hand and foot based gestures are intuitive for providing input.

Touch-less Interaction with Medical Images Using Hand & Foot Gestures

Shahram Jalaliniya
IT University of Copenhagen
2300 København S, Denmark
jsha@itu.dk

Lars Büthe
ETH Zürich
Gloriastrasse 35, 8092 Zürich
lars.buethe@ife.ee.ethz.ch

Jeremiah Smith
Imperial College of London
London, SW7 2RH, UK
jeremiah.smith@imperial.ac.uk

Thomas Pederson
IT University of Copenhagen
2300 København S, Denmark
tped@itu.dk

Miguel Sousa
Future-Shape GmbH
Höhenkirchen-Siegertsbrunn
miguel.sousa@future-shape.com

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UbiComp '13, September 08 - 12 2013, Zurich, Switzerland
Copyright © 2013 ACM 978-1-4503-2215-7/13/09…$15.00.

Abstract
Sterility restrictions in surgical settings make touch-less interaction an interesting solution for surgeons to interact directly with digital images. The HCI community has already explored several methods for touch-less interaction including those based on camera-based gesture tracking and voice control. In this paper, we present a system for gesture-based interaction with medical images based on a single wristband sensor and capacitive floor sensors, allowing for hand and foot gesture input. The first limited evaluation of the system showed an acceptable level of accuracy for 12 different hand & foot gestures; also users found that our combined hand and foot based gestures are intuitive for providing input.

Author Keywords
Gesture-based interaction, Touch-less interaction in Hospital, Wearable sensor, Floor sensor

ACM Classification Keywords
H.5.2 User Interfaces: Interaction Styles; J.3 Life and Medical Sciences: Health

Introduction
Surgeons need to interact frequently with an increasing number of computerized medical systems before and during surgeries in order to review medical images and

records. However, computers and their peripherals are difficult to sterilize, so usually during a surgery, an assistant or nurse operates the mouse and keyboard for such interactions. This indirect interaction suffers from communication problems and misunderstandings [9]. That is one of the main reasons why, in recent years, touch-less interaction has been considered for use in operation theatres. In general, touch-less interaction has been implemented using vocal commands and body gestures. A limitation of voice-based methods is that they cannot usually distinguish between different people speaking in the same room, in addition to being sensitive to environmental noise [7]. As for gesture-based approaches, these systems can detect body gestures using different kinds of cameras or body-worn sensors. The main challenge of vision-based systems is low accuracy in difficult lighting conditions, difficulty in coping with cluttered backgrounds, and occlusions by staff or equipment [4].

In this paper, we designed and implemented a hand gesture-based system based on inertial body-worn sensors to interact with medical images in the operation room. We believe that such sensors could eventually be integrated into the surgeon's garment. We also implemented a foot gesture-based system using a capacitive floor sensor, as a clutch mechanism, to control interaction with different systems. Our approach is motivated by the fact that hand-based gesture interaction has been recognized as a highly natural way of interaction [16], and clinicians already use their foot to control certain medical devices in hospitals; therefore, foot-based gesture could potentially be a suitable input mechanism for interaction in the operation theater. We report on a first limited evaluation of the system both with respect to the accuracy of the hand & foot based gesture recognition system and on usability aspects of the system.

Related Work

Interaction, or more precisely "user input", is touch-less if it can happen without mechanical contact between any part of the system and human body [3]. In past touch-less interaction research, voice and gesture have been investigated by many researchers. Beyond these approaches, HCI researchers have also explored for example gaze tracking and brain-computer interfaces for touch-less interaction.

Voice input can be realized by recording the voice with a microphone and processing it through dedicated algorithms. To ensure a good functionality, training of the system usually is needed and special voice commands have to be remembered. As mentioned before, the main problem of this approach is the fact that the system will react to all people speaking in the room, especially if they have similar voices. This problem has been tackled by installing a microphone array instead of just a single microphone. The microphone array is able to detect where the speaker is standing and through this decide if the speech is an intended command [7]. The advantage of interacting through voice lies in the fact that it is independent of any occlusions, and the user does not need to wear any additional devices.

In general, body gestures could be detected in different ways from using wearable sensors to environmental sensors. The most common approaches for touch-less interaction in operation rooms fall into two main categories: detecting gestures either with a vision-based approach or with the help of body-worn sensors. In the vision-based approach, as with vocal interaction, the user does not need to wear any additional devices. However, a direct line of sight is needed for the interaction, where the users typically have to hold their hand in an unnatural position in order for the system to detect the gestures. The detection of the gestures can for instance be done with regular webcams [17], a stereo camera [6], a time of flight camera [15], or the Microsoft Kinect [5], [14]. The latter is becoming increasingly popular thanks to its low cost and easy implementation.

Body-worn sensors pose a good alternative to vision-based systems, as they do not require a direct line of sight. Furthermore, this type of sensor only allows a designated person to interact with the system, avoiding the potential confusion associated with having multiple people in the room the system is deployed in. A combination of inertial orientation sensors [4], gyroscope [19], and accelerometer are some of the most common body-worn sensors, which have been used to detect hand gestures for touch-less interaction in the hospital setting. For detecting foot gestures, several approaches can be used mainly divided into on-body sensors [1] and environmental sensors (i.e. sensitive floors). The latter category features different technologies including pressure [13], light refraction [2] or capacitive [18]. While capacitive approaches usually have less resolution, when compared to the others, these do not require for a soft or transparent surface, and easily allow for swipe-like foot gestures.

With body-worn sensors it is important to differentiate between gestures performed as intentional input the interactive system and gestures that take place as part of other activities (and should be ignored by the interactive system). One approach is to detect dedicated, easily distinguishable, gestures out of a continuous data stream [10] or enabling the system with another modality like user voice commands [4].

The main difference of our approach compared to previous work is that we are using a combination of hand and foot gestures to interact with several medical systems using both wearable and environment embedded sensors. Hand gesture commands have been designed for interacting with medical images, while foot gestures can enable, disable, and switch interaction between different systems.

Hand Gesture recognition method

Our hand-gesture detection is based on a wrist-worn IMU (Inertial Measurement Unit), which features a three-axis accelerometer, gyroscope and magnetometer, and transmits the data wirelessly, in real-time.

Modeling gestures

For interaction with the system a set of six well-distinguishable gestures have been identified with the aim of being intuitive for the user. During interaction, the initial state of the right arm is in a 90-degree angle. All hand gestures are depicted in Figure. 1. Apart from four regular movements "up" (1), "down" (2), "left" (3) and "right" (4) that are used for any menu-like navigation, an additional two gestures are implemented, i.e. tilting the hand to the left (5) and the right (6). These are used for special actions, such as a zoom-in or zoom out.

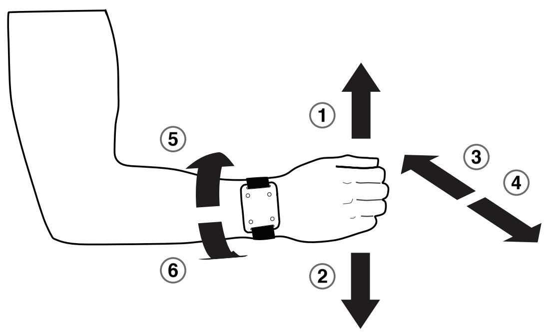


Figure 1. Hand gestures

The inertial sensor is attached, as shown in Figure 1, to the outer side of the right wrist in order to ensure that all movements are detected with maximum amplitude by the device while at the same time the hand and fingers are not obstructed.

Learning gestures
We had two different options for training our system to detect hand gestures: a person independent learner or training system for each individual. The latter strategy increases the accuracy of the system while the system needs to be calibrated for each user separately, which is time consuming. Since our hand gestures are simple and easy to learn, we followed the person independent strategy. For each gesture a training set of around 20 gestures is recorded and labeled. These are used as the input for a learner component, which is trained on the specific gestures. The trained learner is then used for the online gesture recognition.

Gesture recognition
The six gestures we chose for our system were best recognized using a two-stage strategy with a set of single-output artificial neural networks. The first stage used a single neural network to discriminate between the *idle* class and any of one the *gesture* classes. If a gesture class was detected in the first stage, a further set of 6 neural networks was used to predict the exact gesture class. This was done using 6 neural networks each trained to recognize one of the target gestures. The overall predicted gesture class was chosen to be the one for which the network output score was closest to the value assigned during training to its target class. All neural networks used default Matlab settings with a single hidden layer of size 4.

Features were computed from the ETHOS IMU [8] using a moving sample window of size 25, displaced by 9 instances between each collection (FIFO with 64% overlap between samples). Samples are drawn at ~21hz with about 2% error due to missing packets while transmitting the data via the Ant protocol.

The features extracted from each sample window were the following:

(Var(acc.x), Var(acc.y), Var(acc.z), Var(gyr.x), Var(gyr.y), Var(gyr.z), dGyr.x, dGyr.y, dGyr.z)

Where *Var()* stands for the variance measure taken over *acc* (accelerometer) and *gyr* (gyroscope) signals in the sample window. *dGyr* signals whether a gyroscope signal initially increases or decreases with respect to the signal baseline of the sample window. This was marked by a +1 or -1 respectively in the corresponding feature. The axes are specified after the '.' of each signal source. We note that the neural network responsible for discriminating between the idle and gesture class only used the first six features while the 6 gesture-specific neural networks used all 9. The features were obtained by analyzing raw IMU signals, observing that each gesture had a characteristic gyroscope signal signature.

Foot gesture recognition method

Foot gestures are used for two main purposes in the system. In the first case, these serve as an enabler for user interaction with the displays. That is, different foot gestures are used to activate/deactivate the hand gesture detection sub-system. This allows the user to *freeze* the displays with a desired view on the display and start performing a given real-world task that requires hand movements, which otherwise would trigger false (unintentional) gesture commands.

The second case concerns switching between the multiple screens. This task could be accomplished by a more *exotic* hand gesture (such as circling) or a combination of the simple gestures (e.g., three times *gesture_right*). However, the first case requires movements that may not feel natural to the user, whilst in the second one, the user needs to memorize certain gesture combinations, in addition to taking more time to perform the switching task.

Six foot gestures are implemented in total. These are divided into two types: *tapping* and *swiping*. The detection of the latter type is eased by the use of a proximity sensitive floor (vs. pressure sensitive). Furthermore, both left and right feet can be used for interaction. The complete gesture set consists of: *swipe_left* [Figure 2, (1)], *swipe_right* [Figure 2, (2)], *double_tap_left*, *double_tap_right* [Figure 2, (3)], *swipe_down_left*, *swipe_down_right* [Figure 2, (4)]. Detection of the defined gestures on the floor is straightforward. Each of the triangular-shaped sensor cells is considered active or inactive (*ON/OFF*) depending on the proximity of the foot. As such, a *double-tap* is simply translated into a sequence of *ON-OFF-ON-OFF-ON* on a given cell (within a specific timing). The *(x,y)*-location of the cell encodes which foot was used for the tapping. As for the swipe movements, a set of *OFF-ON* sequences of neighboring cells is used, with the swipe direction depending on which cells are activated in a row.

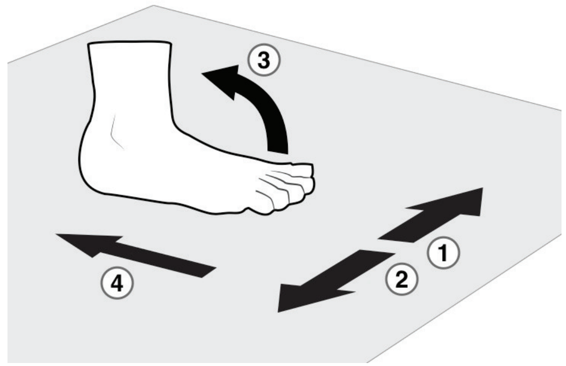


Figure 2. Foot gestures

System architecture

A schematic representation of the main components of the implemented system and the interconnection between different components is illustrated in Figure 3. The system consists of a wristband-style inertial sensor, a capacitive floor sensor, and the gesture recognition system. The inertial sensor and the floor sensor transmit data to the gesture recognition system wirelessly.

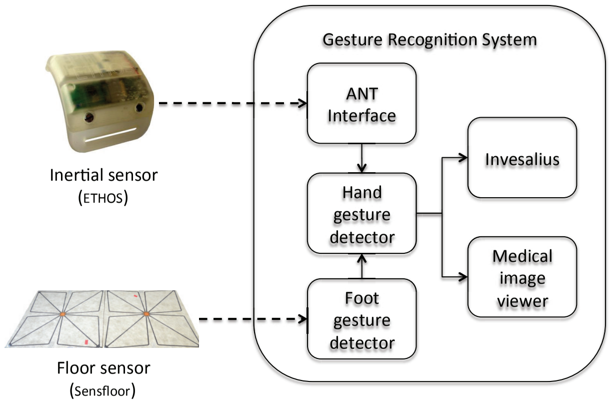


Figure 3. System Architecture

Inertial orientation sensor
In order to detect hand movements, we used ETHOS, which is a miniaturized IMU sensor developed at ETH Zurich for unobtrusive measurement of body movement. The ETHOS measures 3-D acceleration, 3-D rate of turn, 3-D magnetic field. The ANT+ module of the ETHOS provides wireless connection to transmit data to the server.

Capacitive floor sensor
SensFloor [11] is a textile-based underlay with embedded electronic modules equipped with radio transceivers. The floor uses a capacitive measurement approach to detect conductive objects, such as a *human foot*, placed upon itself (or hovering up to a few centimeters). A typical SensFloor unit (0.5x0.5m) has one electronic module with eight surrounding triangular sensor pads (Figure 4).

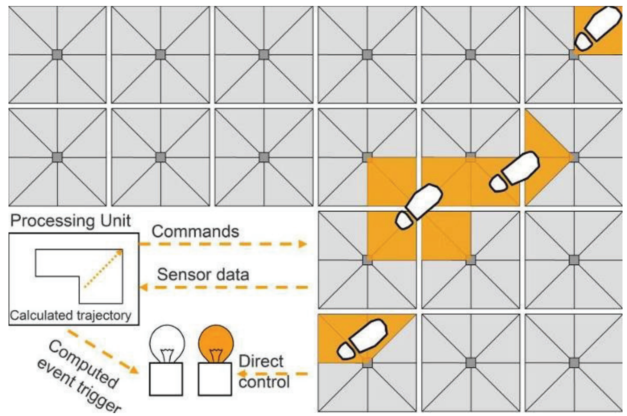


Figure 4. Floor sensor

Each time a significant change in capacitance is detected in one of the sensor pads, the module sends a wireless message to a central receiver, designated as *Smart Adapter*. Such message contains the (x,y) position of the module and the measurement data.

Gesture Recognition System
The entire system is running on a laptop with a dual core 1.4 GHz CPU and 4 GB RAM, running Windows 7. Matlab is used for processing the acquired data and interfacing with visualization applications. The laptop is connected to two screens that are used for displaying the image viewer as well as the InVesalius. Both applications are stand-alone and are controlled with commands received through TCP-IP. All parts of the system are described in more detail in the following subsections.

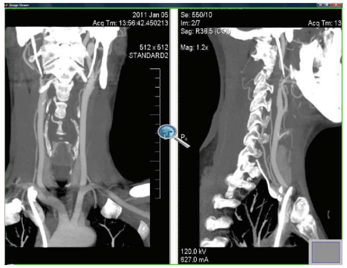


Figure 5. Screenshot of the Medical image viewer system including the magnifier symbol and colored border as feedback mechanisms of the system

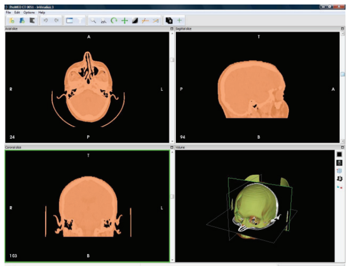


Figure 6. Screenshot of the Invesalius system

ANT INTERFACE MODULE
In order to connect the ETHOS to the gesture recognition system wirelessly, we developed an interface application, which receives the IMU data through the ANT protocol. This application corrects for any lost data and writes the received data on a virtual COM port, which is accessible by the Matlab application.

FOOT GESTURE RECOGNIZER
The sensor data transmitted from the floor modules is received by a Smart Adapter, which streams it to a COM port on the PC. A Matlab application was implemented to read and process such data, in order to detect the relevant gestures (see Figure 2). Each time a foot gesture is detected, the state of the system is changed. Specifically, the application encodes whether a screen is active, and which screen is currently in focus. The foot gesture recognizer gives a vocal feedback to the user after detecting a gesture, which is helpful for users to repeat a non-detected gesture immediately. Without this feedback, the user needs to wait for a longer time to see a visual feedback from one of the two displays.

HAND GESTURE RECOGNIZER
The data stream from the IMU input into the classifier, which processes the data and returns the predicted gesture. In order to avoid that one real gesture is detected as two subsequent gestures by the classifier, a gesture is only output when the previous three windows did not identify any gesture at all.

MEDICAL IMAGE VIEWER
Two different applications are used to demonstrate the user's interaction with the system. One consists of an image viewer, where the user can flip through medical pictures, as well as zoom in, zoom out, and pan in all directions (Figure 5). We used the color of the window around the image as a feedback to show if the image viewer application is activated or not. The green color illustrates the activated status while the gray window means it is inactive. When the user sends a command by performing a specific gesture the image viewer displays appropriate symbols on the screen for each action in addition to performing the action. (i.e. displaying a magnifier symbol for zooming in Figure 5). The other application is a public and open-source software for visualization of 3D medical images, *InVesalius* [12]. Here, the user can navigate through 2D slices of a CAT scan, in different views (Figure 6). When the InVesalius application is activated, the color of the activated window changes from black to green.

Evaluation
The main goal of our evaluation is answering to the following questions: To what degree does our combined hand and foot gesture recognition system correctly identify the commands from users of the system? Do users find the proposed combination of hand and foot gestures a viable alternative to the traditional indirect control of visualizations in operation theaters?

Evaluation of the classifier
The overall gesture recognition system was evaluated in two different ways. We first used a 3-fold cross-validation experiment over a recorded dataset from one of the authors and then confirmed the results in our user study where we recorded the live performance of the system over 5 other users. In both cases, the learner component was oblivious as to whether the data was coming from our recorded dataset or from a live sensor stream.

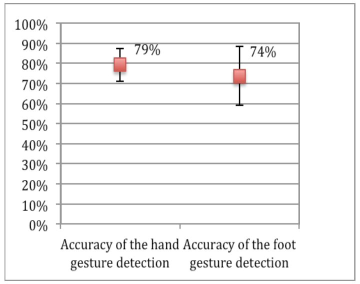


Figure 8. Accuracy of gesture detection in the user study

RESULTS FOR THE 3-FOLD CROSS-VALIDATION EXPERIMENT

In this experiment, we have a dataset of approximately 350 gestures, equally distributed among the 6 classes. Because we are streaming the data to the learner, a change in the sample window size results in a change in the number of points (and their features) received by the learner. We chose to segment the data into 3 roughly equal sized sub-datasets, i.e. 3 data collection runs taken over 3 different days, which determined the parameter for our cross-validation experiment. We report the confusion matrix in Figure 7 and precision and recall in Table1.

Gestures	Up	Down	Left	Right	Tilt_L	Tilt_R	Idle
Precision	0.91	0.93	0.96	1	0.95	0.98	0.99
Recall	1	0.98	1	1	1	1	0.99

Table 1. Precision and recall of the classifier

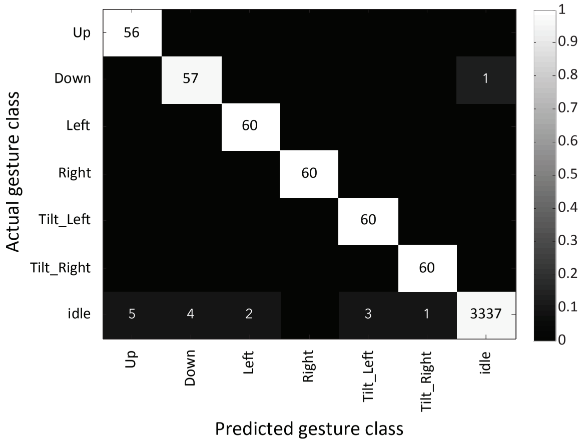


Figure 7. Gesture classification confusion matrix of the cross-validation experiment

The performance on the recorded data shows that the recall rate through classes is very high (low number of deletions), with a single missed 'down-gesture' throughout the experiment. The precision is slightly lower as some idle classes have been falsely predicted as gestures (insertions).

Evaluation of usability

In order to evaluate the usability and intuitiveness of the implemented system we conducted a qualitative user study in our (non-clinical) research lab with five participants (two females and three males), mainly computer science researchers (Figure 10). After introducing the whole setting and a short training on how to do the gestures, we asked users to try the system for about one minute, and then they were asked to go through a predefined scenario including ten steps which took in average three and half a minutes for each participant. Finally, we had a short interview with participants and they were asked to answer a list of seven questions with a five-point Likert scale that ranged from strong disagreement (1) to strong agreement (5). The result of the questionnaire (Figure 9) shows that in general users found that the combination of the foot and hand gestures is a valuable alternative for the indirect interaction. However during the evaluation, the average accuracy of the foot gesture detection was about 74% and this value was 79% for the hand gesture detection (Figure 8). In order to obtain the ground truth, we recorded the video of the participants during the evaluation, and the accuracy of the gesture detection has been calculated based on reviewing the recorded videos.

Figure 10. A user interacting with the system

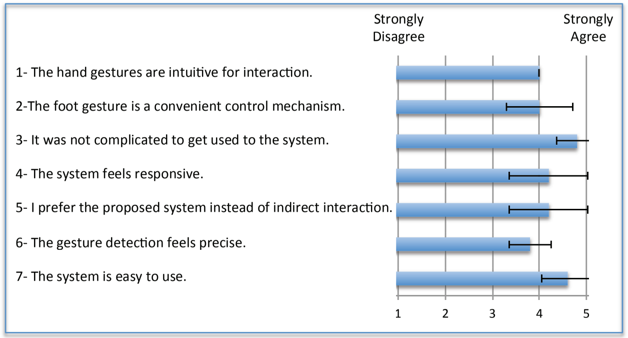


Figure 9. Result of the usability evaluation

Discussion & Conclusions

As the previous studies [4] also showed, using inertial body-worn sensors for interaction in the operation room gives us the possibility of interacting with several displays and distributing interaction between different users while it is independent of the user location. Moreover, since we are using a finger-free method surgeons can interact with the system when they have surgical instruments in their hands.

The combination of hand and foot gestures helped us implement several gestures (6 hand gestures with just one wristband inertial sensor and 6 foot gestures) with a high accuracy of detection. In addition from our limited user study, we can argue that using our foot gesture detection mechanism could be an intuitive and accurate alternative for other mechanisms of controlling interaction such as user's voice. However, our users found some of the foot gestures more intuitive. For example, the double tap gesture was preferable to the swipe gestures because of the user's balance problem during the swipe gesture.

Our gesture detection method showed a high accuracy in theory; however, we observed some gesture confusion during the user study. Because our hand gesture recognition system was calibration-free, which means we used a fixed training dataset sampled by one person for all participants. But each participant did the gestures slightly different that decreased the accuracy of classification. In the further steps, we expect that by adding more training data from different people or calibrating the system for each user, the accuracy of the system would be improved. Also for the foot gesture recognition, we observed some none-detected gestures since we used a capacitive floor sensor, and our users wore dissimilar shoes with different thicknesses. In order to avoid wrong positive gestures, we set the threshold of the capacitive sensor on the minimum value; however, for a real application, this value could be adjusted based on the standard shoes in the operation theatre. As it is clear in the Figure 10, in our current implementation we used a small piece of floor sensor contacting two cells that limited users to stand on a specific position to perform foot gestures. But as a future step, it is possible to detect gestures independent of the user's location on a bigger SensFloor. In the next step, we will ask real surgeons to evaluate our system for a specific surgery.

ACKNOWLEDGEMENT

This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

References

[1] Alexander, J., Han, T., Judd, W., Irani, P., Subramanian, S. Putting Your Best Foot Forward: Investigating Real-World Mappings for Foot-based Gestures. In Proc. *CHI'12*, 1229-1238.

[2] Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff., T., Becker, T., Holz, C. and Baudisch, P. Multitoe: High-Precision Interaction with Back-Projected Floors Based on High-Resolution Multi-Touch Input. *UIST '10*, 209-218.

[3] Barré, R., Chojecki, P., Leiner, U., Mühlbach, L., & Ruschin, D. Touchless Interaction- Novel Chances and Challenges. In Proc. of the *13th International Conf. on Human-Computer Interaction, Part II*, *San Diego, CA, USA, 2009*, 161-169.

[4] Bigdelou, A., Schwarz, L., Navab, N. An adaptive solution for intra-operative gesture-based human-machine interaction. In Proc. Of the *ACM international IUI '12 Conf.*, *2012, p. 75.*

[5] Ebert, L.C., Hatch, G., Ampanozi, G., Thali, M. J., Ross, S. "You Can"t Touch This: Touch-free Navigation Through Radiological Images", *SAGE*, 2011.

[6] Graetzel, C., Grange, S., Fong, T., and Baur, C., A Non-Contact Mouse for Surgeon-Computer Interaction, NCCR-COME Research Networking Workshop, Brauwald, Switzerland, August 2003.

[7] Hands-free interaction in the hospital. Retrieved April 29,2013 from: http://www.healthcare.philips.com/pwc_hc/main/about/assets/Docs/medicamundi/mm_vol49_no3/mm_493_Technology%20News.pdf

[8] Harms, H., Amft, O., Winkler, R., Schumm, J., Kusserow, M., Troester, G., ETHOS: Miniature orientation sensor for wearable human motion analysis, Proceedings of *IEEE Sensors Conf.*, 2010, 1037-1042.

[9] Johnson, R., O'Hara, K., Sellen, A., Cousins, C., and Criminisi, A. Exploring the potential for touchless interaction in image-guided interventional radiology. In Proc. *SIGCHI Conf. CHI,* 2011, 3323-3332.

[10] Junker, H., Amft, O., Lukowicz, P., Troster, G. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recogn.* 41, 6, June 2008, 2010-2024.

[11] Lauterbach, C., Steinhage, A. and Techmer, A., Large-area wireless sensor system based on smart textiles, *9th International Multi-Conf. on Systems*, Signals and Devices, 2012, 1-2.

[12] Martins, T.A.C.P., Santa Ba´rbara, A., Silva, G.B., Faria, T.V., Cassaro, B.,Silva, J.V.L., InVesalius: Three-dimensional medical reconstruction software. In Proc. of the *3rd International Conf. on Advanced Research in Virtual and Rapid Prototyping*, 2008, 135-141.

[13] Paradiso, J., Abler, C., Hsiao, K., Reynolds, M. The Magic Carpet - Physical Sensing for Immersive Environments. In Proc. *CHI' 97*, 277-278.

[14] Ruppert, G. C. S., Reis, L. O., Amorim, P. H. J., de Moraes, T. F., & da Silva, J. V. L. Touchless gesture user interface for interactive image visualization in urological surgery. *World Journal of Urology*, 30(5), 2012, 687-691.

[15] Soutschek, S., Penne, J., Hornegger, J., Kornhuber, J. 3-D gesture-based scene navigation in medical imaging applications using time-of-flight cameras, in: Workshop on Time-of-Flight based Computer Vision, Anchorage, Alaska, June, 2008.

[16] Varona, J., Jaume-i-Capó, A., Gonzàlez, J., Perales, F.J. Toward natural interaction through visual recognition of body gestures in real-time. *Interacting with Computers* 21(1), 2008, 3–10.

[17] Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., Handler, J. Gestix: A Doctor-Computer Sterile Gesture Interface for Dynamic Environments, *Soft Computing in Industrial Applications* vol. 39, Springer Berlin / Heidelberg, 2007,30-39.

[18] Zimmerman, T. G., Smith, J. R., Paradiso, J. A., Allport, D., Gershenfeld, N. Applying electric field sensing to human-computer interfaces, In Proc. *SIGCHI Conf. CHI*, Denver, CO, 1995, 280–287.

[19] Zinnen, A., Schiele, B., Ziegert, T. Browsing patient records during ward rounds with a body worn gyroscope. In Proc. of the *ISWC '07*, 1-2.

Paper 5: EyeDroid: An Open Source Gaze Tracker on Android

Title of Paper

EyeDroid: an open source mobile gaze tracker on Android for eyewear computers

Authors:

Shahram Jalaliniya, Diako Mardanbegi, Ioannis Sintos, Daniel Garcia Garcia

Published:

Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers (UbiComp/ISWC'15 Adjunct), ACM, 2015

Abstract

In this paper we report on development and evaluation of a video-based mobile gaze tracker for eyewear computers. Unlike most of the previous work, our system performs all its processing workload on an Android device and sends the coordinates of the gaze point to an eyewear device through wireless connection. We propose a lightweight software architecture for Android to increase the efficiency of image processing needed for eye tracking. The evaluation of the system indicated an accuracy of 1.06 degrees and a battery lifetime of approximate 4.5 hours.

EyeDroid: An Open Source Mobile Gaze Tracker on Android for Eyewear Computers

Shahram Jalaliniya
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S, Denmark
jsha@itu.dk

Daniel Garcia Garcia
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S, Denmark
dgac@itu.dk

Diako Mardanbegi
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S, Denmark
dima@itu.dk

Ioannis Sintos
IT University of Copenhagen
Rued Langgaards Vej 7
2300 Copenhagen S, Denmark
isin@itu.dk

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
Ubicomp/ISWC '15 Adjunct, September 07-11, 2015, Osaka, Japan
©2015 ACM. ISBN 978-1-4503-3575-1/15/09...$15.00
DOI: http://dx.doi.org/10.1145/2800835.2804336

Abstract
In this paper we report on development and evaluation of a video-based mobile gaze tracker for eyewear computers. Unlike most of the previous work, our system performs all its processing workload on an Android device and sends the coordinates of the gaze point to an eyewear device through wireless connection. We propose a lightweight software architecture for Android to increase the efficiency of image processing needed for eye tracking. The evaluation of the system indicated an accuracy of 1.06 degrees and a battery lifetime of approximate 4.5 hours.

Author Keywords
Gaze tracking; Eyewear computer; Android; Google Glass

ACM Classification Keywords
H.5.2. [Information interfaces and presentation: User Interfaces]: Input devices and strategies

Introduction
By emerging new generation of unobtrusive eyewear computers, such as Google Glass[1] and Vuzix smart glass[2], it seems feasible that eventually these eyewear devices play role in everyday tasks. Due to the special form factor of eyewear devices, the delay between intention and action

[1] $https://developers.google.com/glass/$
[2] $http://www.vuzix.com/consumer/products_m100/$

is very short compared to other mobile devices[10]. This opens new opportunities for eyewear computers to be used more on the move and in parallel with real world tasks. However, mobile interaction with eyewear devices is still challenging. For example, in a mobile scenario, sometimes hands of the user are busy with a manual activity, or the user might be doing a visually demanding task. This means the eyewear device needs to provide several channels for interaction to support users in different situations. State of the art eyewear devices already support for head gesture input, voice commands, and touch-based gestures. Eye gaze has also been studied as an input modality for head-mounted displays [3]. However, due to the technical limitations, gaze-based interaction is not still supported by state of the art eyewear devices. One of the main challenges is the fact that the image processing required for gaze tracking is extremely complex and power demanding. Unfortunately, this computational demand is very far from what can be accomplished on existing eyewear devices such as Google Glass. In this paper, we investigate the possibility of using an Android smartphone to process eye images captured by a head-mounted camera to calculate the gaze coordinates for an eyewear computer.

Related Work

Most of the recent mobile gaze trackers use a laptop in user's backpack [5, 6] or a remote computer [9, 1] to analyze the eye image and calculate gaze coordinate. The dependency of gaze tracking systems to a local or remote computer decreases the mobility of users. There are also some commercial products from companies such as EyeTribe[3], Tobii[4], and Umoove[5] which support eye tracking on mobile and wearable devices. But the commercial mobile gaze trackers are usually so expensive and hard to afford. That is the reason why some of the recent studies have tried to use cheap and small processors such as Raspberry Pi [2] and micro-controllers [7] for eye tracking. Ferhat et.at [2] have presented a cheap eye tracking solution running on a Raspberry Pi device. They based their work on the open source Opengazer [8]. The average gaze estimation error of their system is about $1.4°$ for an image size of 640 × 480 pixels with the frame rate of 3Hz. Although their system was running on a small device, it was only tested on a stationary setup for gaze tracking on a computer screen. A more relevant work to our paper is the iShadow eye tracker by Mayberry et.al [7] that focuses on head-mounted gaze tracking. They have presented a fully mobile eye tracking solution using a low-power ARM Cortex M3 micro-controller.

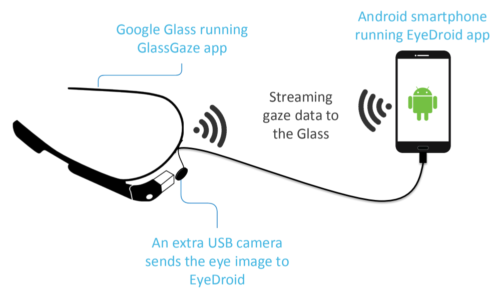


Figure 1: A schematic view of the system Architecture.

[3] $https://theeyetribe.com/$
[4] $http://www.tobii.com/$
[5] $http://www.umoove.me/$

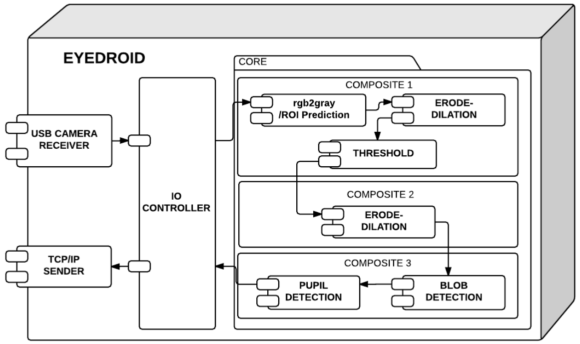


Figure 2: EyeDroid software arcgitecture. Eye tracking algorithm inside the core is decomposed into steps (filters) and connected by pipes (arrows). Each composite is executed on a separate thread.

The focus of the iShadow system was mainly implementing a very efficient video-based eye tracking approach that can run on a small micro-processor. They achieved real-time gaze tracking in an image captured by a front-view camera and they have reported an error of about 3 degrees for their system. Since in eyewear computers the display size is very small (less than 15 degrees), the accuracy of the eye tracker should be higher to provide a graceful interaction. In this paper, to achieve a higher accuracy in eye tracking (about 1 degree), we rely on the processing capacity of commonly used mobile devices. The proposed eye tracker on mobile device is an open-source affordable solution for gaze tracking on eyewear computers.

System Architecture

The proposed system comprises two main components: (1) our gaze tracking application (EyeDroid) running as a server on an android smartphone, and (2) the client application on Google Glass (GlassGaze)[6]. A schematic view of the system architecture is represented in Figure 1.

EyeDroid: Gaze Tracker Server on Android

Hardware

The hardware requirements in the current implementation of the EyeDroid eye tracker are an Android mobile device (minimum API level 15) and a head mounted USB 2.0 infrared camera connected directly to the Android phone through a USB cable. The first hardware prototype of the EyeDroid is shown in Figure 5. The recommended camera resolution is 640×480 pixels. Because the Android platform does not provide support to connect an external USB camera, the operating system needs to own root access to the phone and use customized camera video drivers. To develop the EyeDroid gaze tracker, open source third party drivers are used [4].

Software Architecture

The software architecture of the EyeDroid application is designed based on pipes and filters design pattern. This architecture helped us test different algorithm configurations easily during system development. Also in the EyeDroid software platform we built and used Java Lightweight Processing Framework (JLPF)[7] (Figure 3) as an external library. This design allows for a fully configurable algorithm in terms of decomposability and scheduling of the steps for execution on the available processing resources, instead of a monolithic algorithm that would perform poorly. Finally, in order to divide the algorithm in steps of equal execution time, the composite pattern was implemented to allow composition of individual steps (see Figure 2). Since performance is a

[6] *https://github.com/centosGit/GlassGaze*
[7] *https://github.com/centosGit/JLPF*

critical issue, we used the Android NDK support for C++ instead of the regular Android SDK for java. This allowed the algorithm code to run directly on the processing resources and access system libraries directly, unlike Java which would run on a virtual machine.

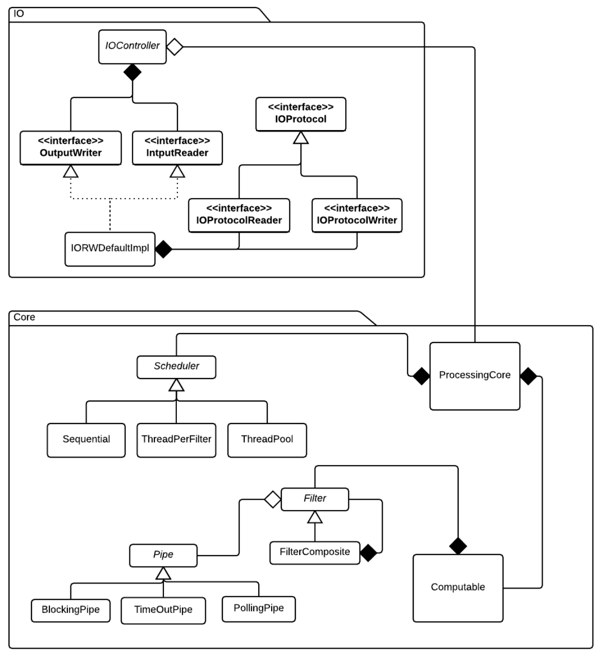


Figure 3: Java Lightweight Processing Framework (JLPF) software architecture

Gaze tracking method
In order to achieve real-time image processing, we have skipped detecting cornea reflection in the image which could compensate for the small movements of the camera relative to the eye. Only the pupil center obtained from the eye image is used for gaze estimation. Pupil detection is done by applying a simple blob detection algorithm on the eye image as follow: (step 1) The image is first converted to grey-scale (step 2) and then a morphological operation is done on the resulting image. (step 3) Then a threshold was applied at a constant level of around 70. (step 4) After thresholding, a morphological operation is done on the image before applying blob detection (4). To reduce the computation time, in each frame, we have defined a region of interest (ROI) for which the image processing is applied for. In the first frame, the ROI will be defined as the entire image. Once the pupil is found on previous processed frames, the ROI is reduced to 30% of the image size and is moved to the most recently computed pupil coordinates (the last frame whose processing is completed).

Calibration
In the current implementation of the system, a homography transformation is used as our gaze mapping function. The mapping function is obtained from a calibration process consisting of a minimum of 4 calibration markers on the display.

GlassGaze: Gaze Tracker Client on Google Glass

GlassGaze [1] is an android app developed for Google Glass that was originally developed to work as a client for the open-source Haytham gaze tracker [1]. GlassGaze provides a convenient user interface that can be controlled by voice and finger gesture. This client has an Android background service that receives the gaze data from the server and allows different applications, to subscribe to its messages. This background service also allows applications on Glass to communicate with the gaze tracking server. By applying the same messaging protocol defined in the GlassGaze we could easily use this app as our client.

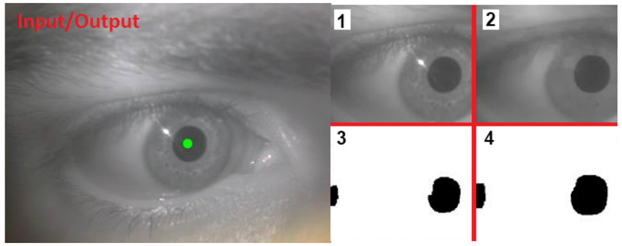


Figure 4: Pupil detection steps

System Performance

Accuracy of the Gaze Tracker

To measure the accuracy of our gaze tracker we conducted an experiment with 10 participants recruited among students from our university. Participants were asked to wear the Google Glass and run the GlassGaze application. First, we had a training session in which they tried the system for a while until they felt comfortable with the system. We started the experiment with a four-point calibration. After calibration, 15 markers were displayed randomly on the Google Glass display for one second. The participants were asked to look at the markers immediately after marker appearance (see Figure 5). We had 15 markers distributed evenly in 3 rows and 5 columns (Figure 6).

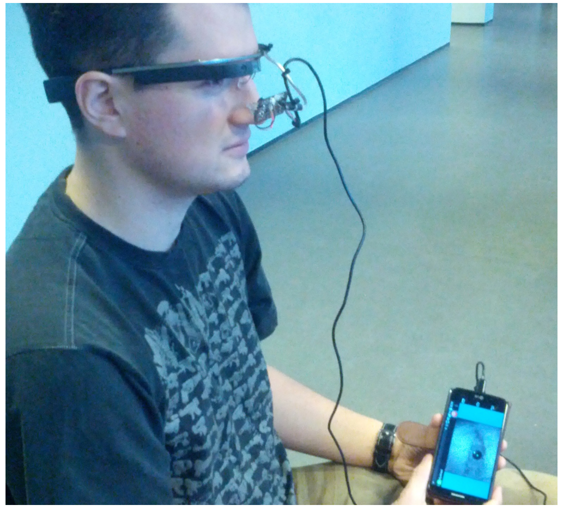


Figure 5: A participant perfomring the task.

To measure the gaze coordinate, the average coordinate of the gaze for the last 700 milliseconds of looking at each marker was calculated. The average of deviation from actual marker position was equal to 52.61 pixels with standard deviation of 35.26 pixels. This means that the error of our gaze tracking system is equal to 1.06 degrees. The distribution of the error for each marker is illustrated in Figure 6. The X and Y dimensions of the graph in Figure 6 represent two dimensions of the display on Google Glass (maximum 640 $\times$ 320), and the gray circles around each marker show the average error of the detected gaze point for each marker.

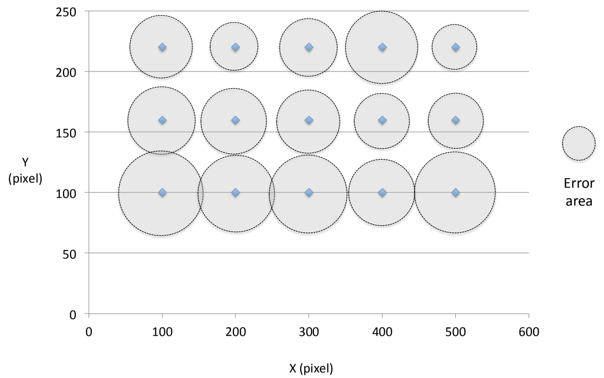


Figure 6: The average error of the gaze estimation for each marker (dots) is illustrated by circles around the markers

Battery Life
To calculate the battery life of the mobile device (a brand new LG-G2 smartphone with 2 GB RAM, a Quad-core 2.26 GHz Krait 400 processor, an Adreno 330 GPU and running Android 4.4 version) while running the EyeDroid application, we measured the charge of the battery (given by the Android built-in battery level indicator) every hour for 3 hours. To compensate the inaccuracy of the built-in indicator, the device was fully charged before conducting each experiment, any other apps were closed but default Android services, and the brightness of the screen was minimized. Each measurement was repeated three times and results were averaged.

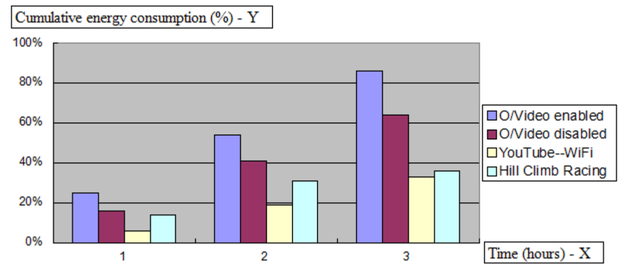


Figure 7: Comparison between EyeDroid and two other popular applications showing cumulative energy consumption (%) per hour

Since EyeDroid can optionally show the resulting coordinates drawn in top of the input video streamed on the device display, two different experiments were conducted. First in the video preview enabled mode and second when the preview mode is disabled. To have a baseline for our comparison, the battery life of the device running two other popular applications was measured in the same way: YouTube video streaming and Hill Climbing racing game. The results suggest that EyeDroid behaves similar to Hill Climbing game but deviating approximately 10% per hour. The maximum battery life estimation running EyeDroid with in preview-disabled-mode is approximately 4.5 hours.

Discussion & Conclusion
In this paper we presented a monocular mobile gaze tracker on Android smartphone to support gaze-based interaction with eyewear devices. We used an efficient and lightweight software architecture to divide image processing task into parallel threads. Using our approach, we reached to 6.41 fps performance in the image processing task. The experimental study showed the

accuracy of 1.06 degree for our gaze tracker. The error areas (gray zones) around each marker in the Figure 6 show that our gaze tracker can be used for interaction with Google Glass since users are able to accurately point to (at least) 15 different objects on the display. Although, the head gear was fixed on the head, small movements of the camera relative to the eye could create a relatively large error in the gaze tracking result. This was due to the fact that gaze mapping was using only pupil center. As future work, we will add glint detection to increase robustness of the system.

Acknowledgments

This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

References

[1] Haytham gaze tracker. http://itu.dk/research/eye/, June 2014.
[2] Ferhat, O., Vilarino, F., and Sánchez, F. A cheap portable eye-tracker solution for common setups. *Journal of Eye Movement Research 7*, 3 (2014), 2.
[3] Jalaliniya, S., Mardanbeigi, D., Pederson, T., and Hansen, D. Head and eye movement as pointing modalities for eyewear computers. In *BSN Workshops, 2014 11th International Conference on* (June 2014), 50–53.
[4] Lab, K. Usage of usb webcam with customized galaxy nexus (android 4.0.3). http://brain.cc.kogakuin.ac.jp/research/usb-e.html.
[5] Li, D., Babcock, J., and Parkhurst, D. J. openeyes: A low-cost head-mounted eye-tracking solution. In *Proceedings of the 2006 Symposium on Eye Tracking Research &Amp; Applications*, ETRA '06, ACM (New York, NY, USA, 2006), 95–100.
[6] Lukander, K., Jagadeesan, S., Chi, H., and Müller, K. Omg!: A new robust, wearable and affordable open source mobile gaze tracker. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, ACM (New York, NY, USA, 2013), 408–411.
[7] Mayberry, A., Hu, P., Marlin, B., Salthouse, C., and Ganesan, D. ishadow: Design of a wearable, real-time mobile gaze tracker. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '14, ACM (New York, NY, USA, 2014), 82–94.
[8] Nel, E., MacKay, D., Zieliński, P., Williams, O., and Cipolla, R. Opengazer: open-source gaze tracker for ordinary webcams.
[9] Rantanen, V., Vanhala, T., Tuisku, O., Niemenlehto, P., Verho, J., Surakka, V., Juhola, M., and Lekkala, J. A wearable, wireless gaze tracker with integrated selection command source for human x2010;computer interaction. *Information Technology in Biomedicine, IEEE Transactions on 15*, 5 (Sept 2011), 795–801.
[10] Starner, T. Project glass: An extension of the self. *Pervasive Computing, IEEE 12*, 2 (April 2013), 14–16.

Paper 6: Head and eye movement as pointing modalities

Title of Paper

Head and Eye Movement as Pointing Modalities for Eyewear Computers

Authors:

Shahram Jalaliniya, Diako Mardanbegi, Thomas Pederson, Dan Witzner Hansen

Published:

11th International Conference on Wearable and Implantable Body Sensor Networks Workshops (BSN Workshops), IEEE, 2014

Abstract

While the new generation of eyewear computers have increased expectations of a wearable computer, providing input to these devices is still challenging. Hand-held devices, voice commands, and hand gestures have already been explored to provide input to the wearable devices. In this paper, we examined using head and eye movements to point on a graphical user interface of a wearable computer. The performance of users in head and eye pointing has been compared with mouse pointing as a baseline method. The result of our experiment showed that the eye pointing is significantly faster than head or mouse pointing, however, our participants thought that the head pointing is more accurate and convenient.

2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops

Head and eye movement as pointing modalities for eyewear computers

Shahram Jalaliniya
IT University of Copenhagen
2300 Copenhagen S, Denmark
Email: jsha@itu.dk

Diako Mardanbeigi
IT University of Copenhagen
2300 Copenhagen S, Denmark
Email: dima@itu.dk

Thomas Pederson
IT University of Copenhagen
2300 Copenhagen S, Denmark
Email: tped@itu.dk

Dan Witzner Hansen
IT University of Copenhagen
2300 Copenhagen S, Denmark
Email: witzner@itu.dk

***Abstract*—While the new generation of eyewear computers have increased expectations of a wearable computer, providing input to these devices is still challenging. Hand-held devices, voice commands, and hand gestures have already been explored to provide input to the wearable devices. In this paper, we examined using head and eye movements to point on a graphical user interface of a wearable computer. The performance of users in head and eye pointing has been compared with mouse pointing as a baseline method. The result of our experiment showed that the eye pointing is significantly faster than head or mouse pointing; however, our participants thought that the head pointing is more accurate and convenient.**

Keywords*—*Gaze tracking, Eye pointing, Head tracking, Head pointing, Head-mounted display, Wearable computing.

I. INTRODUCTION

Advances in hardware and software technologies has resulted in developing new generation of eyewear computers, such as Google Glass and Vuzix smart glass, and it seems feasible that eventually these unobtrusive eyewear devices play role in everyday tasks. However, design for wearable devices is associated with a lot of known and unknown challenges. An important design challenge of interactive wearable computers is the need for novel interaction techniques since the classical WIMP desktop metaphor do not support users mobility. That is the reason why other alternative interaction metaphors, such as Personal Assistant [1], has been introduced for interactive wearable computers.

Head-mounted display is one of the main components of an eyewear computer which means the wearers of these devices need to interact with graphical user interfaces. Interaction with standard graphical user interfaces involves pointing at the object of interest and selecting the object. Some suggested interaction techniques that use hand for pointing are often inconvenient for mobile settings and sometimes require external devices like smart phones, joysticks or hand-held point-and-click devices [2]. Hand gestures recognized by a front-view camera or other wearable sensors have also been used as a mechanism for interaction with wearable devices [3]. However, for a mobile user, hands-free interaction is a big advantage in many situations.

Head and gaze-based input mechanisms are two modalities that can be useful for situations that prohibit the use of the hands, such as when the users' hands are disabled or occupied with other task. These two techniques are among the less-explored hands-free pointing mechanisms for the eyewear computers.

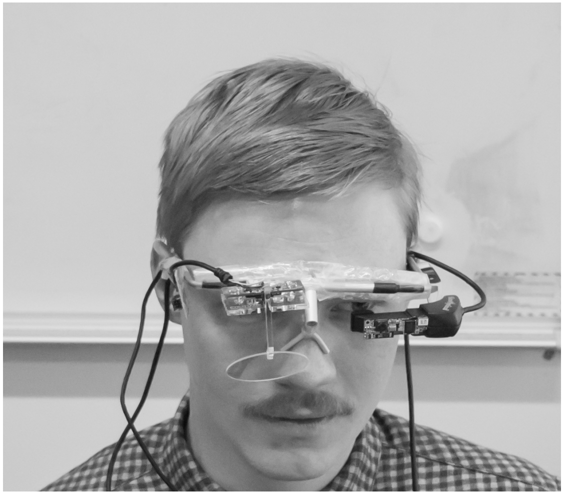


Fig. 1. A subject wearing the eyewear setup including the gaze tracker (tracking the right eye) and a HMD on the left eye. A head tracker sensor is mounted on the right side of the Glasses frame.

In this paper, we investigate the possibility of using head and gaze movements as a pointing mechanism for wearable computers through an experiment. In the experiment, participants were asked to point and select different targets on a head-mounted display by moving the head, gaze, or a mouse trackball.

II. RELATED WORK

Using the eye gaze as a source of input has long been a topic of interest in HCI and it is due to the fact that humans naturally tend to direct the eyes toward the target of interest [4], [5], [6], [7]. In fact, gaze pointing is one of the possible ways of pointing, and the typical use of gaze as a pointing mechanism is to control the cursor position on the screen. Gaze pointing has also been used for interaction with head-mounted displays [8], [9].

Head movement is another possible way of controlling the cursor on the screen and can be measured through a camera [10], [11] or other wearable sensors [12], [13]. Moreover,

978-1-4799-6136-8/14 $31.00 © 2014 IEEE
DOI 10.1109/BSN.Workshops.2014.14

Fig. 2. The hand-held finger mouse used in the experiment

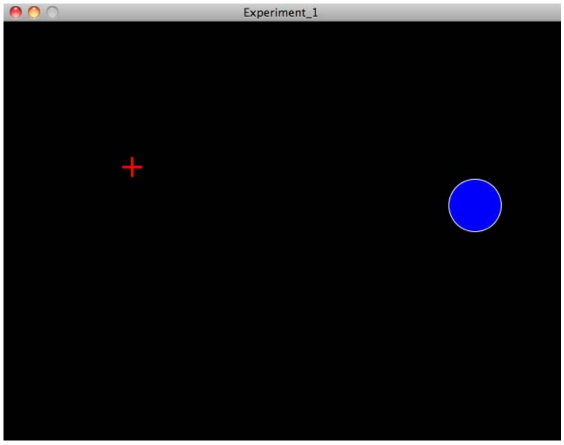


Fig. 3. A screen shot of the system: the blue circle is a 60 pixels width target, and the red cross is illustrating the pointer. By moving the pointer over the target the color of the target changes to red and the user clicks on the target.

head gesture has been used as an input modality in upcoming eyewear computers, such as Google Glass. But to the best of our knowledge, using head movement as a pointing modality is not investigated for wearable computers.

Bates and Istance [14] investigated the usability problems associated with eye and head-based pointing for direct manipulation on a standard graphical user interface. They compared the quality of interaction using these two input modalities during an interaction task.They found that an eye mouse is generally faster than head mouse and it could exceed the performance of a head mouse if target sizes were large and users sufficiently well practiced. While previous works have explored the eye and head pointing for stationary screens, the focus of our paper is to evaluate eye and head pointing for head-mounted displays (HMD).

III. Experiment Design

Given the known challenges of using WIMP desktop metaphor techniques for wearable computers, in this paper, we investigate whether gaze and head tracking methods can be used as a viable alternative to classical techniques such as mouse for pointing purpose. To answer this question we conducted an experiment to measure the user performance using a head-tracker, a gaze-tracker, and a mouse given the same pointing task.

IV. Method

A. Participants

8 participants (mean age = 32, no female) were recruited among local university students to participate in the experiment. Most of the participants were highly skilled computer users ($\bar{X} = 4.87$, $\sigma = .35$, where the range was 1 to 5), and all of them had perfect visual acuity. All participants were experienced hand mouse users; however, only two of them were used to use the finger hand-held mouse (see Fig. 2) . Also three participants had the experience of using gaze-tracker.

B. Apparatus

In order to examine head and gaze movement as an input modality for eyewear computers, we developed a wearable prototype including a gaze tracker, a head tracker, and a HMD connected to an ordinary MacBook laptop (see Fig. 1). The HMD was a MicroOptical SV-9 (640 $\times$ 480 pixels), and the head tracker sensor was a Sparkfun Razor (9DoF) including accelerometer, gyroscope, and magnetometer sensors. The size of the HMD was about $15 \times 10mm$ and the average distance between the eye and the HMD was about $35mm$. For tracking gaze, we used a custom-built hardware platform including a small infrared camera and a hot mirror reflecting the infrared light back to the camera to capture eye image. Gaze tracking is done by the open-source Heytham gaze tracker [15].

C. Procedure

The experiment started with a short introduction to the purpose of the experiment and the use of the apparatus. To keep the physical condition of the participants equal for all conditions, all of the participants were asked to wear the whole device during all conditions; however, they did not need to use all components in each condition. After participants were prepared for the experiment, they were asked to use the system until they felt comfortable. This usually took 2-3 minutes for each condition. Then each participant was asked to complete the task in three different conditions. The task was selecting the targets displayed on the HMD by using gaze-pointing, head-pointing, and mouse-pointing. The targets were blue circles with three different diameters of 60, 80, and 100 pixels displayed on a black background one after each other (see Fig. 3). The pointer was illustrated by a red plus which could be moved by moving head, eye, or trackball of the mouse. When the pointer was on the target the color of the target changed from blue to red and users had to click on the target. After the experiment, the users were asked to complete a short questionnaire with 5-point likert scale questions polling their experiences completing the task and using the system. The experimental setup was randomized to balance conditions and avoid the order effect. The conditions in which the task was completed were as follows:

1) Gaze pointing: Gaze tacker needed to be calibrated prior to the start of each trial. The calibration procedure required the user to look at 9 points shown in the HMD. After calibration, participants completed the task for 24 targets (8 instances of 3 different sizes) by moving the eye.

2) Head pointing: Before starting the task using head movements, the head tracker needed to be calibrated to set the starting position of the pointer in the center of the screen when the head of the user was in the neutral position (facing straight ahead). After a short warm-up trial, participants accomplished the task for the targets similar to the gaze pointing condition.

3) Pointing with the mouse: As a base-line condition, the participants were asked to complete the task using a hand-held finger mouse (Fig. 2) which is typically used for wearable computers.

D. Design

The experiment was an 8 $\times$ 3 within-subjects design, and each participant completed all above-mentioned conditions in one experimental session that lasted for approximately half an hour. Aside from training the amount of entry was: 8 participants $\times$ 3 conditions $\times$ 3 target size $\times$ 8 repetition = 576 trials.

V. Result

A. User performance

We recorded the task completion time and the number of wrong selections (errors) for each pointing and selecting task. The average speed of each pointing was calculated based on the distance between the departure point and the target divided by the task completion time.

In total, we had 192 samples for each condition. A one-way between subjects ANOVA was conducted to compare the pointing speed for each target size in different conditions. There was a significant difference at the $p < .05$ level for all 9 groups of trials (three conditions $\times$ three target widths) [$F(8, 567) = 12.005$, $p < .05$]. Also for each target group the user performance was significantly different in each condition: for the target size of 60 pixels [$F(2, 189) = 4.87$, $p = .008$], for the 80 pixels target [$F(2, 189) = 3.70$, $p = .02$], and for the 100 pixels target [$F(2, 189) = 6.71$, $p = .001$]. Post hoc comparisons using the t-test indicated that the pointing speed in gaze condition was significantly higher than head and mouse conditions for all targets (see Fig. 4). However, there was no significant difference between head pointing and using mouse.

As illustrated in the Fig. 5), the accuracy of the eye pointing was less than other conditions, but the statistical tests indicated no significant difference between error rates in the different conditions.

B. Questionnaire

From the questionnaire, 5 out 8 participants preferred using the head pointer over the eye-tracker and mouse since they found it easier to point with head ($\bar{X} = 4.12$, $\sigma = .99$) than eye pointing ($\bar{X} = 3.62$, $\sigma = .74$) and using mouse ($\bar{X} = 3.87$, $\sigma = 1.12$).

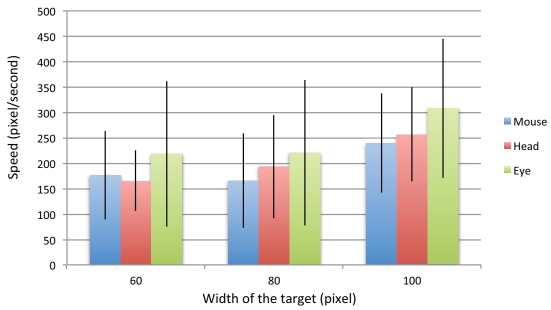


Fig. 4. The speed of completing the task for different sizes of the target.

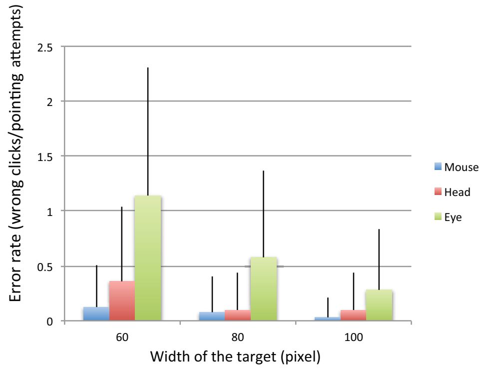


Fig. 5. The accuracy of completing the task for different sizes of the target.

VI. Discussion

In this paper we compared the performance of three different input modalities (mouse, gaze, and head) as a mechanism for pointing while interacting with a HMD. These three input mechanisms can be compared in terms of speed, accuracy and comfort. However, there are some differences between gaze pointing and head or mouse pointing that somehow makes it difficult to compare the gaze modality with the others. The main difference is the fact that the user can look at the target or the cursor while moving the cursor with either mouse or the head movements. This provides the user a visual feedback which is needed for the target acquisition. However, this is not the case with gaze pointing. With the gaze pointing the cursor always follows the user's gaze point and the user always sees the cursor at his/her fixation point. Therefore, the target acquisition can be very fast compared to the other pointing methods. Our experiment also showed the higher speed of the gaze pointing; however, we observed less accuracy in gaze pointing in comparison with head and mouse pointing. Since gaze tracking is inherently not a highly accurate pointing mechanism [16]. This means for eye-pointing to the small targets the user needs to keep looking at the target for a long time or correct the error by moving the eye which decreases the convenience of eye pointing method. The result of our questionnaire also indicated the lower user acceptance of the eye pointing compared to the other methods.

An important challenge of using camera-based gaze trackers for monocular head mounted displays is the fact that the pointer eye is mostly covered by the head mounted display, so that we need to track the other eye. In our experiment, the gaze-tracker was used in the right eye regardless of the dominant eye of the participants. This might also be another source of error for the gaze tracking approach.

Another critical issue with using gaze as an input modality for eyewear computers is that head-mounted eye trackers have the maximum accuracy of about 0.5 degrees and this limits how small the display can be and how small targets can be selected.

Regarding the low accuracy of the gaze trackers specially for small HMDs, one possibility to improve the user acceptance can be using the eye pointers for big targets on the graphical user interfaces.

Unlike the gaze tracking, the head pointing is a relatively stable approach for pointing (see fig. 5). That is probably why most of our users preferred the head pointing method to other approaches. Another advantage of using head movements as an input modality for eyewear computers is availability of the inertial sensors in most of the existing commercial products such as Google Glass and Vuzix smart glass while to track the eye we usually need additional hardware and software platforms. However, the mass of the head can reduce the speed of pointing, and it can be tiring for the neck muscles [14].

VII. Conclusion

In this paper, we compared two input modalities for eyewear computers: eye pointing and head pointing. Our experiment showed that the eye pointing is significantly faster than head pointing and pointing with hand-held mouse; however, head pointing is more accurate and convenient for users.

As a future work, we will repeat the experiment after improving the hardware platform. In the next wearable prototype, an eyewear computer (Vuzix smart glass) will be mounted in front of the dominant eye which is tracked by the gaze tracker. Furthermore, in the next step we will try to combine the gaze and the head tracking mechanisms so that the pointing can take advantage of both speed of the eye-based approach and the accuracy of the head tracking.

VIII. Acknowledgement

This work was supported by the EU Marie Curie Network iCareNet under Grant Number 264738.

References

[1] A. Clark, N. Newman, A. Isaakidis, and J. Pagonis, "What do we want from a wearable user interface?" in *Eurowearable, 2003. IEE*, Sept 2003, pp. 23–27.

[2] S. Nelson, J. Lee, and M. Helgeson, *Handheld computer apparatus*. Google Patents, Jun. 2005, US Patent 6,911,969. [Online]. Available: https://www.google.com/patents/US6911969

[3] A. Colaço, A. Kirmani, H. S. Yang, N.-W. Gong, C. Schmandt, and V. K. Goyal, "Mime: Compact, low power 3d gesture sensing for interaction with head mounted displays," in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '13. New York, NY, USA: ACM, 2013, pp. 227–236. [Online]. Available: http://doi.acm.org/10.1145/2501988.2502042

[4] C. Ware and H. H. Mikaelian, "An evaluation of an eye tracker as a device for computer input2," in *ACM SIGCHI Bulletin*, vol. 17, no. SI. ACM, 1987, pp. 183–188.

[5] R. A. Bolt, "Eyes at the interface," in *Proceedings of the 1982 conference on Human factors in computing systems*. ACM, 1982, pp. 360–362.

[6] R. J. Jacob, "The use of eye movements in human-computer interaction techniques: what you look at is what you get," *ACM Transactions on Information Systems (TOIS)*, vol. 9, no. 2, pp. 152–169, 1991.

[7] W. James, *The Principles of Psychology, Vol. 1*, reprint ed. Dover Publications, Jun. 1950.

[8] data eyeglasses. (2014, Jun.) Data glasses controlled by eye movements. [Online]. Available: http://www.fraunhofer.de

[9] A. T. Kurauchi, C. H. Morimoto, D. Mardanbegi, and D. Witzner Hansen, "Towards wearable gaze supported augmented cognition."

[10] NaturalPoint. (2014, Jun.) Smartnav. [Online]. Available: http://www.NaturalPoint.com

[11] H. Kim and D. Ryu, "Computer control by tracking head movements for the disabled," in *Computers Helping People with Special Needs*, ser. Lecture Notes in Computer Science, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Springer Berlin Heidelberg, 2006, vol. 4061, pp. 709–715.

[12] D. G. Evans, R. Drew, and P. Blenkhorn, "Controlling mouse pointer position using an infrared head-operated joystick," *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 1, pp. 107–117, 2000.

[13] Y. Chi, S. Ong, M. Yuan, and A. Nee, "Wearable interface for the physical disabled," in *Proceedings of the 1st international convention on Rehabilitation engineering & assistive technology: in conjunction with 1st Tan Tock Seng Hospital Neurorehabilitation Meeting*. ACM, 2007, pp. 28–32.

[14] R. Bates and H. O. Istance, "Why are eye mice unpopular? a detailed comparison of head and eye controlled assistive technology pointing devices," *Universal Access in the Information Society*, vol. 2, no. 3, pp. 280–290, 2003.

[15] (2014, Jun.) Haytham gaze tracker. [Online]. Available: http://itu.dk/research/eye/

[16] R. H. Carpenter, *Eye movements*. CRC Pr I Llc, 1991, vol. 8.

Paper 7: MAGIC Pointing for Eyewear Computers

Title of Paper

MAGIC Pointing for Eyewear Computers

Authors:

Shahram Jalaliniya, Diako Mardanbegi, Thomas Pederson

Published:

Proceedings of the ACM International Symposium on Wearable Computers (ISWC '15), 2015.

Abstract

In this paper, we propose a combination of head and eye movements for touchlessly controlling the "mouse pointer" on eyewear devices, exploiting the speed of eye pointing and accuracy of head pointing. The method is a wearable computer-targeted variation of the original MAGIC pointing approach which combined gaze tracking with a classical mouse device. The result of our experiment shows that the combination of eye and head movements is faster than head pointing for far targets and more accurate than eye pointing.

MAGIC Pointing for Eyewear Computers

Shahram Jalaliniya
jsha@itu.dk

Diako Mardanbegi
dima@itu.dk

Thomas Pederson
tped@itu.dk

IT University of Copenhagen
2300 Copenhagen S, Denmark

ABSTRACT
In this paper, we propose a combination of head and eye movements for touchlessly controlling the "mouse pointer" on eyewear devices, exploiting the speed of eye pointing and accuracy of head pointing. The method is a wearable computer-targeted variation of the original MAGIC pointing approach which combined gaze tracking with a classical mouse device. The result of our experiment shows that the combination of eye and head movements is faster than head pointing for far targets and more accurate than eye pointing.

Author Keywords
Eyewear computer; gaze tracking; head tracking; MAGIC pointing;

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces: Input devices and strategies

INTRODUCTION
Even if industry lately is trying to push Head-Mounted Display(HMD)-based wearable computers to the masses for everyday use, interaction challenges remain. The need for interaction on the move [11] and using eyewear devices in parallel with real world tasks require novel hands-free interaction techniques. For example, when hands are busy with real world tasks or in sterile environments such as operation theatre, providing touchless input modalities to the users is a big advantage for eyewear devices. Since eyewear computers sit on the users' head and in front of the users' eyes, head and eye movements are among the most interesting touchless input modalities. While head gesture-based interactions have already been supported by eyewear providers such as Google and Vuzix companies, eye tracking is not still available in commercial eyewear computers.

Just like previous mass-market user interface paradigms used in smartphones and PCs, interaction with eyewear devices relies heavily on the visual modality where point-and-select operations are fundamental. Previous studies of head and eye-pointing for eyewear computers [8, 4] have shown that while eye-pointing (letting eye movements control the mouse cursor) is faster than head pointing and mouse-pointing on HMD-based platforms, the inaccuracy of existing eye tracking methods limits eye-pointing to be used only for large targets on the display [3, 10]. On the contrary, head pointing has been found to exhibit higher accuracy [14, 8] but be limited by ergonomic challenges [4]. In this paper, we try to extend the old idea of the MAGIC (Manual And Gaze Input Cascaded)-pointing [17] to eyewear computers by combining head and eye movements for a target acquisition task. We conducted an experiment to compare the proposed MAGIC pointing approach with head pointing and eye pointing methods. We found that the proposed MAGIC approach benefits from both the speed of eye pointing and the accuracy of head pointing. In addition, the MAGIC method can decrease the amplitude of head-movements and thus ergonomic problems.

Figure 1. A participant performing the target acquisition task

RELATED WORK
Using eye gaze as an input modality has always been an interesting topic in the HCI community. The typical use of gaze in graphical user interface is a pointing mechanism to control the cursor position on the screen [4, 10]. Gaze pointing has also been explored for interaction with head-mounted displays [2, 8]; however, due to the inaccuracy of existing gaze tracking approaches and the subconscious jittery motions of eye [17], using eye-pointing is limited to the pointing towards large targets on the screen [3, 10, 4]. Aside from the target size limitation, eye-pointing has in several studies been found to be an inconvenient way of pointing [8, 4]. In fact, overloading the visual channel with a motor control task can be the main reason for eye-pointing to be recognized as an inconvenient pointing technique [17].

Another possible method for controlling the cursor is head tracking. The head movements can be detected by a camera [12, 9] or other wearable inertial sensors [7, 5]. Even if up-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ISWC '15, September 7-11, 2015, Osaka, Japan.
Copyright 2015 ©ACM 978-1-4503-3578-2/15/09…$15.00.
http://dx.doi.org/10.1145/2802083.2802094

coming eyewear computers such as Google Glass and Vuzix Smart Glass are able to detect head gestures, the head movement as a pointing modality is not much explored for eyewear computers. Previous studies have proved that head pointing method is more accurate and convenient for users compared to eye pointing or trackball mice[8]. However, the mass of the head can reduce the speed of pointing, and it can be tiring for the neck muscles [4].

The MAGIC pointing concept was firstly proposed by Zhai et al [17] to utilize eye movements for a mouse pointing task. In their proposed approach, the cursor is initially placed within the boundary of gaze area, and after the cursor appears, the user completes the target acquisition using a mouse. In this method, users do not need to know that the initial point is tied to their eye gaze; therefore, the whole pointing task seems more intuitive to the users compared to the gaze-pointing method [17]. MAGIC pointing have been explored in different ways in the HCI community. For instance, the combination of MAGIC pointing with a touch-sensitive mouse in the MAGIC-Touch System [6], is proved to be faster for a pointing task on a complex background compared to a normal mouse. Also in the Satellite Cursor System [16], the MAGIC pointing approach has been implemented without gaze tracking with the help of multiple cursors. Stellmach and Dachselt [13] extended the "conservative" method presented in [17] by introducing MAGIC Touch and MAGIC Tab pointing techniques. Their proposed techniques require extra input from users to activate the cursor. On the contrary, we used a "liberal" [17] method where the cursor moves to the new gaze location whenever the eye gaze moves more than a predefined distance from the initial point.

The most relevant work to our study is [15] where the benefits of using head movements to adjust gaze cursor position in a desktop settings is investigated. However, we investigate head-assisted gaze pointing on wearable_ and _near_to_eye_ displays that cover a small portion of the users field of view. While [15] compares a version of MAGIC pointing with gaze-only pointing, we compare MAGIC pointing with head-only pointing. In [15], head movements are directly derived from eye movements obtained from the eye tracker which is not applicable to the eyewear computers without using an additional scene camera. In our study we used the Google Glass' inbuilt inertial sensors for head tracking. The main novelty in our work is to apply MAGIC pointing technique to an eyewear computer setting, while previous works have mainly explored desktop settings. We believe that MAGIC technique can be even more useful for interaction with an eyewear device compared to stationary screens. Because the size of the display is relatively small in an eyewear device which increases the inaccuracy problem of the eye pointing.

EXPERIMENT DESIGN

In this paper, we investigate whether the combination of gaze and head tracking (MAGIC pointing) can reduce the limitations of eye and head only pointing modalities. To answer this question, an experiment is designed where the user accomplishes a target acquisition task by head only and MAGIC methods. In this experiment, we have two different target sizes (30 and 70 pixels) since the accuracy of our gaze tracker is about 1 degree, and the minimum selectable target using only our gaze tracker is about 50 pixels. The design of our experiment covers both larger and smaller targets than this limit. Also we defined two different distances (280 and 100 pixels) for the pointing task.

METHOD

Participants

16 participants (mean age = 29, 2 females) were recruited among local university students to participate in the experiment. Most of the participants were highly skilled computer users ($\bar{X} = 4.62$, $\sigma = .5$, where the range was 1 to 5), and all of them had perfect visual acuity or wearing contact lens. Two participants had the experience of using gaze-tracker systems before.

Apparatus

Since the main focus of our study is developing a novel interaction technique for new eyewear computers, we developed a prototype on the Google Glass platform (see Figure. 1). The inbuilt inertial sensors of the Glass was used to track head movements. While to detect the eye gaze, an external infrared camera was added to the Glass and positioned under the display. The camera sends the eye image wirelessly to a remote server [1]. The server analyzes the eye image in real-time and calculates the eye gaze in two dimensions. The server sends the calculated gaze to the client application on the Google Glass through WiFi connection. The client application receives the gaze data and adjusts the user interface accordingly (see Figure. 2). The accuracy of our home-made gaze tracker system is about 1 degree.

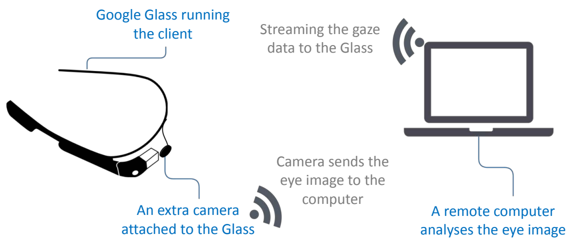


Figure 2. The system architecture of the prototype

Head only pointing technique

To control the cursor by head movements, we used the internal fusion function of the Google Glass ($RotationVector$). This function combines the data from accelerometer, gyroscope, and magnetometer to calculate yaw, pitch, and roll. The yaw value is used to calculate the horizontal position of the cursor (X), and the vertical position of the cursor (Y) is calculated based on the pitch value. The yaw and pitch values are converted to degree and multiplied by 10 to increase the sensitivity of the cursor motion. When the experiment starts the head position of the participant is in neutral state and the cursor is positioned in the middle of the screen. But after performing some head pointing tasks, the user's head might move from the neutral position. The participants were asked to return their head to the neutral position whenever they needed. In such situations, the cursor follows the screen borders.

MAGIC pointing technique
Our MAGIC pointing technique is similar to the Zhai's "liberal" approach [17]. In the MAGIC pointing condition, the gaze data is used as an implicit input to adjust the initial position of the pointer as close as possible to the target. After appearing the target, the user immediately moves the eye gaze towards the target. The gaze tracker calculates the gaze position, and the cursor appears close to the target in the area of $3°$ around the target. At this point the user is able to control the cursor by head movements to reach the target.

Procedure
The experiment started with a short introduction to the purpose of the experiment and the use of the apparatus. After preparing participants for the experiment, they were asked to use the system for a while until they felt comfortable with all conditions. This usually took 2-3 minutes for each condition. Then each participant was asked to complete the task in three different conditions. The task was a simple target acquisition in which the targets were displayed sequentially on the Glass prism, and the users were asked to point to the targets by combined head and eye movements (1st condition) and head movements (2nd condition), after which the target was conclusively selected by tapping on the Glass touchpad. The targets were red circles with two different diameters of 30, 70 pixels (equal to $0.6°$ and $1.4°$) displayed randomly on a black background one after each other. Since the new target appears immediately after selecting the previous one, the previous target is taken as the start point for the next pointing task. The pointer was illustrated by a white cross controlled by head and eye movements. When the pointer was on the target, users had to tap on the Google Glass touchpad, to select the target and accomplish the task. After the experiment, the participants were asked to complete a short questionnaire with 5-point likert scale questions to reflect on their experience in each condition. The conditions were counterbalanced to avoid the learning effect.

Design
The experiment was an within-subjects design with 16 participants, and each participant completed all conditions in one experimental session that lasted for approximately half an hour. In each condition (head pointing and Magic pointing), participants completed the task for two target sizes (30 and 70 pixels) and two different distances (100 and 280 pixels equal to $2°$ and $5.6°$). In order to remove outliers from the experiment, the participants were asked to repeat each task for 15 times and the median of the 15 trials was taken.

RESULT
We recorded the task completion time and error rate (average of the number of misses by tapping off the target) for each target acquisition task. Figure 3 (a and b) illustrates the mean and standard deviation of the task completion time and error rate for each condition. A repeated measure ANOVA is used to investigate the differences in task completion time and error. Post-hoc paired samples t-tests with a Bonferroni correction were used for pairwise comparisons. (α = .05)

Pointing Modality. The result of statistical analysis showed that the task completion time significantly varied with pointing technique: F(1, 15) = 9.27, $p < .008$. The post-hoc t-tests revealed that head pointing is significantly faster than MAGIC pointing for the short distance (100 pixels) where target = 30 t(15) = 2.42 $p < .029$ and target = 70 t(15) = 5.196 $p < .0001$. However, participants were faster in MAGIC pointing condition when they pointed to the far distance (280 pixels) for target = 30, t(15) = 7.15 $p < .0001$ and target = 70, t(15) = 3.014 $p < .009$.

Target size. As expected, the effect of target size was significant in task completion time in all conditions: F(1,15) = 285.92, $p < .0001$. Participants selected smaller targets at the same distance significantly slower in both conditions.

Distance. Also the distance factor affected the task completion time significantly: F(1,15) = 274.16, $p < .0001$. However, there was no significant difference between pointing to a big target (70) located in different distances in the MAGIC pointing condition t(15) = .037 $p = .971$.

Error rate. Analysis of error rate showed no significant difference between different modalities, target sizes, and distances. The result of questionnaire is represented in the Figure 3 (c). In the MAGIC pointing condition, we also calculated the portion of gaze in the total task completion time. In average only 8% of the total pointing time was spent by gaze tracker to detect the eye gaze and move the cursor close to the target. Moreover, in 33% of the MAGIC pointing trials with big target (70) the gaze point was exactly on the target, while for the small targets (30) just in 4% of the trials the detected gaze point was on the target.

DISCUSSION AND CONCLUSION
In this paper we empirically evaluated a pointing modality for eyewear devices using a combination of head and eye movements. All of the participants were able to complete the task. Using our gaze tracker for pointing, users would be able to select only targets larger than 50 pixels, while in our experiment all of the small targets were successfully selected using MAGIC pointing. This means MAGIC pointing technique makes it possible to select targets smaller than accuracy of our gaze tracker system.

Findings from the experiment, showed that MAGIC pointing is faster than Head pointing just for long distances. One reason for this can be the delay of the gaze tracker system to detect the gaze coordinate and communicate it to the eyewear device. If the target is too close, the head only pointing can start immediately, but in the MAGIC pointing condition, user should wait until the gaze point is detected close enough to the target. This means MAGIC pointing is faster than Head pointing only for the far targets which is in line with our initial goal which was to reduce amplitude of head-movements and its ergonomic problems. In fact, for pointing to the close targets the head does not need to move a lot. Another observation is the fact that the speed of MAGIC pointing method

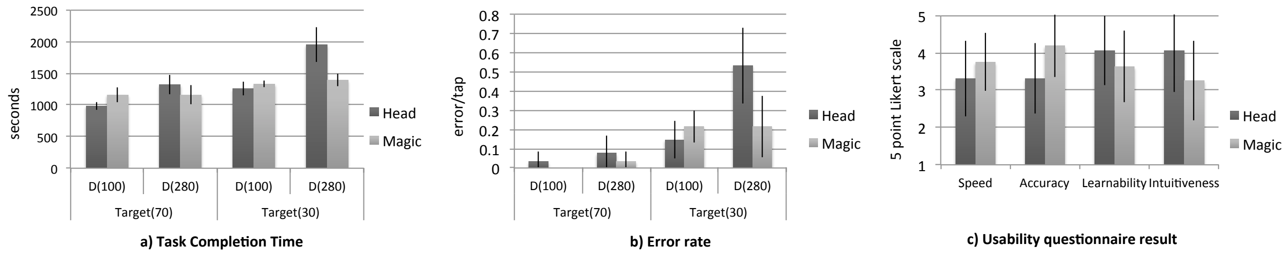


Figure 3. a) Mean of the task completion time for two distances: D(100) & D(280), two target sizes: T(30) & T(70), and two modalities: Head & MAGIC pointing, b) Mean of the error rates for all conditions, c) Results of the usability questionnaire

does not depend on the distance for big targets. This can be due to the fact that in MAGIC pointing condition, most of the distance between initial position of the pointer and the target is gone by eye movements, and the manual part of the pointing task is the distance from warped cursor position to the target. Which means the manual part of the pointing task is independent from the initial position of the pointer.

Our prototype is based on the state of the art technology in eyewear computers (Google Glass) to evaluate practicality of the MAGIC pointing as a novel target acquisition technique for these devices. Our findings indicate that 1) the MAGIC pointing looks very promising technique for target acquisition in eyewear computers, 2) probably in the emerging gaze informed (attention-aware) user interfaces, the traditional design guidelines based on the Fitts' Law (e.g. minimizing the cursor movements etc.) cannot be directly transferred to this new interface paradigm.

ACKNOWLEDGEMENTS

This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

REFERENCES

1. Haytham gaze tracker, online: http://itu.dk/research/eye/, retrived sep 2014.
2. Data glasses controlled by eye movements, online: http://www.fraunhofer.de, retrived sep, 2014.
3. Ashmore, M., Duchowski, A. T., and Shoemaker, G. Efficient eye pointing with a fisheye lens. In *Proc. of GI '05*, Canadian Human-Computer Communications Society (2005), 203–210.
4. Bates, R., and Istance, H. O. Why are eye mice unpopular? a detailed comparison of head and eye controlled assistive technology pointing devices. *UAIS 2*, 3 (2003), 280–290.
5. Chi, Y., Ong, S., Yuan, M., and Nee, A. Wearable interface for the physical disabled. In *Proc. of i-CREATe07*, ACM (2007), 28–32.
6. Drewes, H., and Schmidt, A. The magic touch: Combining magic-pointing with a touch-sensitive mouse. In *INTERACT 2009*. Springer, 2009, 415–428.
7. Evans, D. G., Drew, R., and Blenkhorn, P. Controlling mouse pointer position using an infrared head-operated joystick. *Rehabilitation Engineering, IEEE Transactions on 8*, 1 (2000), 107–117.
8. Jalaliniya, S., Mardanbeigi, D., Pederson, T., and Hansen, D. W. Head and eye movement as pointing modalities for eyewear computers. In *Proc. of the BSNWORKSHOPS '14*, IEEE Computer Society (2014), 50–53.
9. Kim, H., and Ryu, D. Computer control by tracking head movements for the disabled. In *ICCHP '06*, vol. 4061 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2006, 709–715.
10. Kumar, M., Paepcke, A., and Winograd, T. Eyepoint: Practical pointing and selection using gaze and keyboard. In *Proc. of CHI '07*, ACM (2007), 421–430.
11. Marshall, J., and Tennent, P. Mobile interaction does not exist. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, ACM (2013), 2069–2078.
12. NaturalPoint. Smartnav online: www.naturalpoint.com, retrived sep 2014.
13. Stellmach, S., and Dachselt, R. Look & touch: Gaze-supported target acquisition. In *Proc. of the CHI '12*, ACM (2012), 2981–2990.
14. Stellmach, S., and Dachselt, R. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proc. of the CHI '13*, ACM (2013), 285–294.
15. Špakov, O., Isokoski, P., and Majaranta, P. Look and lean: Accurate head-assisted eye pointing. In *Proc. of the ETRA '14*, ACM (2014), 35–42.
16. Yu, C., Shi, Y., Balakrishnan, R., Meng, X., Suo, Y., Fan, M., and Qin, Y. The satellite cursor: Achieving magic pointing without gaze tracking using multiple cursors. In *Proc. of the 23nd annual ACM UIST*, ACM (2010), 163–172.
17. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (magic) pointing. In *Proc. of the CHI '99*, ACM (1999), 246–253.

Paper 8: EyeGrip

Title of Paper

EyeGrip: Detecting Targets in a Series of Uni-directional Moving Objects Using Optokinetic Nystagmus Eye Movements

Authors:

Shahram Jalaliniya, Diako Mardanbegi

Published:

In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 5801-5811.

Abstract

EyeGrip proposes a novel and yet simple technique of analysing eye movements for automatically detecting the user's objects of interest in a sequence of visual stimuli moving horizontally or vertically in front of the user's view. We assess the viability of this technique in a scenario where the user looks at a sequence of images moving horizontally on the display while the user's eye movements are tracked by an eye tracker. We conducted an experiment that shows the performance of the proposed approach. We also investigated the influence of the speed and maximum number of visible images in the screen, on the accuracy of EyeGrip. Based on the experiment results, we propose guidelines for designing EyeGrip-based interfaces. EyeGrip can be considered as an implicit gaze interaction technique with potential use in broad range of applications such as large screens, mobile devices and eyewear computers. In this paper, we demonstrate the rich capabilities of EyeGrip with two example applications: 1) a mind reading game, and 2) a picture selection system. Our study shows that by selecting an appropriate speed and maximum number of visible images in the screen the proposed method can be used in a fast scrolling task where the system accurately (87%) detects the moving images that are visually appealing to the user, stops the scrolling and brings the item(s) of interest back to the screen.

EyeGrip: Detecting Targets in a Series of Uni-directional Moving Objects Using Optokinetic Nystagmus Eye Movements

Shahram Jalaliniya
IT University of Copenhagen
Copenhagen, Denmark
jsha@itu.dk

Diako Mardanbegi
IT University of Copenhagen
Copenhagen, Denmark
dima@itu.dk

ABSTRACT
EyeGrip proposes a novel and yet simple technique of analysing eye movements for automatically detecting the users objects of interest in a sequence of visual stimuli moving horizontally or vertically in front of the user's view. We assess the viability of this technique in a scenario where the user looks at a sequence of images moving horizontally on the display while the user's eye movements are tracked by an eye tracker. We conducted an experiment that shows the performance of the proposed approach. We also investigated the influence of the speed and maximum number of visible images in the screen, on the accuracy of EyeGrip. Based on the experiment results, we propose guidelines for designing EyeGrip-based interfaces. EyeGrip can be considered as an implicit gaze interaction technique with potential use in broad range of applications such as large screens, mobile devices and eyewear computers. In this paper, we demonstrate the rich capabilities of EyeGrip with two example applications: 1) a mind reading game, and 2) a picture selection system. Our study shows that by selecting an appropriate speed and maximum number of visible images in the screen the proposed method can be used in a fast scrolling task where the system accurately (87%) detects the moving images that are visually appealing to the user, stops the scrolling and brings the item(s) of interest back to the screen.

Author Keywords
Gaze tracking, Optokinetic Nystagmus (OKN) eye movements, Implicit interaction, Scrolling

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
We are living in the digital information age where companies, organizations, and even end users are producing an enormous and rapidly growing flow of digital information. Users of Internet applications such as social networks have already been overloaded by tremendous amount of digital information ranged from textual to graphical contents. This has resulted in us to make our browsing more efficient by quickly moving our eyes across the contents and picking the contents that seem more interesting to us. The fact that our brain processes images significantly faster than text [4] might be one of the reasons of why we are often more engaged with images than textual information and why viewing pictures is among the most popular functions in social networks such as Facebook [19].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI'16, May 07-12, 2016, San Jose, CA, USA.
© 2016 ACM. ISBN 978-1-4503-3362-7/16/05...$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858584

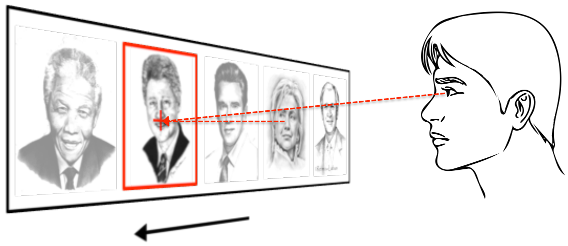


Figure 1. EyGrip technique to detect an object of interest among horizontally moving images

When people are browsing their Facebook [1] page on their mobile device, it's often that they quickly scan the Newsfeed by scrolling down or up the Facebook page until they find some interesting information. However, scrolling for navigation on small-screen devices has its own usability and inefficiency problems [9]. The three steps of a) scrolling, b) stopping the page, and c) bringing the desired content back to the display by scrolling back up are the main parts of browsing the contents. We go through the same steps when we search for a particular image in our photo gallery. Our ability to rapidly scan and process the visual cues that are quickly moving across our eyes, enables us to speed up the scrolling task. However, the third step (bringing the desired content back) can be a cumbersome task for users in a fast scrolling task since it requires a very high coordination between eyes, brain and our motor control system (e.g. touching the display with our fingers). Finding the target image that has gone out of the screen during a fast scrolling is not always easy and it sets a limitation to how fast the scrolling can be done.

[1] www.facebook.com

This paper proposes the EyeGrip method that enables computer systems to automatically detect the moving content that seems to be interesting for a user by monitoring and analysing the user's eye movements. Depending on the application, such systems can for example tag the content of interest in a series of scrolling contents or they can immediately react by stopping the content of interest in front of the user's view. EyeGrip provides an attentive scrolling mechanism which analyses the user's natural eye movements (Optokinetic Nystagmus) subtly in the background and it does not require any explicit command from the user or any change in their gaze behavior. Optokinetic Nystagmus (OKN) is a type of eye movement that occurs when a person tracks a moving field. OKN stabilizes images on the retina while viewing a sequence of moving objects. OKN has a sawtooth-like pattern that consists of alternating pursuit movements made in the direction of stimulus (slow phase) followed by saccacdes (fast phases). Generally, two forms of OKN have been described in the literature [26]. One is called Stare OKN which is a reflexive response that occurs when a viewer passively follows a moving visual field [16] and the other one is called Look OKN when a viewer voluntarily tracks moving stimulus in the visual field.

The principle behind the EyeGrip method is to analyze the combination of the saccades and smooth pursuits in the OKN eye movements to detect deviations in the OKN signal which is related to the long smooth pursuits or slow phase in the OKN eye movements (peaks in Figure 2). We used a machine learning approach to detect these peaks by feeding a window of the horizontal eye movement signal as a feature into the WEKA classifier. As we discuss this further in the paper, implementing EyeGrip does not necessarily require gaze estimation or any gaze calibration between the eye tracker and the display. However, depending on what approach is used for detecting a peak in the signal, we might need some algorithm calibration (not gaze calibration) or a learning phase to build a classifier as we did in our implementation.

In this paper, we show the feasibility of the EyeGrip method by detecting the images of interest in an image scrolling application. We further investigate the effect of two independent variables on the accuracy of the classification through a lab experiment. The first independent variable is *speed of scrolling*, and the second one is *maximum number of visible images in a single frame*. We manipulated the latter variable by changing image width. Based on our findings from the experiment, we propose some design guidelines for implementing EyeGrip. Finally, to demonstrate the utility of the EyeGrip technique in interactive systems, two follow up usability studies have been presented: 1) a picture selection application and 2) a mind reading game.

RELATED WORK

Gaze-based interaction

Using gaze as an input modality for computing devices has long been a topic of interest in HCI community, and it is due to the fact that humans naturally tend to direct eyes toward the target of interest. Gaze can be used both as an explicit and implicit input modality. Implicit input are actions and behaviors of humans, which are done to achieve a goal and are not primarily regarded as interaction with a computer, but captured, recognized, and interpreted by a computer system as input [23]. While explicit input are our intended commands to the system through mouse, keyboard, voice commands, body gestures, and etc.

Gaze for explicit input

One of the most explored explicit ways of using gaze to interact with computers is to use gaze as a direct pointing modality instead of mouse in a target acquisition task [12]. The target can be selected either by fixating the gaze for a while on a particular area (dwell-time) [25] or using a mouse click [14]. However, controlling cursor with eye movements is limited to pointing towards big targets due to the inaccuracy of gaze tracking methods and subconscious jittery motions of the eyes [29]. Eye-gesture is another explicit approach for gaze-based interaction where user performs predefined eye-strokes [8]. Previous studies [3, 14] have shown that using gaze as an explicit input modality is not always a convenient method for users. In fact, overloading eyes as humans' perceptual channel with a motor control task is not convenient [29].

Gaze for implicit input

In implicit method of using gaze in user interface design, natural movements of the eyes can be used to detect context, for example looking at certain objects in an environment can reveal interest of humans to those objects [17]. Gaze can also be used to infer about user's behaviour, for instance which objects attracts user attention during an everyday activity like cooking [20]. Another example of using gaze as an implicit input is to detect user's attention point and react to the users eye contact [24], or adapt user interface behavior [11] accordingly. The gaze data can also be used indirectly for interaction purposes [29, 13, 18, 28]. For instance, in the MAGIC pointing technique [29, 13], gaze data is used to move the cursor as close as possible to the target. Mardanbegi et al. [18] proposed a gaze-based interaction technique where the gaze data is used indirectly for head-gesture recognition. The other relevant work to our study is Pursuits interaction technique [28] which enables users to select an object on the screen by correlating eye pursuit movements with objects moving on the screen. The accuracy of their proposed technique depends on the difference of trajectories which means it fails to detect uni-directional moving objects due to the similarity of the trajectories in a uni-directional movement. On the contrary, our proposed EyeGrip method enables computer devices to detect the object of interest among uni-directional moving objects. EyeGrip is an implicit way of using gaze since we do not ask users to perform any kind of predefined eye-strokes or fixating on a particular target. The EyeGrip technique is based on analyzing natural eye movements for automatically detecting object of interest in a user interface.

Smooth pursuit recognition

The main part of our proposed approach is to automatically detect a deviation in the OKN eye movements when a particular object grabs user's attention. This deviation is related

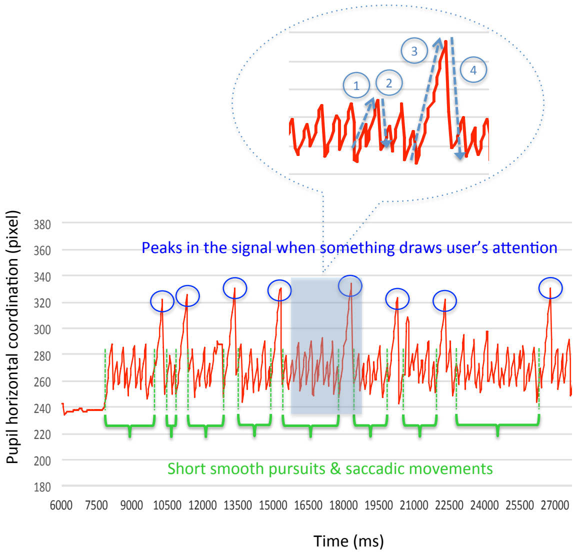


Figure 2. OKN signal generated from horizontal eye movements in a visual search task among uni-directional moving objects that move from the right side of the screen to the left. 1- Short smooth pursuit movements when eyes are scanning pictures, 2- Short saccade after a short pursuit when eyes are about to scan the next picture, 3- Long smooth pursuit (which may be supplemented by saccadic movements for fast moving objects) when an object draws user's attention, 4- Long saccade that takes the gaze back to the right area of the screen

to the slow phase of the OKN which is basically a combination of long smooth pursuits and saccadic movements. To the best of our knowledge this is the first study on using OKN in HCI. In the earlier studies, Kalman filters was used to process smooth pursuits [5, 1] while more recent works have analyzed both dispersion and velocity of the signal to classify smooth pursuits [21, 15]. Vidal and et al. [27] used a machine learning-based approach to detect pursuits by analyzing a combination of different features. In this study, we also use a machine learning algorithm to recognize patterns in the eye movement data. However, in contrast to the Pursuits [21, 15] method, we just use a single feature for classification. Our approach is explained in the next section.

THE EYEGRIP METHOD

When an object catches our visual attention, the eyes try to follow that moving object closely. This type of eye movements is called smooth pursuit. In contrast to other types of eye movements such as saccades and micro-saccades and also fixations that occur between saccades, pursuit parameters are generally more difficult to measure and are not as stereotyped as saccades [15]. Smooth pursuit consists of two phases: initiation and maintenance. Measures of initiation parameters can reveal information about the visual motion processing that is necessary for pursuit. Maintenance involves the construction of an internal, mental, representation of target motion which is used to update and enhance pursuit performance.

When we look at a series of linearly moving images, and we search for a particular image, our eyes perform a combination of saccadic and smooth pursuit movements (OKN). The smooth pursuit movements are relatively short when our eyes do not see an interesting image. As soon as an image draws user's attention, the maintenance phase of the smooth pursuit movement gets longer. In the EyeGrip technique, we exploit the difference between smooth pursuit lengths when the eyes are looking for an interesting object and when an object catches user's attention. In a visual search task among a series of uni-directional moving images, the viewer's eyes mainly move in the same direction as the moving contents. If we record the amplitude of the user's horizontal eye movements while looking at a series of moving images in the horizontal direction on the display, the generated signal looks like Figure 2 that illustrates a sawtooth like OKN signal. This figure shows the short saccadic and smooth pursuit eye movements that happen in a visual search task. The longer smooth pursuit movements occur when an object draws users' attention. In this phase of the visual search task (slow phase in OKN), eyes follow the object of interest for a longer time which generates a peak (deviation) in the signal. By detecting the moment and location of this peak (deviation), we are able to detect the object of interest among other moving objects.

We used a machine learning approach (Multilayer perceptron classifier) to detect these peaks by feeding a window of the horizontal eye movement signal as a feature into the WEKA classifier. To generate the OKN signal, we only need to detect eye movements which means there is no need for any gaze estimation or gaze calibration. In our experiment, we used a camera-based eye tracker to detect eye movements; however, to generate the OKN signal it is also possible to use other eye tracking methods such as Electrooculography (EOG) [6]. In our implementation, the classifier needs to be trained first. We collected training data from 15 participants in the experiment, and we used the same trained classifier for new participants in the follow up usability studies without adding any new training data. Since the accuracy of EyeGrip in the both usability studies for unseen data remained in the same range as the accuracy of EyeGrip in the experiment, we can conclude that the EyeGrip does not necessarily need any training phase for new users.

EXPERIMENTAL DESIGN

To characterise the eye movements in different conditions and investigate the accuracy of different algorithms, we conducted an experiment with two independent variables: 1) the speed of scrolling, and 2) the maximum number of images visible in the view-port (visible part of the sequence on the screen). To manipulate the maximum number of visible images, we can change either the size of the view-port, offset between images, or image width. Assuming that the offset between images and the width of the view-port are fixed, we changed the image width to manipulate the maximum number of images visible in the view-port. This means in some conditions the images are squeezed (Figure 4 (a) and (b)) however, since humans are extremely good in detecting faces even when they are deformed, we believe slightly squeezing images by only changing the image width has not a considerable effect on the recognition rate.

Con	Speed	Image width
1	Slow (1400 pixel/s $\simeq 26.5°/s$)	Small (480 pixels)
2	Slow (1400 pixel/s $\simeq 26.5°/s$)	Big (960 pixels)
3	Med (2000 pixel/s $\simeq 37.5°/s$)	Small (480 pixels)
4	Med (2000 pixel/s $\simeq 37.5°/s$)	Big (960 pixels)
5	Fast (2600 pixel/s $\simeq 49°/s$)	Small (480 pixels)
6	Fast (2600 pixel/s $\simeq 49°/s$)	Big (960 pixels)

Table 1. 6 different conditions used in the experiment

The dependent variables in our experiment are: 1) accuracy of the classification for detecting the moment when an image draws users' visual attention and 2) the error rate which is defined as number of target images missed by the participants divided by total number of target images.

Method

Participants

20 participants (mean age = 28, ranged from 20 to 56 years old, and 2 females) were recruited among local university staff and students to participate in the experiment. After preprocessing the data we removed the data of 5 participants which was not usable due to the inaccuracy of the eye detection for them. All of the participants had perfect visual acuity or wearing contact lens.

Apparatus

We used a home-made wearable monocular gaze tracker and the open-source Haytham gaze tracking software [2] to record the eye movement data (see Figure 4 (c)). The eye tracker was set to track the left eye for all the participants without considering the eye dominance. We assume that any possible difference between the movements of the left and right eyes will not be significant for our study. However, investigation of whether left and right eyes move differently in OKN due to the eye-dominance could be an interesting subject for the future research. The accuracy of our eye tracker is about 1 degree, and the frequency of the sampling eye data is 20 Hz. Although, in our experiment, we have not used the gaze data provided by the software. In fact, we did not calibrate the gaze tracker to calculate the gaze coordinate on the screen. We developed an application to display a series of horizontally moving images at a certain speed. The speed, direction, and the size of the images in the screen can be adjusted in the application. Both our gaze tracker software and the picture display application run on a HP laptop with a 8G RAM, Corei7 processor with the speed of 2.6 GHz, and a display with 1600×900 pixels resolution and 34.5×19.5 cm dimensions. The viewing distance form the display is about 60cm.

Procedure

The experiment started with a short introduction to the purpose of the experiment and the use of the apparatus. Then participants were asked to wear the gaze tracker, and we controlled if the gaze tracker is positioned appropriately in front of the participant's eye. Then each participant was asked to complete the task in six different conditions. The task was to look at a series of horizontally moving images of famous people's face (e.g. politicians, athletes, actors/actresses) on the screen and find the Bill Clinton's picture as target image. As soon as the participant recognizes Bill Clinton's face among other faces he/she should press space bar on the keyboard. Before starting the task, the participants were asked if they are familiar with Bill Clintons face or not. All of the participants mentioned that they know Bill Clinton, and they are able to recognize his face.

[2]http://eyeinfo.itu.dk

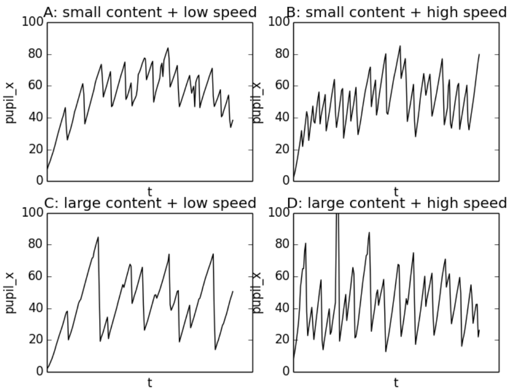


Figure 3. Optokinetik Nystagmus pattern sampled for 4 extreme conditions while viewing a set of scrolling images. W is defined as the size of the image divided by the size of the display and S is defined as the scrolling speed measured in degrees of visual field per second. The four images show the OKN pattern for conditions A) $\{W = 0.2, S = 19°/s\}$ **B)** $\{W = 0.2, S = 50°/s\}$ **C)** $\{W = 0.8, S = 19°/s\}$ **D)** $\{W = 0.8, S = 50°/s\}$

During each condition 40 pictures were displayed where 7 of them were target images. We recorded the eye movement data in the horizontal direction and the moment participants pressed the space key. In our study, the eye movement data is defined as pupil horizontal position in the eye image. The conditions were counterbalanced to avoid any learning effect. Also the position of the target images were counterbalanced in each condition.

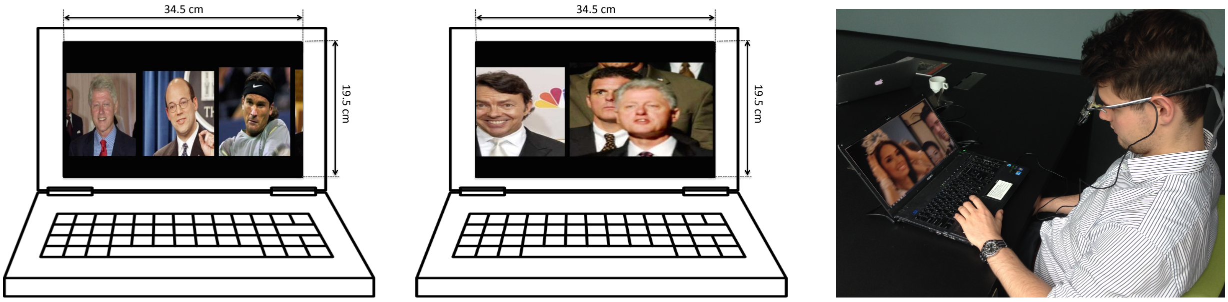
Design

The experiment was an 3×2 within-subjects design with 15 participants, and each participant completed all conditions in one experimental session that lasted for approximately 10 minutes. In each condition, participants completed the task with three different speeds (1400, 2000, 2600 pixel/s which are respectively equal to 26.5, 37.5, and 49 degrees/s) and two different image widths (480, 960 pixels equal to 9.1 and 18.2 degrees). To change the image width we kept the height of the image fixed and just rescaled the image width. All combinations of speed and image width parameters generated 6 different conditions (see Table 1.)

Event detection algorithm

In order to recognize the moment and the location of the peak in the eye movement signal when a user performs a visual search task, we used machine learning algorithms in WEKA software [3]. A comparative analysis between different classifiers in WEKA showed that Multilayer perceptron algorithm

[3]http://www.cs.waikato.ac.nz/ml/weka/



Small width conditions: 1, 3,and 5
(a)

Big width condition: 2, 4, and 6
(b)

(c)

Figure 4. a) a screen-shot of the system in small width conditions: 1, 3, and 5, b) a screen-shot of the system in big width conditions: 2, 4, and 6, c) a participant wearing the home-made mobile eye tracker performing the task

is the most accurate and reliable classifier among other available classifiers in WEKA. We used the default setting for the Multilayer perceptron algorithm in the WEKA with a single hidden layer. The eye movement data is used as the only feature in our classification. We used a sliding window to detect the moment when something draws users' visual attention. Since the experiment included three different speeds and two different widths of moving images, the best window size needed to be found for each condition. In the following sections, we briefly explain the data preparation and classification steps.

Pre-processing data

Removing outliers: as mentioned in the participant section, 20 participants were recruited for the experiment. First of all, the eye movement data of each participant is reviewed to investigate whether the eye tracker detected the pupil of the user appropriately or not. If pupil of the participant is not detected more than 25% of times, we removed the data of the participant from the experiment. After analyzing data from 20 participants, 5 participants were removed from the experiment. For the remaining participants, the missing values of pupil coordination are calculated based on the linear regression method.

Data cleaning & normalization: Before starting the experiment and after performing the task, participants were asked to look at the center of two red circles on the left and right sides of the screen. Each circle was displayed for 3 seconds. These two targets were later used for determining the lower and upper bounds of the eye movement signal. In order to prepare the data for aggregated data sets for each condition, the lower and upper values were used to normalize the eye movement data for all participants using the min-max method. The process of removing noise and outliers from the eye signal was easier after data normalization. We also used these two target points and the corresponding pupil positions while looking at each target, for roughly estimating the gaze area in the screen and locating the image of interest in small size image conditions.

Data aggregation: To calculate the performance of the classifier for each condition, we aggregated the normalized data from all participants in 6 data sets.

Sliding window & classification

To detect the event when an image draws user's attention, we used a sliding window with 50% overlap between two neighbor windows. To find the best window size for each condition, we used 4 different window sizes (10, 16, 20, 30). These window sizes have been chosen to cover the minimum and maximum duration that takes for an image to appear on the screen and disappear from the screen. This time period depends on the speed of the moving images (ranged from 1400 to 2600 pixels/s), the image widths (ranged from 480 to 960 pixels), the screen width (1600 pixels), and the sampling rate (20 Hz). The time needed for appearing an image on the screen and disappearing from the screen can be calculated using this equation: $time = (screenwidth + imagewidth)/speed$. Using the above values for screen width, image width, and speed the maximum time duration can be calculated as $time_{max} = (1600 + 960)/1400 = 1.8 seconds$, and the minimum time is equal to $time_{min} = (1600 + 480)/2600 = 0.8 seconds$. Since the sampling rate is 20 Hz, the minimum window size is equal to $0.8 \times 20 = 16$ and the maximum window size is equal to $1.8 \times 20 = 36$.

The performance of the classification is calculated for each condition using the aggregated data sets. The data sets are labeled based on the moment of pressing space key by participants as the center of each window with "Event" label. The accuracy and precision of the classification is measured using 10 folds cross-validation method. Figure 5-a illustrates the performance of classification for each condition with 4 different window sizes. We tried to find the highest classification performance where precision of classification is high for both classes: 1) "Event" and 2) "No event". We finally selected window size 30 for the first, second, and fourth conditions, window size 20 for the third and sixth conditions, and window size 16 for the fifth condition.

Results

To analyze the effect of speed and maximum number of visible images in the view-port on the classifier, the performance of classification is calculated for each participant in different conditions using 10 folds cross-validation method. Figure 6 (a) shows the mean and standard deviation of the performance of the classifier in each condition. A repeated measure

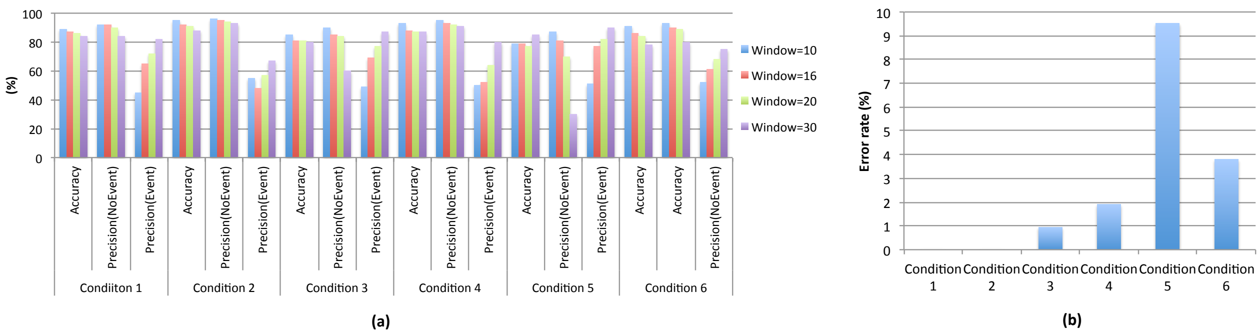


Figure 5. a) Accuracy and precision of the classification for each condition with 4 different window sizes, b) Error rate (percentage of missing targets)

ANOVA is used to investigate the differences in performance of the classifier. Post-hoc paired samples analysis with a Bonferroni correction is used for pairwise comparisons ($\alpha = .05$).

In order to measure the robustness of the classifier against unseen data, the performance of the classifier is also evaluated for the condition 4 taking a leave-one-out approach where the data of each participant was removed from the training data and used as test data. The leave-one-out evaluation for condition 4 reported an average accuracy of 87.2% ($\sigma = 11.5$) for the classification. The results of leave-one-out evaluations is illustrated in a box plot diagram (Figure 7).

The error rate (total missing target images by participants divided by total number of targets) is represented in Figure 5-b.

Effect of image width

The result of statistical analysis showed that the classification performance significantly varied with the image width: $F(1, 14) = 34.9$, $p < .0001$. The post-hoc pairwise comparisons revealed that the accuracy of the classifier is significantly higher when the image width is bigger. Figure 6 (c) illustrates changing the average accuracy of the classification when the image width changes.

Effect of speed

The statistical analysis indicated no significant effect of speed on the classification performance (see Figure 6 (b)). However, the medium speed shows a higher performance specially for small images. Moreover, participants missed more target images in the high speed conditions. Moreover, some of the participants mentioned after the experiment that it was difficult for them to complete them the task in high speed conditions specially in condition 5 where the speed was maximum and the image width was minimum.

Discussion

The results of the experiment indicates that our EyeGrip technique is more accurate for the lower number of visible images in the view-port where the image width is 960 pixels (equal to 60% of the screen width) moving with the medium speed (2000 pixel/s). As it is visible in Figure 3, increasing number of visible images in the view-port makes the sawtooth shapes in the OKN signal more homogeneous which decreases the accuracy of the classifier. Increasing the speed of moving images has a similar effect on the OKN signal (Figure 3). When images move faster on the screen, even the smooth pursuit component of the OKN eye movements have a sacadic characteristics. This makes harder for the classifier to detect slow phase of the OKN. Apart from the classification challenges, the high number of missing images in the fast conditions (see Figure 5-b) shows that following and processing fast-moving objects is harder for humans particularly when they need to see more complex images. On the other hand, there is also a lower limitation for the speed. Very low speeds let users follow all images one by one which means the shape of the smooth pursuit component of the OKN becomes more homogeneous and harder to detect for the classifier.

DESIGN GUIDELINES

We believe that the EyeGrip method is applicable to different application areas. To increase the usability of the EyeGrip technique and minimize the limitations of using the EyeGrip method, we propose the following guidelines for user interface designers.

Uni-directional moving objects

In the EyeGrip method, there is no limitation for the number of detectable objects. The important assumption is that objects need to move next to each other in the same direction at a certain speed. The objects might be placed dynamically in the queue but the system needs to know the position of each object within the sequence. In some applications, we may not be interested in detecting which content has grabbed the user's attention, and we only want to know which part of the sequence was visible in the display at the time the system has detected a long slow phase. In this case, the system can bring that part of the sequence back to the display even though there might be multiple contents visible in that moment.

Balance between moving speed & number of images

As we mentioned in the discussion section before, there are upper and lower limits for the speed of moving objects. If the objects move at lower speed the accuracy of the event detection classifier decreases. Also higher speeds increases the human error rate and risk of missing objects by user. Since the effect of increasing number of image in the view-port is similar to the increasing the speed, to maximize the accuracy of EyeGrip we need to find a balance between speed of the objects and the number of images in the view-port. Actually

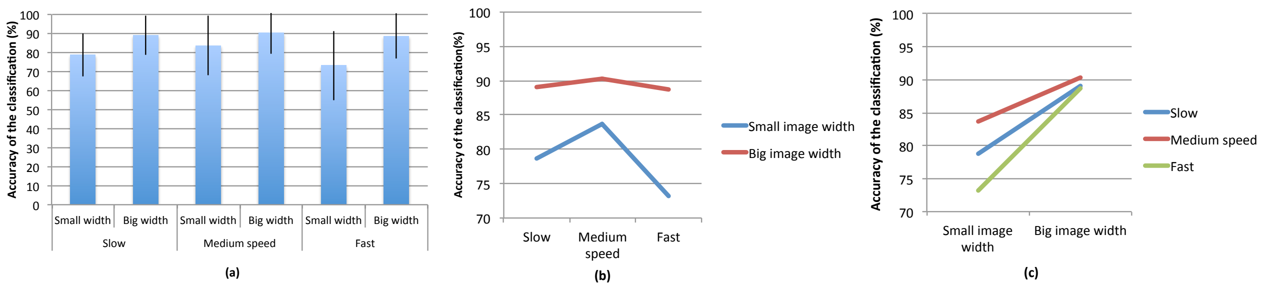


Figure 6. a) Performance of classification for each participants in different conditions, b) no significant effect of speed is observed, c) the effect of image width on the accuracy of the classifier is significant.

the EyeGrip technique works when there is a temporal tension in the visual search task. We need to be sure that we generate enough temporal tension by adjusting an appropriate speed and number of images in the view-port. At the same time, the speed should not exceed the upper limit to let users easily follow images on the screen.

Complexity of the visual search task

One of the limitations of using EyeGrip in user interface design is the fact that when a lot of images draw users' visual attention, the number of false positives increases. In other words, if users spend equal visual attention on each image in the line, the classifier cannot differentiate between target images and other objects. This limitation might be important for some applications where there is a need for high accurate recognition such as a visual inspection task. In such occasions, we can implement a two-stage algorithm where the system filters out some irrelevant objects at the first stage, and in the next step the user reviews the remaining objects to control the false positive detections.

POTENTIAL APPLICATIONS

Most of the existing gaze-based interfaces use gaze location as input. Which means for a graceful interaction, they need a very accurate gaze tracker with a cumbersome calibration procedure. In contrast, EyeGrip uses just one dimension eye movement which is much easier to achieve specially in mobile and wearable settings. This opens up a wide range of application areas that can use EyeGrip. In the following sections, we explain some of the applications that can use the EyeGrip technique for interaction.

Mind reading game

The EyeGrip technique helps the system know what attracts users' attention. This can be used in a mind reading game where the user is asked to select a person among some faces displayed on the screen. Then the user is asked to count the number of repetitions in displaying the face of that particular person among other faces while all images move horizontally in one direction with a fixed speed. The main purpose of asking users to count the number of repetitions is to draw their visual attention to a particular object. At the end the system predicts the identity of the selected person. Since EyeGrip does not need calibration and an accurate gaze tracker, the mind reading game can be installed on public displays to entertain passers-by in public places such as train stations, airports, or waiting halls.

Picture explorer on head-mounted display

One of the main challenges of interaction with eyewear computers such as Google Glass is providing input to the device. There are many situations where the hands of the user are busy with real-world task and providing a hands-free input channel can be a big advantage. The EyeGrip technique helps users with a fast and hands-free method for browsing graphical contents in eyewear computers. If we assume that in mobile scenarios, interaction with an eyewear computer should not take so much time [2], the EyeGrip method seems to be a promising technique for fast scanning visual contents on the HMD. For instance, users of the social network applications such as Facebook will be able to scan many graphical contents in a short time without any explicit input to the eyewear computer. In this case, to start scrolling the user can perform a head gesture to the left or right side or use voice commands. The graphical or textual content will start scrolling in one direction at a fixed speed. As soon as something draws the user's visual attention, the system stops scrolling and lets the user to look at that particular image or Facebook post. The user can continue scanning other contents by performing head gestures or voice commands. The beauty of using EyeGrip technique for implicitly finding users' interests is the fact that it does not need to work 100% accurate since users will always have an explicit way such as voice commands, hitting the touch-pad in Google Glass, or performing head gestures to stop scrolling. In this application, if the EyeGrip method detects object of interest in 80% of cases, it means EyeGrip has reduced the need for providing an explicit command in 80% of the times which can be a big success.

Public displays

Public displays have long been used for advertisement purposes. However, they have always been in a one-direction communication with passers-by. The EyeGrip method can help the public displays get feedback from users. One example could be to show a series of mono-directional moving images of different products on a public display where the scrolling stops whenever an image attracts attention of a user who is standing in front of the display and his/her eye is being tracked by a stationary eye tracker. EyeGrip method can

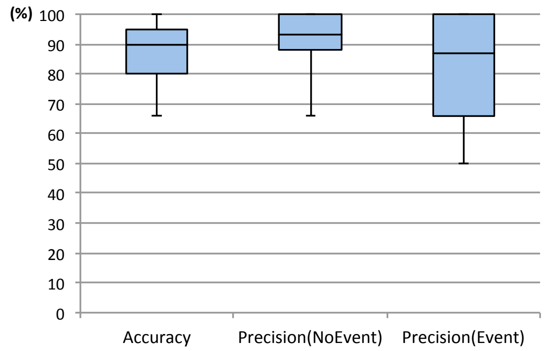


Figure 7. The performance of the classifier for condition 4 taking a leave-one-out approach.

for example be implemented using the Pupil-canthi-ratio approach [30] which is an interesting calibration-free approach for interaction with public displays. Because the relative movement between the user and the display may change the range of the horizontal movements of the eye, such a system requires the users to only move their eyes and to keep their head direction towards the center of the display. This challenge can be solved by placing a stationary infrared light source and using pupil-corneal reflection method. It is also possible that within a few seconds of recording the eye movement data while the user is looking at the moving (scrolling) contents on the display, the system figures out the lower and upper range of the eye movement signal. This can be an implicit way of calibrating the gaze direction and makes it possible to detect the images attracted users' visual attention after the scrolling has stopped.

Text reading assistant for small displays

Reading large amount of texts on small displays such as mobile devices, smart glasses, or smart watches is still challenging. One of the common approaches to facilitate reading in small displays is to enlarge the text and move it based on reader's eye movement [22]. The EyeGrip technique can be applied to such applications in order to give feedback to the system about the words which are harder to read or understand for the reader. When a user follows a word for a longer time the system can slow down the speed of moving text on the screen and provide some help, e.g. synonyms, to the user to better understand the challenging part of the text.

Assistant for visual inspection in production lines

Visual inspection is still part of quality control process in many production lines. In many cases, one or more workers control the appearance properties of the products while product move on a conveyor belt with a fixed speed. In a visual inspection task, quality controllers detect the potentially unqualified products and manually separate them from the others. If we use a gaze tracker to capture the eye movements of the quality controllers, the EyGrip technique can automatize the detection of unqualified products. If the system detects the target objects, a robot or other machines can separate them automatically. EyeGrip can potentially increases the speed of inspection by removing the manual part of the task.

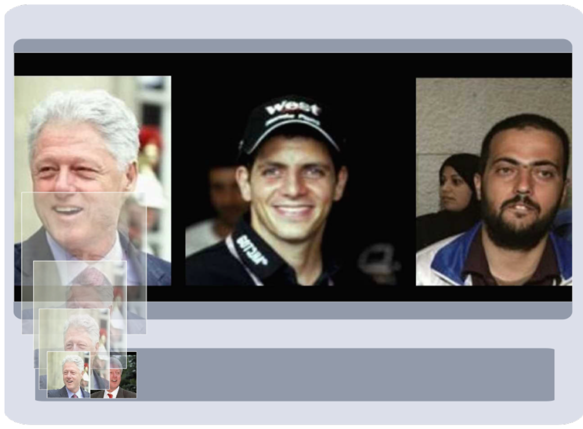


Figure 8. A screen-shot of the picture selection system (study1).

USABILITY STUDIES

To evaluate the usability of the EyeGrip method from users' point of view, we conducted two user studies: 1) a picture selection system and 2) a mind reading game. The picture selection system utilizes EyeGrip in a live interaction scenario, while the mind reading game uses EyeGrip as a context recognition method to detect what draws users' visual attention. In the following sections, we report the results of the usability evaluation in each study.

Study1: A picture selection system

We designed a desktop application to select a predefined set of images among scrolling pictures on the screen. A screenshot of the system user interface is illustrated in Figure 8. The system starts scrolling by pressing the space bar on the keyboard. To use the picture selection system in a mobile scenario, the start mechanism can change to head gestures, voice commands, etc.

Participants

8 participants (mean age = 25, ranged from 20 to 37 years old, and 1 female) were recruited among local university students to try the system. All of the participants had perfect visual acuity.

Procedure

The session started with a short introduction to the purpose of the experiment and the instruction of using the system. After preparing participants for the experiment, they were asked to wear the eye tracker apparatus and perform the task. To maximize the accuracy of the EyeGrip, we adjusted the scrolling speed equal to 2000 pixel/s and the number of visible images in the view-port (image width = 960 pixels) based on the results of our experiment. The task was similar to our experiment. The participants were asked to look at a series of moving images in the upper rectangle (see Figure 8) and count the number of Bill Clinton's pictures. Whenever the participant pays extra attention to an image, the image is selected and moved to the thumbnail panel at the bottom of the page. To give users a visual feedback about the selection mechanism, the moving procedure is animated in the user interface.

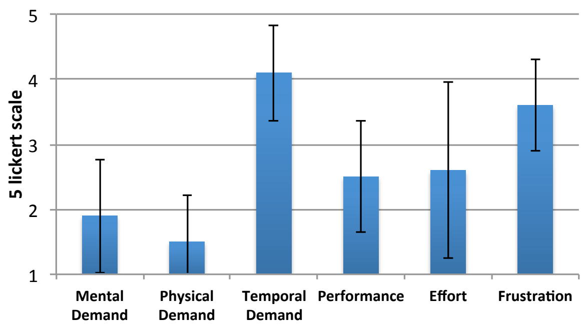


Figure 9. The result of usability questionnaire (NASA-TLX) for the picture selection system (study 1).

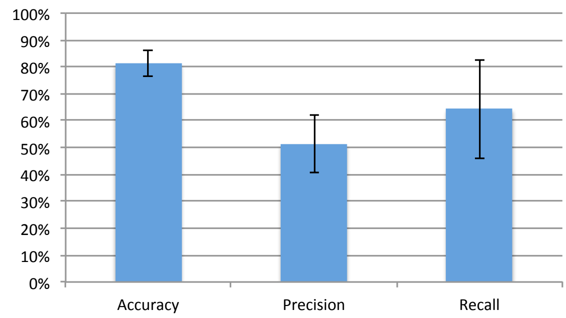


Figure 10. Performance of the EyeGrip classifier in picture selection application (study 1).

Evaluation results

To calculate the accuracy of the system, we recorded the number of correct selections, missed pictures, and wrong selections. The average accuracy, precision, and recall of the classification for all 8 participant is illustrated in Figure 10. After performing the task, the participants were asked to complete a usability questionnaire designed based on NASA-TLX [10] to reflect their experience. The result of the questionnaire is illustrated in Figure 9. The participants' general impression was also asked in an open question. They found the EyeGrip interaction technique *different* and *fun*. However, some of the participants found the EyeGrip method a bit confusing since they do not exactly know how the system selects images. Moreover, animating the image selection procedure was distracting for some users.

As it is illustrated in Figure 10, the performance of EyeGrip in the picture selection task (mean = 81%, $\sigma = 5$) is relatively close to the performance of EyeGrip in our controlled experiment (87% in Condition 4). This shows the robustness of the classifier to detect the object of interest even for the unseen data which means EyeGrip can be trained only once.

The result of the NASA-TLX questionnaire, indicates that using EyeGrip for picture selection puts time pressure on the users. This might be the reason why they felt a relatively high amount of frustration while performing the task and the accuracy of EyeGrip was slightly lower than what we observed in the experiment. Nevertheless, the task has not been physically and mentally demanding for users because the picture selection happens automatically based on their natural OKN eye movements without providing any explicit input.

Study2: A mind reading game

We also developed a mind reading game based on the software and hardware platform that we used as apparatus in the experiment. We adjusted the speed and the maximum visible number of images in the view-port similar to the condition 4 in the experiment and the picture selection application in the first usability study.

Participants

10 participants (mean age = 29, ranged from 21 to 44 years old, no female) among local university students and staff participated in the study. All of the participants had perfect visual acuity or wearing contact lens.

Procedure

We asked the participants to select a person among 4 faces printed on an A4 paper without telling us who has been chosen. Next we asked participants to wear the eye tracker hardware and sit in front of the laptop screen. They were asked to count the number of images of the selected person among other moving images on the screen. After finishing the task, the name of the selected person is displayed to the participants. All of the target images are repeated 4 times in the queue among 50 images of other people.

Evaluation results

The mind reading game was 100% accurate, and users got exited when they saw the result. Some of the participants even asked to repeat the game. Since the mind reading game has the chance to guess the selected person in 4 different occasions the probability of guessing the right person increases significantly.

Top-down & bottom up attention mechanisms

The two above-mentioned applications for EyeGrip utilize a top-down attention mechanisms in the brain. In both applications, the user knows what s/he is looking for; therefore, the visual attention is directed based on the user's longer-term cognitive strategies which is more like a top-down mechanism [7]. The EyeGrip technique can also be useful in applications where the user does not have any predefined plan for the visual search such as the Facebook Newsfeed reader explained earlier in the paper. In such applications the user's attention can be directed based on raw sensory input such as an attractive colour or fast movements (bottom-up mechanism).

DISCUSSION & CONCLUSIONS

In this paper, we introduced EyeGrip which is a novel interaction technique to support users in a visual search task in desktop, mobile, and wearable settings. EyeGrip analyzes Optokinetic Nystagmus eye movements to detect the object or area of interest among a sequence of uni-directional moving objects. This information enables users to potentially select an object without providing any explicit input to the computer.

Since OKN is a natural reaction of the eyes to the moving visual field, EyeGrip opens room for designing more intuitive methods of eye-based interaction.

We also tried to characterise the EyeGrip technique by empirically investigating the effect of scrolling speed and maximum number of visible images in the view-port (manipulated by changing image width) on the accuracy of the system and users' performance. The results of our experiment indicated a significant effect from number of image in the view-port on the performance of the classification while the effect of speed on the classification accuracy was not statistically significant. However, increasing the speed of moving images indicated a significant effect on the users' performance. But there is also a lower limit for the speed of moving objects. If the objects move very slow the user has enough time to pay equal visual attention to all of the objects. This makes the sawtooth shapes of the OKN signal more homogeneous which means it will be difficult to detect a deviation in the OKN signal when something draws user's attention.

EyeGrip utilizes the limitation of humans visual perception system in temporally intensive visual tasks where the user's visual perception mechanism needs to prioritize the time spent on following visual cues. To use the EyeGrip technique in user interface design we need to find an optimum speed and number of images in view-port to create a temporal intention, but we need to keep the speed low enough in order to minimize users' error. The temporal intention might seem to be a limitation for EyeGrip, but considering the increasing pace of producing visual contents in the Internet, we will need such mechanisms in the future to support users in quickly scanning a lot of visual contents.

In this paper, we used a home-made eye tracker, a very simple eye movement feature and classification algorithm to demonstrate the concept of EyeGrip. Using this setting we reached the accuracy of 87% where the scrolling speed is equal to 2000 pixels/s and the maximum number of visible images in the view-port is 3 (image width = 960 pixels). We believe by using more advanced features and classification models the accuracy of EyeGrip can be improved even more than what we reached in this study.

The results of the usability studies and the leave-one-out evaluation indicated an acceptable level of classification performance for the unseen data. The leave-one-out evaluation for condition 4 reported an average accuracy of 87.2% (σ = 11.5) for the classification. Furthermore, in the picture selection study, as a real-time interactive application, the average accuracy was 81% (σ = 5) and in the mind reading game the EyeGrip technique was 100% accurate. This shows that the EyeGrip technique can be used pretty accurate without any additional training for new users.

In the future work, we will implement the EyeGrip method by capturing the eye movement data from stationary eye trackers and other sensing technologies such as EOG for wearable systems. In that case, the OKN signal will be generated based on only eye movement data, and other peak detection algorithms can be used for finding local peaks in the OKN signal.

ACKNOWLEDGEMENTS
We would like to thank our fellow PIT Lab members for their valuable input at various stages of the work.

REFERENCES
1. W. Abd-Almageed, M.S. Fadali, and G. Bebis. 2002. A non-intrusive Kalman filter-based tracker for pursuit eye movement. In *American Control Conference, 2002. Proceedings of the 2002*, Vol. 2. 1443–1447 vol.2. DOI: http://dx.doi.org/10.1109/ACC.2002.1023224
2. Daniel L. Ashbrook. 2010. *Enabling Mobile Microinteractions*. Ph.D. Dissertation. Atlanta, GA, USA. Advisor(s) Starner, Thad E. AAI3414437.
3. Richard Bates and Howell O Istance. 2003. Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices. *UAIS* 2, 3 (2003), 280–290.
4. L Burmark. 2004. Visual Presentations That Prompt, Flash & Transform Here are some great ways to have more visually interesting class sessions. *Media and methods* 40 (2004), 4–5.
5. T. Cecchin, D. Sauter, D. Brie, and B. Martin. 1990. On-line Separation Of Smooth Pursuit And Saccadic Eye Movements. In *Engineering in Medicine and Biology Society, 1990., Proceedings of the Twelfth Annual International Conference of the IEEE*. 777–778. DOI: http://dx.doi.org/10.1109/IEMBS.1990.691327
6. M. Cheng and J.S. Outerbridge. 1975. Optokinetic nystagmus during selective retinal stimulation. *Experimental Brain Research* 23, 2 (1975), 129–139. DOI:http://dx.doi.org/10.1007/BF00235455
7. Charles E Connor, Howard E Egeth, and Steven Yantis. 2004. Visual attention: bottom-up versus top-down. *Current Biology* 14, 19 (2004), R850–R852. DOI: http://dx.doi.org/doi:10.1016/j.cub.2004.09.041
8. Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the Computer Using Gaze Gestures. In *Human-Computer Interaction INTERACT 2007*. Lecture Notes in Computer Science, Vol. 4663. Springer Berlin Heidelberg, 475–488. DOI: http://dx.doi.org/10.1007/978-3-540-74800-7_43
9. Johannes Harms, Martina Kratky, Christoph Wimmer, Karin Kappel, and Thomas Grechenig. 2015. Navigation in Long Forms on Smartphones: Scrolling Worse than Tabs, Menus, and Collapsible Fieldsets. In *INTERACT 2015*. Lecture Notes in Computer Science, Vol. 9298. Springer, 333–340. DOI: http://dx.doi.org/10.1007/978-3-319-22698-9_21
10. Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183. DOI: http://dx.doi.org/doi:10.1016/S0166-4115(08)62386-9

11. Aulikki Hyrskykari, Päivi Majaranta, and Kari-Jouko Räihä. 2005. From gaze control to attentive interfaces. In *Proceedings of HCII*, Vol. 2.

12. Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '90)*. ACM, New York, NY, USA, 11–18. DOI: http://dx.doi.org/10.1145/97243.97246

13. Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. 2015. MAGIC Pointing for Eyewear Computers. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15)*. ACM, New York, NY, USA, 155–158. DOI: http://dx.doi.org/10.1145/2802083.2802094

14. Shahram Jalaliniya, Diako Mardanbegi, Thomas Pederson, and Dan Witzner. 2014. Head and Eye Movement as Pointing Modalities for Eyewear Computers. In *Wearable and Implantable Body Sensor Networks Workshops (BSN Workshops), 2014 11th International Conference on*. 50–53. DOI: http://dx.doi.org/10.1109/BSN.Workshops.2014.14

15. Do Hyong Koh, Sandeep Munikrishne Gowda, and Oleg V. Komogortsev. 2010. Real Time Eye Movement Identification Protocol. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 3499–3504. DOI: http://dx.doi.org/10.1145/1753846.1754008

16. Lori A Lott and Robert B Post. 1993. Up-down asymmetry in vertical induced motion. *PERCEPTION-LONDON-* 22 (1993), 527–527.

17. PaulP. Maglio, Teenie Matlock, ChristopherS. Campbell, Shumin Zhai, and BartonA. Smith. 2000. Gaze and Speech in Attentive User Interfaces. In *Advances in Multimodal Interfaces ICMI 2000*. Lecture Notes in Computer Science, Vol. 1948. Springer Berlin Heidelberg, 1–7. DOI: http://dx.doi.org/10.1007/3-540-40063-X_1

18. Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-based Head Gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 139–146. DOI: http://dx.doi.org/10.1145/2168556.2168578

19. Thomas Paul, Daniel Puscher, and Thorsten Strufe. 2015. The User Behavior in Facebook and its Development from 2009 until 2014. *arXiv preprint arXiv:1505.04943* (2015).

20. Ravikrishna Ruddarraju, Antonio Haro, Kris Nagel, Quan T. Tran, Irfan A. Essa, Gregory Abowd, and Elizabeth D. Mynatt. 2003. Perceptual User Interfaces Using Vision-based Eye Tracking. In *Proceedings of ICMI '03*. ACM, New York, NY, USA, 227–233. DOI: http://dx.doi.org/10.1145/958432.958475

21. Javier San Agustin. 2009. *Off-the-shelf gaze interaction*. Ph.D. Dissertation. IT-Universitetet i KøbenhavnIT University of Copenhagen, DirektionenManagement, InstituttetThe Department, Innovative CommunicationInnovative Communication.

22. Christopher A Sanchez and James Z Goolsbee. 2010. Character size and reading to remember from small displays. *Computers & Education* 55, 3 (2010), 1056–1062. DOI: http://dx.doi.org/doi:10.1016/j.compedu.2010.05.001

23. Albrecht Schmidt. 2005. Interactive context-aware systems interacting with ambient intelligence. *Ambient intelligence* 159 (2005).

24. Jeffrey S Shell, Roel Vertegaal, and Alexander W Skaburskis. 2003. EyePliances: attention-seeking devices that respond to visual attention. In *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 770–771.

25. Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of Eye Gaze Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, New York, NY, USA, 281–288. DOI: http://dx.doi.org/10.1145/332040.332445

26. JWG Ter Braak. 1936. Untersuchungen über optokinetischen Nystagmus. *Arch Neerl Physiol* 21 (1936), 309–376.

27. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2012. Detection of Smooth Pursuits Using Eye Movement Shape Features. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 177–180. DOI: http://dx.doi.org/10.1145/2168556.2168586

28. Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of UbiComp '13*. ACM, New York, NY, USA, 439–448. DOI: http://dx.doi.org/10.1145/2493432.2493477

29. Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proc. of the CHI '99*. ACM, 246–253. DOI: http://dx.doi.org/10.1145/302979.303053

30. Yanxia Zhang, Andreas Bulling, and Hans Gellersen. 2014. Pupil-canthi-ratio: A Calibration-free Method for Tracking Horizontal Gaze Direction. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. ACM, New York, NY, USA, 129–132. DOI: http://dx.doi.org/10.1145/2598153.2598186

Paper 9: Seamless Interaction with Scrolling Contents

Title of Paper

Seamless Interaction with Scrolling Contents on Eyewear Computers Using Optokinetic Nystagmus Eye Movements

Authors:

Shahram Jalaliniya, Diako Mardanbegi

Published:

In Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications (ETRA '16). ACM, New York, NY, USA, 295-298, 2016.

Abstract

In this paper we investigate the utility of an eye-based interaction technique (EyeGrip) for seamless interaction with scrolling con- tents on eyewear computers. EyeGrip uses Optokinetic Nystagmus (OKN) eye movements to detect object of interest among a set of scrolling contents and automatically stops scrolling for the user. We empirically evaluated the usability of EyeGrip in two different ap- plications for eyewear computers: 1) a menu scroll viewer and 2) a Facebook newsfeed reader. The results of our study showed that the EyeGrip technique performs as good as keyboard which has long been a well-known input device. Moreover, the accuracy of the EyeGrip method for menu item selection was higher while in the Facebook study participants found keyboard more accurate.

Seamless Interaction with Scrolling Contents on Eyewear Computers Using Optokinetic Nystagmus Eye Movements

Shahram Jalaliniya*
Diako Mardanbegi†

Abstract

In this paper we investigate the utility of an eye-based interaction technique (EyeGrip) for seamless interaction with scrolling contents on eyewear computers. EyeGrip uses Optokinetic Nystagmus (OKN) eye movements to detect object of interest among a set of scrolling contents and automatically stops scrolling for the user. We empirically evaluated the usability of EyeGrip in two different applications for eyewear computers: 1) a menu scroll viewer and 2) a Facebook newsfeed reader. The results of our study showed that the EyeGrip technique performs as good as keyboard which has long been a well-known input device. Moreover, the accuracy of the EyeGrip method for menu item selection was higher while in the Facebook study participants found keyboard more accurate.

Keywords: Eye tracking, Optokinetic Nystagmus eye movements (OKN), eyewear computers, scrolling, implicit input

Concepts: •**Human-centered computing** → **Interaction techniques;**

Figure 1: *A right-to-left fast scrolling menu on a head-mounted display where it stops on the content that has attracted the user's visual attention*

1 Introduction

Scrolling for navigation on small-screen devices (e.g. smartphones) has its own usability and inefficiency problems [Harms et al. 2015] which can be even more challenging on eyewear computers such as Google Glass. The only mechanism for scrolling the main menu in Google Glass UI, is to perform touch gestures on the touch-sensitive surface on the right side of the device. However, this manual mechanism is not always the best modality where the users' hands are busy with other tasks. Moreover, on a small screen, finding the desired content that has gone out of the screen requires a lot of touch gestures which is not always easy and sets a limitation to how fast the scrolling can be done. An alternative approach is to use eye-based techniques for hands-free interaction with scrolling contents in eyewear computers. But despite the great potentials of using our eyes for interaction, eye-based interaction techniques are still not widely used. Several challenges need to be tackled to make gaze interaction more pervasive. First, existing eye trackers need to be calibrated for each user due the differences between individual eye geometries. Furthermore, other factors such as relative movements of the eye and the eye tracker, ambient light conditions, and calibration quality affect the accuracy of gaze tracking. Finally, eye is a perceptual organ and is not suitable to use as an explicit input [Jacob 1990].

*e-mail: jsha@itu.dk
†e-mail:dima@itu.dk

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. © 2016 ACM.
ETRA 2016, March 14 - 17, 2016, Charleston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2857539

Variety of gaze interaction techniques have been proposed in the recent years to overcome the above-mentioned challenges of eye-based interaction. Overall, we see trends towards more implicit way of using eye input in non-command interfaces [Nielsen 1993; Mardanbegi et al.]. The *EyeGrip* method is a novel implicit eye-based method for interaction with scrolling contents that addresses all of the above-mentioned challenges [Jalaliniya and Mardanbegi]. EyeGrip is a calibration-free method that uses natural reflexive Optokinetic Nystagmus eye movements for hands-free interaction with dynamic user interfaces and helps the user intuitively stop a sequence of moving (scrolling) visual contents displayed on the computer screen. In this paper, we investigate the utility of the EyeGrip method for eyewear computers that are becoming increasingly popular, to allow the user to seamlessly select an item among a series a scrolling contents in the near-eye display (Figure 1).

2 Related Work

The idea of using eye input in a scrolling task was originally suggested by Jacob et al. [Jacob 1990] and further studied (e.g. by [Kumar and Winograd]) but these work were limited to only enhancing the task of reading digital documents by automatically scrolling the page based on where the user is looking at. [Vidal et al. 2013] proposed the Pursuits method for interaction with dynamic user interfaces. Pursuits allows users to select an object among several moving objects on a display by following the object with their eyes which leads to a smooth pursuit eye movement. Since in the Pursuits method the trajectory of the moving objects should not to be identical, it is not possible to use Pursuits for interaction with scrolling contents with an identical trajectory. On the contrary, the EyeGrip method [Jalaliniya and Mardanbegi] enables us to detect an object of interest among a set of moving objects that all move in the same direction at the same speed. The earlier study on EyeGrip [Jalaliniya and Mardanbegi] was done on a desktop computer and a machine learning approach was used to detect the object of interest while in this paper, we implemented the EyeGrip method on a mobile setup with a head-mounted display (HMD) and a different classifier that analyses users' eye-strokes.

3 EyeGrip Method

Optokinetic nystagmus (OKN) has a sawtooth-like pattern that consists of alternating pursuits movements made in the direction of stimulus (slow phase) followed by saccades (fast phases). When a user is doing a visual search by looking at a scrolling sequence of contents on a computer screen we can see that the eye follows some objects longer than the others. The longer the user is following an object of interest, the higher the peak of the slow phase appears in the eye movement signal. In Figure 2 the blue signal (E) represents the horizontal eye movements (over time) recorded by an eye tracker in a visual search task where a user looks at horizontally (right to left) scrolling images and searches for a particular image (target) among other pictures. The upper/lower bounds of the signal (marked on the vertical axis of the graph) correspond to the left and right sides of the screen. The slow and fast phases of the Look OKN are visible in the figure as well as the longer slow phases that has happened when the target images have drawn user's attention two times (marked by green circles). These longer slow phases in the signal are denoted by *target-peak* in this paper. Implementing the EyeGrip method requires a tight communication and synchronization between the eye tracker and the UI. The eye tracker tracks the user's eye movements and generates a set of feature vectors (e.g. pupil center or even gaze coordinates). For the sake of synchronization later these vectors could be time-stamped (e.g. $[t_i, E_i]$ where t_i is time and E_i is the eye feature vector). The scrolling engine that updates the UI, controls what part of the moving sequence is within the display at each time instant (e.g. $[t_i, S_i]$ where S_i is the position an the state of the scroller). These two sets of vectors are pushed to a classifier that detects the target-peaks in the OKN signal. Once the classification is done and a target-peak is detected, the UI needs to be updated accordingly e.g. by stop scrolling and bringing the area or content of interest back to the screen.

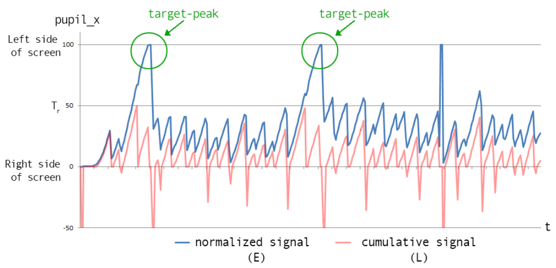


Figure 2: *The blue color signal (E) is generated from horizontal eye movements in a visual search task among images scrolling from right to the left side of the screen. The red color signal (L) is generated in the classification process. The target-peak is detected when L exceeds the T_r limit.*

3.1 Our approach for detecting target-peaks

In this study, we used pupil center as the main feature obtained from the eye image and we were only interested in the horizontal OKN movements. The moving stimuli on the display were scrolling from right to left, and only the horizontal component of the pupil center is captured. We adopted a sliding window approach for detecting the target-peaks and once a peak in the signal is detected, the system reacts by stop scrolling the content. Although, the EyeGrip technique itself does not rely on the absolute position of the gaze point and it does not require any gaze estimation, our classification algorithm needs to be calibrated (not gaze calibration) because it relies on an absolute threshold operator in the decision process that needs to be adopted for each user. In order to calculate a default threshold value for different users, we have normalized the eye signal for all users and in our implementation, we asked each participant to look at two red circles to detect right and left borders of the screen. This is used to determine the lower/upper bounds of the OKN signal and to obtain the parameters of a linear mapping function (F_{norm}). This function is used to map the pupil coordinates to a normalized range of $[LB, UB]$ where LB and UB respectively correspond to the pupil position when a user looks at the right and left sides of the screen. For each incoming frame, we find the pupil center in the image denoted as E_{new}. Then we clean the data by removing zero values caused by pupil tracking failures. After applying a smoothing window and normalizing the new data ($F_{norm}(E_{new})$), the pre-processed data E_N is passed to the classifier. The classifier always keeps a window of the recent N observations $E_{1:N} = \{E_t | 1 \leqslant t \leqslant N\}$ (a set of observations of E_t in a sliding window of N frames within time span of $[t - N + 1, t]$). After a new observation the classifier updates the buffer. The classifier also buffers two other sets of features both generated from the main observation input of E_N. The first set is defined as $\Delta E_{1:N} = \{\Delta E_t = E_{t+1} - E_t | 2 \leqslant t \leqslant N\}$ which is basically a difference between adjacent items in the sliding window $E_{1:N}$. The other set is a cumulative sum of the items in the data set $\Delta E_{1:N}$ and is defined as $L_{1:N} = \{L_t | 2 \leqslant t \leqslant N\}$ where L_t is defined based on the following rules where $direction = \Delta E_t \times \Delta E_{t-1}$.

$$L_t = \begin{cases} \Delta E_t + L_{t-1} & \text{if } direction > 0 \\ 0 & \text{if } direction \leq 0 \end{cases} \quad (1)$$

Finally using a general threshold T_r the current frame E_N is classified into target-peak class (TP) for when $L_N > T_r$ and $E_N < 0.5 \times UB$ otherwise to no-interest-area class (NIA). Where the term $E_N < 0.5 \times UB$ is to ensure that when a target-peak is detected, the user has been looking at the left side of the screen. Whenever a target-peak is detected the two $\Delta E_{1:N}$ and $L_{1:N}$ get empty to ensure that the event is not detected twice. Based on a preliminary study, we have derived the optimized constant values used in the classification process as: $\{T_r = 50, LB = 0, UB = 100, N = 100\}$ The main reason for using the value L_N instead of E_N for the classification is to ensure that a fast movement from right to left side of the screen or noise in the data is not considered as a slow phase (see the last peak of the signal in Figure 2). Figure 2 (red line) shows the result of applying the classification process on an example signal.

3.2 Voluntary vs involuntary interactions

As humans, we can attend to objects one by one. Our visual attention can be attracted by salient stimuli that 'pop out' in our surroundings which is called bottom-up attentional mechanism. Attention can also be voluntarily directed to an object based on our longer-term cognitive strategies which is more like a top-down mechanism [Connor et al. 2004]. In EyeGrip, we detect users' object of interest in a scrolling UI where users' attention can be directed to a certain object either voluntarily due to a predefined task (top-down mechanism) or involuntarily to operate on raw sensory input such as an attractive colour or fast movements in the user interface (bottom-up mechanism). We exploit these two attentional mechanisms to implement EyeGrip interaction technique for two different types of applications. In the first application, users have a predefined plan to search for a particular menu item on the screen. In this case, EyeGrip supports the top-down attention mechanism by stopping the menu scroller when a menu item draws users' attention. We call this type of using EyeGrip, *voluntary interaction* since the user is voluntarily searching for a specific object. In the second application, there is no predefined goal in the visual search task. Users might look at the computer screen with no specific goal or task. In such cases they might be attracted to an image due to the novelty of the image or transients such as motion, and change [Pashler and Harris 2001]. In this type of *involuntary interactions*, the bottom-up attention mechanism directs users' attention.

4 User Study

To characterise involuntary and voluntary types of interactions, we developed two different applications where users can select a particular item among scrolling visual contents using EyeGrip: 1) a menu scroll-viewer and 2) a Facebook newsfeed reader. In order to maximize the performance of EyeGrip the velocity and image width should be adjusted carefully [Jalaliniya and Mardanbegi]. Based on the earlier study on EyeGrip and our preliminary study we selected the optimized values for the size of the content ($W_{content}$), size of the constant offset between the contents (W_{offset}), and scrolling speed ($Speed$) as: $W_{content} = 0.6 \times W_{display}$, $W_{offset} = 0.1 \times W_{display}$, and $Speed = 35°/second$

4.1 Participants

We recruited 11 participants (mean age = 35, from 28 to 52 years old, and 3 females) among local university staff to try both systems. Three participants wear glasses and one uses contact lenses. Rest of the participants have perfect visual acuity. Also 9 of 11 participants are Facebook users. Each participant completed the tasks for both studies in a single session in approximately 30 minutes.

4.2 Apparatus

Due to the limitation in processing power of Google Glass, we simulated the Google Glass info cards in a desktop application with an HMD. The screen of the desktop application is mirrored on a binocular HMD (ICUITI DV920) with the resolution of 640 × 480. For tracking eye movements, we used a home-made wearable monocular gaze tracker and the open-source Haytham gaze tracking software[1]. The eye tracking camera is mounted on the HMD. The frequency of the sampling eye data is 20Hz. In the Facebook reader app the newsfeed is scrolled horizontally unlike the original Facebook newsfeed in mobile devices which scrolls vertically. Due to the dimensions of the display in state of the art eyewear computers it is easier to explore Facebook posts horizontally.

4.3 Study 1: gaze enabled menu scroll-viewer

The main concept in the Google Glass UI is a side-scrolling stream of info cards providing updates for different categories of interests, such as email, messages, weather, etc. The main challenge of using info cards is that if you want to find a particular card, you have to scroll through every existing card which requires a lot of touch gestures on the touchpad. EyeGrip can be a solution to this problem. The gaze-enabled menu scroll-viewer system helps users stop scrolling info cards whenever target card passes in front of the users' eyes. We investigate the utility of EyeGrip for such menu selection tasks. To have a baseline for comparison, we implemented a manual method (keyboard) for menu selection and the participants were asked to select info cards with both manual and EyeGrip methods.

4.4 Study 2: Facebook newsfeed reader

When people are browsing their Facebook page, particularly in smart phones, it's often the case that they quickly scan their newsfeed by scrolling down or up until they find something interesting to stop on. If Google Glass users want to explore their Facebook newsfeed, they would need to keep swiping back and forth on the touchpad of the Glass. Just like the menu scroll-viewer app, EyeGrip can provide a hands-free automatic mechanism to stop scrolling when an interesting Facebook post draws users' attention. To investigate the utility of EyeGrip in such applications, we developed a Facebook newsfeed reader app for eyewear computers. The participants, tried the Facebook app immediately after the menu scroll-viewer. To have a baseline for comparison, the participants were asked to use both manual (keyboard) and EyeGrip methods.

[1]http://eyeinfo.itu.dk

4.5 Procedure

The session started with a short introduction on purpose of the study. Since none of the participants were familiar with the concept of info cards in Google Glass, first we asked them to wear a Google Glass and scroll between different cards using the touchpad of the Glass. Next, we asked them to wear the HMD and eye tracker set.

Menu scroll viewer: In a menu selection task usually users are familiar with the menu items; therefore, we showed our sample 14 menu items (14 is an arbitrary big enough sample size) to the user before starting the task. Then the participant started the manual selection mode where the menu items (cards) start scrolling on the screen and the user is asked to press the space key to stop scrolling when the target item appears on the screen. The participants could correct their error using left and right arrow keys if they stop before or after the target item by mistake. The task is repeated five times for five different target items. In the next condition, the participants performed the same task using EyeGrip. In this step, the EyeGrip method automatically stopped the scrolling content as soon as the participant found the target item. Also in this condition, the task was repeated five times for five different target items, and we recorded the accuracy of the automatic menu selection and the number of corrections. In order to remove the order effect, the manual and EyeGrip conditions are counterbalanced and the cards were randomly positioned in the queue.

Facebook reader: this part started immediately after the first part and participant went through a similar procedure (manual and EyeGrip conditions). The only difference was that instead of menu items, 50 Facebook posts (randomly selected from public Facebook pages such as CNN, National Geographic, etc.) was scrolling on the screen, and participants were not familiar with the posts. In contrast to the study 1, there was no plan for stop scrolling. The users were allowed to stop scrolling whenever they found something interesting among the scrolling content. After finishing each part, the users were asked to complete a questionnaire with 5-point likert scale questions polling their experience completing the task. After filling out the questionnaire, the participants were interviewed briefly.

4.6 Results

Study1: We defined the error as the total number of items between the selected item and the target item. This means if the target card is selected correctly the error is zero. The statistical paired t-test with Holm-Bonferroni corrections indicates a significant difference in the accuracy of menu selection for manual (mean = 2, $\sigma = 1.6$) and EyeGrip (mean = .90, $\sigma = .94$) conditions; t(10) = 2.12, p = .02 $< \alpha = .05$. The results of the usability questionnaire is illustrated in Figure 3-b. Pairwise comparisons showed that, participants found EyeGrip significantly more intuitive than the manual for selecting menu items; t(10) = 2.6, p = .01 $< \alpha = .05$. Moreover, the users evaluated EyeGrip as a more comfortable method for selecting items compared to the manual method; t(10) = 5.16, p = .0002 $< \alpha = .05$. Finally, EyeGrip is also recognized as a faster

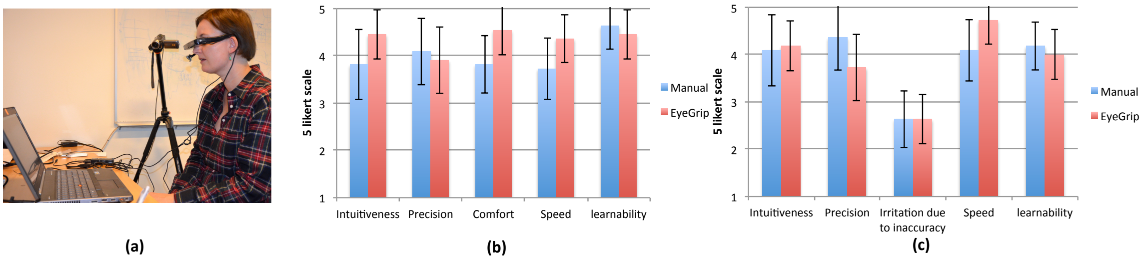


Figure 3: *a) A participant performing the task, b) Result of the questionnaire for study1, and c) Result of the questionnaire for study2*

method for selecting menu items compared to the manual method; t(10) = 2.05, p = .03 $< \alpha = .05$.

Study2: The results of the questionnaire is represented in Figure 3-c. We compared different aspects of the usability for each condition using statistical paired t-tests with Holm-Bonferroni corrections. The statistical analysis indicated that participants experienced the manual method significantly more precise than EyeGrip for the Facebook app, t(10) = 1.88, p = .04 $< \alpha = .05$. However, EyeGrip is evaluated significantly faster compared to manual method for reading the Facebook newsfeed, t(10) = 2.6, p = .01 $< \alpha = .05$. Other usability aspects indicated no significant difference.

In both studies, we asked participants how much temporal tension they felt during the task. They felt significantly more temporal tension in the Facebook study (mean = 3.09 $\sigma = .69$) compared to the scroll-viewer study; t(10) = 2.69, p = .01 $< \alpha = .05$.

5 Discussion & Conclusions

In the menu selection study, EyeGrip is more accurate than the manual method. It can be due the high speed of the scrolling menu items in the screen which requires a fast reaction to select the target item as soon as it appears on the screen. EyeGrip can be potentially a faster than the manual technique since in EyeGrip, as soon as the eyes react to an item on the screen the system stops scrolling without any additional motor task. But the manual method requires a very high coordination between eyes, brain and our motor control system which decreases the user performance. Due to the same reason, in the first study, participants found EyeGrip faster than the manual method. The participants also found EyeGrip more comfortable and intuitive method for selecting menu items compared to the manual method. The comfort and intuitiveness of the EyeGrip can be explained by being a hands-free interaction technique which is easier to use in eyewear devices. In the second study, the EyeGrip is again evaluated as a faster method for exploring Facebook compared to the manual method. However, the manual method is recognized significantly more accurate than EyeGrip. The reason can be the significant role of the bottom-up attention mechanism in the Facebook reader app which directs users' attention based on properties of the visual contents. Since this type of attention is involuntary, sometimes even the user does not know if s/he is paying attention to a particular object in the scene. This might be the reason why the manual approach is evaluated significantly more precise method in the Facebook app and not in the menu selection app. This reveals the limitations of using EyeGrip in involuntary applications. In both studies, we compared the usability of EyeGrip with keyboard which is a very old and well-known input device. Nonetheless, the EyeGrip technique is evaluated event better in some usability aspects such as speed and intuitiveness. The result of interviews also showed that 90% of the participants preferred the EyeGrip method for both applications. One of the points made by most of the participants was the need for a hands-free modality to start the scrolling movement. In a real application, the manual method that we used to start scrolling can be replaced with another hands-free technique such as dwell-time or even head gestures that are detected by eyewear computers.

Apart from functional utilities of EyeGrip, many participants particularly expressed that the EyeGrip interaction technique is *different* and *fun*: *"It can be relaxing if you just lay down at home, wear a HMD, and let your unconscious attention together with EyeGrip decide what you should see in Facebook newsfeed." (Participant 4)*

In the future, we shall investigate the use of EyeGrip with vertical scrolling stimuli as vertical OKN has different characteristics. The way of handling deviations in OKN movements, once they are detected by the system, could also be the subject of future studies.

References

CONNOR, C. E., EGETH, H. E., AND YANTIS, S. 2004. Visual attention: bottom-up versus top-down. *Current Biology 14*, 19, R850–R852.

HARMS, J., KRATKY, M., WIMMER, C., KAPPEL, K., AND GRECHENIG, T. 2015. Navigation in long forms on smartphones: Scrolling worse than tabs, menus, and collapsible fieldsets. In *INTERACT 2015*. Springer, 333–340.

JACOB, R. J. K. 1990. What you look at is what you get: Eye movement-based interaction techniques. In *Proc. of CHI '90*, ACM, New York, NY, USA, 11–18.

JALALINIYA, S., AND MARDANBEGI, D. Eyegrip: Detecting targets in a series of uni-directional moving objects using optokinetic nystagmus eye movements. In *Proc. of CHI '16*.

KUMAR, M., AND WINOGRAD, T. Gaze-enhanced scrolling techniques. In *Proc. of UIST '07*.

MARDANBEGI, D., HANSEN, D. W., AND PEDERSON, T. Eye-based head gestures. In *Proc. of ETRA '12*.

NIELSEN, J. 1993. Noncommand user interfaces. *Commun. ACM 36*, 4 (Apr.), 83–99.

PASHLER, H., AND HARRIS, C. R. 2001. Spontaneous allocation of visual attention: Dominant role of uniqueness. *Psychonomic Bulletin & Review 8*, 4, 747–752.

VIDAL, M., BULLING, A., AND GELLERSEN, H. 2013. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proc. of UbiComp '13*, ACM, New York, NY, USA, 439–448.

Paper 10: A Wrist-Worn Thermohaptic Device for Interruption

Title of Paper

A Wrist-Worn Thermohaptic Device for Graceful Interruption

Authors:

Frank Bolton, *Shahram Jalaliniya*, Thomas Pederson

Published:

Interaction Design and Architecture(s) Journal, 2015

Abstract

Thermal haptics is a potential system output modality for wearable devices that promises to function at the periphery of human attention. When adequately combined with existing attention-governing mechanisms of the human mind, it could be used for interrupting the human agent at a time when the negative influence on the ongoing activity is minimal. In this article we present our self-mitigated interruption concept (essentially a symbiosis of artificial external stimuli tuned to existing human attention management mechanisms) and perform a pilot study laying the ground for using a wrist-worn thermohaptic actuator for self-mitigating interruption. We then develope a prototype and perform an insightful pilot study. We frame our empirical thermohaptic experimental work in terms of Peripheral Interaction concepts and show how this new approach to Human- Computer Interaction relates to the Context-Aware-systems-inspired approach "Egocentric Interaction" aimed at supporting the design of envisioned Wearable Personal Assistants intended to, among other things, help human perception and cognition with the management of interruptions.

A Wrist-Worn Thermohaptic Device for Graceful Interruption

Frank Bolton[1], Shahram Jalaliniya[2], and Thomas Pederson[2]

[1]Department of Industrial Engineering, Faculty of Engineering, Tel Aviv University, Tel Aviv, 69978, Israel. frankbolton@post.tau.ac.il
[2]Pervasive Interactive Technology Lab, IT University of Copenhagen, Rued Langgaards Vej 7, 2300 København S, Denmark. {jsha,tped}@itu.dk

Abstract. Thermal haptics is a potential system output modality for wearable devices that promises to function at the periphery of human attention. When adequately combined with existing attention-governing mechanisms of the human mind, it could be used for interrupting the human agent at a time when the negative influence on the ongoing activity is minimal. In this article we present our self-mitigated interruption concept (essentially a symbiosis of artificial external stimuli tuned to existing human attention management mechanisms) and perform a pilot study laying the ground for using a wrist-worn thermohaptic actuator for self-mitigating interruption. We then develope a prototype and perform an insightful pilot study.

We frame our empirical thermohaptic experimental work in terms of Peripheral Interaction concepts and show how this new approach to Human-Computer Interaction relates to the Context-Aware-systems-inspired approach "Egocentric Interaction" aimed at supporting the design of envisioned Wearable Personal Assistants intended to, among other things, help human perception and cognition with the management of interruptions.

Keywords. Thermal Haptics, peripheral interaction, notification, interruption.

1 Introduction

Historically, access to information beyond what you can see and hear, here and now, has been something only for the privileged. The printing press, diffusion of literacy and increased level of general education has together with information technology (e.g. mass produced books, TV, computers, smartphones) almost reversed the situation, leading to increasing situations of "information overload" where access to information hampers our actions and decision-making rather than simplifying it. Today, knowledge and resources are needed *to keep information out.* The conscious human mind is not made for concurrent tasks and generally works best when focus can be maintained on one thing at a time. A possibility increasingly diminishing because while the recent advent of mobile and wearable tools for communication has boosted the possibility of sharing information independently from place, it has also brought with it an increased risk for users to be interrupted at a bad moment such as when carefully prepared or focused work is being done.

The work presented in this article belongs to the set of efforts that aim at reducing unwanted interruptions by making the mobile/wearable devices that cause the interruptions "smarter". Inspired by recent work in peripheral interaction, our approach to graceful interruption management is based on the idea of a carefully designed symbiotic interplay between the existing interruption management infrastructure in our brains, and the digital devices that want to draw our attention. The rational behind this approach is that by partially offloading interruptability decision making to biology (the part of our brains that has evolved to do exactly that), the need for sensing and modelling the situational context of the human agent (e.g. ongoing and recently performed activities, nearby entities, nearby other human agents) would be reduced. While this implicit user interface design approach certainly faces its particular challenges due to how volatile, multimodal, and sensitive human attention is, the classical context-awareness approach e.g. to determine interruptability based on tracking and modelling physical phenomena in the vicinity of human agents have also proven to be very hard indeed. While the two approaches obviously complement each other, we do in this article focus on the challenge of interfacing to the human attention system and this through an as of yet very unexplored modality within Human-Computer Interaction: thermohaptics.

The haptic modality has some properties that makes it particularly interesting for graceful interruption and peripheral interaction in general, including the minimal interference with potentially ongoing everyday tasks (perceptionally, cognitively, socially) and the relative ease in which haptic information can be generated by wearable and mobile devices. We present a pilot experiment in which we investigated the feasibility of thermohaptics for graceful interruption, using a wrist-worn thermohaptic actuator through which stimuli with varying intensity were generated.

2 Interruptions

The term "interruption" has received various definitions. In this article, we adhere to Boehm-Davis and Remington's who define an interruption to be "the suspension of one stream of work prior to completion, with the intent of returning to and completing the original stream of work" (p. 1125) [1]. The link between peripheral interaction and interruptions caused by digital devices is strong. We find it reasonable that if the information associated with a specific notification is important and urgent, the interrupting notification should demand focused attention whereas a notification that represents an equally important but not urgent message should be delivered using a method that targets the periphery of attention. In the remaining parts of this section, we will highlight some effects that external interruptions from mobile and wearable devices can have in everyday life as well as the role modality, intensity, and timing of the interruptive stimuli plays.

2.1 An Increasing Problem

While interruptions caused by body-external events can be regarded as natural and unavoidable in a world were human agents switch between individual and collaborative activities as part of everyday life, there is no shortage in research identifying the negative side effects that interruptions can have. These include the negative effects interruptions have on a person's rate of performing tasks [2], their impact on the ability to drive safely [3], and an increase in associated accidents by healthcare providers [4,5]. Mark et al. found that workers compensate for interruptions by increasing work rate and maintaining work quality but experiencing an increase in stress, frustration, time pressure and increase in effort [6]. Much of this research findings stem from investigating subjects performing a focused activity, often with high risk of failure. In more open-ended settings, the findings of interruption studies are less clear. O'Conaill and Frohlich found that the interruptee benefits in some way from most interruptions and that the benefits offered are not balanced between interrupter and interruptee and are not consistent [7].

As mentioned in the introduction section, we believe that the increased use of mobile and wearable communication devices, apart from all the obvious positive effects, potentially threatens the well-being of people by the increased exposure to unwanted interruptions, resulting in for instance increased stress and stress-related diseases.

2.2 Social aspects of interruption

Audio-based mobile device interruptions can have a negative impact on the social context of the person receiving the interruption. Kern and Schiele use the concept of social interruptibility and give the following reasoning, "A notification does not necessarily reach the user only. An audio alarm can also be perceived by the environment; a potentially embarrassing situation, e.g in a lecture." (p. 3) [8] From Hansonn et al. we can see that "Auditory cues for mobile devices are typically designed to attract maximum attention to be able to penetrate even a very noisy sound environment. The notification in itself requires the recipient to, more or less instantly, direct her attention towards it. " (p. 2) [9].

Exteroceptic (body-external) haptic notification stimuli such as vibration or heat changes caused by devices in direct or indirect contact with the skin have clear social advantages over audio for notifications directed to single individuals by being perceivable only by the intended recipient.

2.3 Time

Interruptions (whether triggered internally or as in the case of our experiment, externally generated) can occur more or less frequently, resulting in "concurrent" multitasking when task switches are forced to occur every other second and "sequential multitasking" when task switching happens with hours in between [10].

Our approach to interruption management on envisioned future wearable devices is in practice an indirect manipulation of the point in time when the interruption stimuli is consciously noticed and causes a task switch. The point in time when an interruptive stimuli is noticed is hypothetically to be determined by on the one hand the thermohaptic stimuli intensity controlled by the notification system and on the other hand the human attention management system embedded in our brains which filters out unimportant stimuli.

While cognitive psychologists are still debating the exact inner workings of human attention, as designers of interactive systems we allow ourselves to treat it very much as a black box. In this article, we present our initial attempts in determining how this black box responds to variations in stimuli intensity and cognitive load and whether we can successfully incorporate its behaviour in future design of interactive systems.

3 Talking to human attention

It is clear that graceful interruption demands a more careful approach to human attention than the brute force override taken by most mobile systems to day (e.g. the loud audio-based ringing tone of cellular phones). As designers of interaction systems that perform interruptions, we need to take the human attention system more seriously.

Although our brain, body, and perception system operates in a highly parallel fashion, we consciously attend to only a very limited set of these processes. Much control of body and thought is automated and thus escapes our conscious mind. Human attention is indeed very much a “one thing at a time” phenomenon. However, since different body functions (e.g. the perception of different sensory modalities) are handled by different relatively independent parts of our brains, psychologists have found that perception, cognition, and action involving mutually independent control centres (cognitive resources) can indeed be successfully monitored and controlled by our brains in the grey-area between conscious and unconscious attention. Hausen shows a simplified model of human attention [11] based on divided attention theory [12], illustrating how various factors (including the availability of cognitive resources) influence whether human attention is directed towards a given stimuli or not.

Our sensory system seems to have a modality-based distribution over centres in our brain which allows for multitasking without reduced cognitive performance as long as, put simply, modalities (e.g. touch, vision) used in the different activities do not overlap [13]. Bakker illustrates our ability to distribute attentional resources over more than one task in parallel as long as none of them are too demanding [14]. Most attentional resources are allocated to the primary (center) task of preparing dinner but some are also devoted to secondary (peripheral) tasks: listening to the radio and monitoring the dishwasher; cognitive resource allocation.

3.1 Potato peeling + radio = true

Cognitive resource allocation is very dynamic and changes instantly as a result of for instance external stimuli changes. For instance, if the hypothetical preparing dinner task enters into a habituated/automated phase (e.g. peeling potatoes) the radio the radio program is more likely to substitute the dinner preparation as primary task. If the two tasks are swapped again at a later stage (the dinner preparation task becomes central and the radio listening task becomes peripheral), without negative impact on neither tasks, we have a win-win situation where the dinner both tastes great and was fun to prepare. The radio "interruption" was graceful by adding value without negatively impacting the primary task. It worked, in part, and according to Multiple Resource Theory [13], because the sound modality is not very important as feedback when peeling potatoes and (we assume) our protagonist to be a fairly avid and experienced potato peeler.

Peripheral interaction as proposed by Edge [15] and further developed by Bakker [14] and Hausen [11], is a design approach aimed at allowing for peripheral perception and/or manipulation of artefacts without significantly drawing attention from another primary task. As such, the peripheral interaction design approach is a promising and fresh take on the design of interactive systems that integrate themselves into everyday real world activities. Drawing from the interaction design discipline which traditionally tend to emphasise the shape and behaviour of single artefacts, most existing prototype designs are however limited to HCI dialogues based one a *single* (potentially multimodal) system that interacts with a user.

Coming from the Ubiquitous Computing discipline ourselves, and envisioning an increasing amount of digital services (from different manufacturers and service providers) that call for our attention, we have found it necessary to conceptualise future HCI as an interaction paradigm consisting of *many* artefacts (including everyday objects, digital devices, and services). This is the situation our "users" are facing and we might as well aim at modelling this somewhat chaotic situation also as system designers. The case for interruption management is a good example of an interaction design challenge that calls for this kind of holistic perspective.

4 Wearable Personal Assistants

Any interruption management mechanism that is implemented in a mobile or wearable system will be effective only to the degree it is aware of, and can control, the sources of interruptions affecting the person carrying/wearing the device. Just like today's mobile phones are used as central portal to sources of information on the Internet to provide various services, we envision Wearable Personal Assistants (WPAs) [16] to act as a central bottleneck through which the majority of digital services need to push any potentially interrupting notifications. In fact, modern smartphones can be considered to be early instances of WPAs, having started to offer certain explicit control to their users with respect to which digital services that should be allowed to make use of what notification modality and under which circumstance. We think that this is just the beginning of the development of future more advanced interruption management mechanisms.

4.1 A filter and a torch

Our long term goal is to build Wearable Personal Assistants (WPAs) that facilitate, for a given human agent, performing the activities s/he wants to perform in short and long term by suppressing currently irrelevant information from perception/cognition and by highlighting the relevant. Thus the WPA would need to be somewhat aware of and capable of influencing processes taking place both in the local physical (real environment) and in the digital domain on the Internet or local computing devices. Apart from this need to sense external context, it is clear that a good interface towards human attention processes is necessary as well. We think peripheral interaction theory and our own work on egocentric interaction are good conceptual stepping stones towards framing this kind of human-computer interaction dialogue that partially will take place without the "user" being consciously aware of it.

4.2 Perception-Cognition-Action Loop

To design wearable personal assistants (WPAs, to be discussed in detail in the next section) we developed a simple interaction flow model in the spirit of classical HCI dialogue models but with emphasis on the interplay between human perception, cognition and action on the one hand and the environment and the wearable assistant on the other (Figure 1) [16].

Unconscious Perception By and large, our perception of the world (pathway 2-4) and our perception of body state (arrow 5) is beyond our conscious control. The filters in Hausen's model are some of the mechanisms, symbolized by arrow 12 in the figure, that are at work. However, conscious cognitive processes influence unconscious processes (arrow 7), as in the case when we deliberately address our attention to a certain speaker in a crowd and we automatically (thanks to subconscious processing), to some degree, can single out the voice we are interested in. We can also consciously and indirectly affect unconscious processing by orienting our body sensors (e.g. vision) towards phenomena of interest (pathway 8-10-2-4).

Unconscious Cognition Human cognition is divided into unconscious and conscious processing (arrows 12 and 13 respectively in Figure 1), receiving input from sensors capturing in-body phenomena (e.g. proprioceptive information about limb positions; information for maintaining homeostasis) and from sensors capturing information from the external world. No world phenomena or in-body phenomena is subjected to conscious cognitive processing before having been unconsciously processed (pathways 2-4-6 and 5-6 respectively).

Action Human action is initiated and controlled by a mix of conscious and unconscious cognitive processes. An example of an activity mostly driven by an *unconscious* perception-cognition-action loop could be "walking" along a well-known road with no exposure to obstacles (pathway 2-4-9-10 & 11).

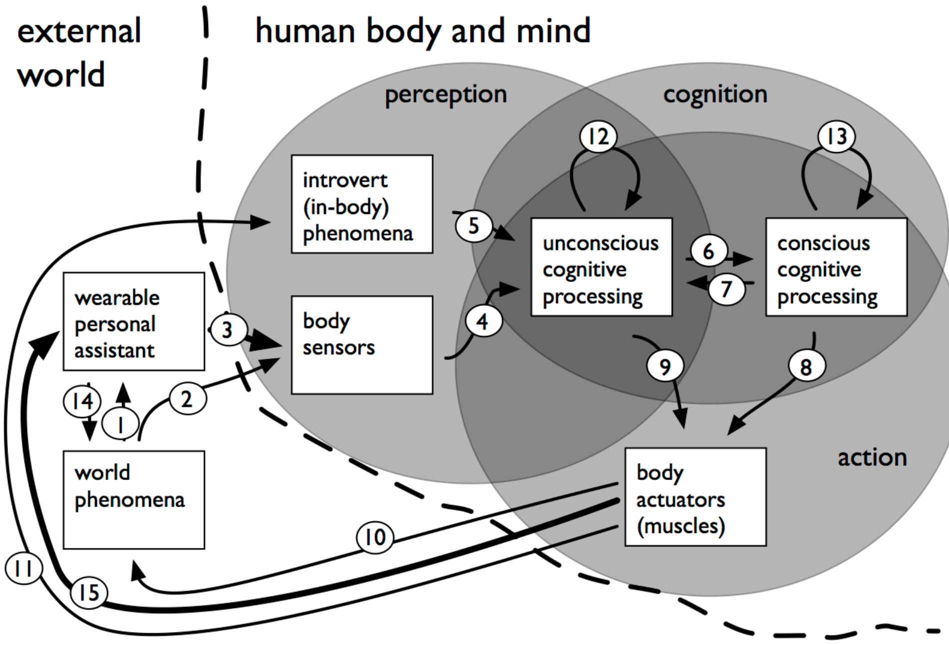


Fig. 1 A simple model of information flows occurring as result of a human agent acting in the world, partially supported by a wearable personal assistant. [16]

4.3 Self-mitigated interruption

There are many ways in which interruption management could be implemented in a wearable personal device such as a WPA. In this article we present our first investigation into using the thermohaptic modality for perceptually and cognitively graceful integration of notifications originating from digital services with ongoing tasks that the human agent is performing. The main idea behind the mechanism which we call "self-mitigated interruption" (see Figure 2) is to let the existing supervisory attentional system [17] play an important part in deciding whether a notification is to interrupt the current primary task or not, instead of primarily let the WPA rely on sensing and interpreting the context and then rather brutally call for attention. Note that we still treat the actual inner workings of the supervisory attentional system as a "black box" (see the discussion in section 2.3) and expect to learn a little bit about its behaviour as part of our experiment.

Of course, the self-mitigated interruption approach is modality agnostic. Our interest in the thermal modality is motivated by the fact that it is a modality rarely explicitly used in everyday activities and thus a potentially very useful information channel for interruption by not intruding directly into modalities that might already "be in use" [13].

4.4 Human haptic perception

Richter [18] distinguishes among three kinds of haptic perception: interoception, proprioception and exteroception. Interoception handles the state of the internal organs which we are most of the time not consciously aware of, driving for instance the feeling of hunger and pain [11]. Proprioception deals with the limb and digit joint position as well as the balance and orientation sense provided by the inner ear. Both interoception and proprioception are represented as “introvert in-body phenomena” in Figure 1. Finally, and of most relevance to us in this article, exteroception relates to stimuli coming from outside of the body and includes the sense of touch, vibration, and temperature differences. Haptic exteroceptic information is acquired by the human sensory system through perceptors sensitive to pressure and temperature changes occurring for instance when we manipulate objects with our hands [19].

4.5 Thermal haptics

Thermal haptics ties together psychophysics (the branch of psychology concerned with the perception of stimuli), engineering and Human Computer Interaction. This subsection introduces psychophysics, prior work in thermal haptics and a motivation for using thermal haptics in a wrist watch type device.

Early psychophysics research showed that human bodies do not sense temperature as thermometers do, but rather sense the change from neutral skin temperature [20]. The slightest change that a person performing a psychophysics experiment reports is called the just noticeable difference (JND).

Temperature changes are perceived by cold receptors and warm receptors. Cold receptors are more numerous than warm receptors by a ratio of up to 30:1, and respond to decreases in temperature over a temperature range of 5-43°C [21]. Warm receptors discharge with increases in skin temperature reaching a maximum at temperatures of around 45°C [22].

The JND changes as a function of rate of stimulus change [23] and with spatial summation [24]. A faster rate of change or a larger surface area allows for a smaller temperature difference to be sensed. JND also depends on the position of stimulation on the body with hairless skin (e.g fingertips and palms) being less sensitive than hairy skin, and areas on or near the trunk more sensitive than the extremities [25].

Psychophysics experiments are normally performed in a controlled setting with the participant dedicating time and attention to the experiment, this is unlike the highly variable set of contexts in which pervasive computing takes place. A factor largely ignored in thermal haptic notification research thus far is the level of cognitive when a thermal stimulus is generated. Arroyo and Selker used heat and light based ambient displays for providing interruptions. They found that heat was slower to sense but harder to ignore once sensed than light, that heat was more disruptive than light, and that heat could be perceived than invasive to private space but was able to communicate to a person privately [32].

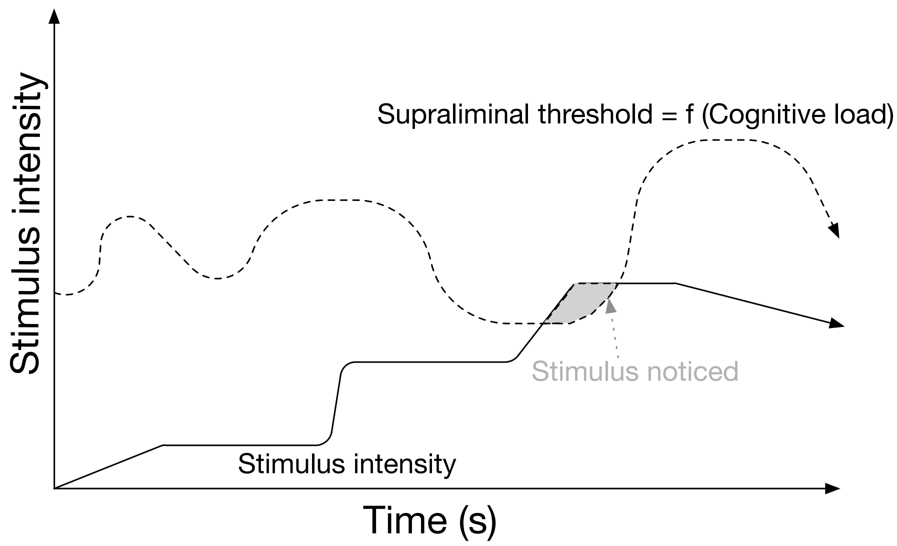


Fig. 2 Principle of self-mitigated interruption: The intensity of the stimulus (represented by the solid line) and the supraliminal threshold (represented by the dashed line), which is dependent on level of cognitive load, determines whether the stimulus gets consciously noticed. By increasing the stimulus intensity level, the likelihood of the stimulus being noticed increases.

5 Why a Wrist-Worn Thermohaptic Device?

While the haptic modality is abundantly used as part of explicit input to interactive computer systems (e.g. we *grab* the computer mouse and we *click* its buttons), haptics is less frequetly used in the other direction. Currently the most common use as a system output is probably as "silent ring tone" vibration on phones.

As pointed out by both Bakker [14] and Hausen [11], wearable interactive computer devices constitute a very interesting hardware platform for investigating peripheral interaction. While opening up the design space for graceful interruption mechanisms, wearable and mobile devices at the same time also contribute themselves to increase the risks for their carriers to become involuntarily interrupted due to the simple fact that these wearable devices are always there.

Wettach explored the use of thermal haptics for navigation and notifications, employing a resistive actuator that could only warm. The notifications worked well but navigation proved problematic [21]. Wilson et al. [25–28] and Halvey et al. [29–31] performed several studies using thermoelectric cooler based thermal haptic devices. Much of their work followed a psychophysics research approach (identifying level of detection, level of discrimination, etc) and they also included realistic device ecology considerations (how does it work while wearing clothing, outdoors, etc). Casio released a series of databank watches starting in 1980 with a wrist worn calculator, a dictionary and later with the 1984 release of a databank watch. The pebble smart watches broke new ground in 2012 when they introduced a consumer grade smartwatch to the market. This device connected to a smartphone and run host stand alone applications or act as a remote interface for applications that ran on the mobile phone. Various Android manufacturers (e.g. Samsung and LG) and Apple have also introduced smartwatches to their product lines. The history and the state of the art are evidence that the wrist is a most viable "home" for wearable computing.

Thermal stimulus was chosen for the prototype as it is a modality which allows for very subtle gradual changes (temperature takes time to change which makes it less startling than other modalities). Thermal stimuli are also completely private as it cannot be heard by somebody nearby the wearer (as is often the case with vibrotactile notifications).

Realistic scenarios where a wrist worn thermal notification could be valuable are for discrete notifications where the importance of the message is non-zero but also not very urgent or very critical (for instance, emails received during a lecture).

6 Experiment design

To investigate the feasibility of using a wrist-worn thermal haptic actuator for notifications, we produced a prototype and performed a pilot study where participants were also asked to respond to occasionally occurring thermohaptic stimuli.

The independent variables are: stimulus intensity, which is function of amount and rate of temperature change and direction (heating or cooling). In total, we have 18 different stimulus intensity for 2 directions (cooling and heating) × 3 change rates controlled by changing the fraction of time that the current is flowing (slow, medium, fast) × 3 temperature change amounts (0.5°C, 1.0°C and 1.5°C). The dependent variable is a binary variable indicating whether a particular stimulus was perceived.

To validate the prototype the following well-established psychophysics behaviours were tested: people are more sensitive to cooling than warming, people are more sensitive to high rate of temperature change than a low rate of temperature change, and lastly, that people are more sensitive to a higher overall temperature change than a low overall temperature change.

We recruited 15 volunteer participants (2 females) among local university students to participate in the experiment. All of the participants were right hand dominant. All of the participants completed the task within approximately 45 minutes.

6.1 Apparatus: wearable wrist worn thermal haptic prototype

We developed and evaluated a prototype wrist-worn thermal haptic system able to provide notifications to the person carrying it. The system consists of several components explained in this section, the wearable wristwatch device (Figure 3-B), the power controller (right hand side of Figure 3-A), and experiment software running on a computer.

Temperature differences are produced by the Peltier principle by passing electrical current through a Thermoelectric cooler (TEC).

Detailed system architecture for our system is shown in Figure 4 showing the components of the wearable device, the power controller and the computer. The power controller and wearable device are both connected to the computer.

Wearable wristwatch component The objective of this prototype was to realise wearable thermal haptic notifications whilst maintaining a somewhat realistic wearable device weight. The wrist device holds a microcontroller (Arduino ® Nano) and four thermistors (to measure the hot and cold sides of the TEC, the skin temperature of the person wearing the device, and the room temperature). Electronics are shown exposed in Figure 3-B and were covered with insulating tape for the experiment (Figure 3-A). Thermal contact with the skin is made by an anodised aluminium heat spreader. Residual heat is dissipated with a heat sink fitted to the back side of the Peltier. The microcontroller is programmed to be inherently safe with a safety feature to turn off the power if a thermistor wire break occurs or if the temperature exceeds 50°C or gets below 5°C.

We tried to minimise the size of the thermal haptic device to fit into a watch form factor; however, due to the physical limitations of the Peltier device and the power electronics, a large heatsink was required on the exposed side of the wearable device (black square in Figure 3-B).

Power electronics The temperature changes are produced using a thermoelectric cooler (described in Section 6.1). The amount heat flow is dependent on the magnitude of the current applied, to drive this a bidirectional current source is needed, this can be achieved using the H-bridge which is a switching device that allows a single power supply to drive an electrical load in both directions. To vary heat transfer rates, pulse width modulation (PWM) is used to control the magnitude of the current. PWM works by a switch being turned on and off, the ratio of time that the switch is closed (current flows) in a given period is varied to control the current to the load. The prototype power controller is shown on the right hand side of Figure 3-A.

Thermal Stimulus selection As explained in Section 4.5, a person's sensitivity to temperature change is a function of more than just the maximum or minimum temperature reached as. It is also a function of rate of temperature change, whether heating or cooling is performed, and the surface area of stimulation. In practice, the sensitivity to stimuli also depends on the contact pressure of the haptic device with the skin. It is also well known that there is between-subject variability in temperature sensitivity.

Initial device testing was using previously used values [25] to warm, and cool the skin temperature by 1°C, 3°C and 6°C, at two rates (3°C per second and 6°C per second).

During lab testing of the thermal haptic device the intensities were reduced to be appropriate to our wearable device (0.5°C, 1.0°C and 1.5°C). Additionally, it proved difficult to have a closed loop controller manipulate the temperature to have a consistent rate of temperature change as there was some heat transfer delay in the system. We decided rather to implement three fixed temperature change rates referred to in the rest of the paper as the heating rates "slow", "medium" and "fast", by manipulating the Duty Ratios for the heating and cooling (i.e. 38%, 50% and 77%), illustrated in the "power controller" in Figure 4.

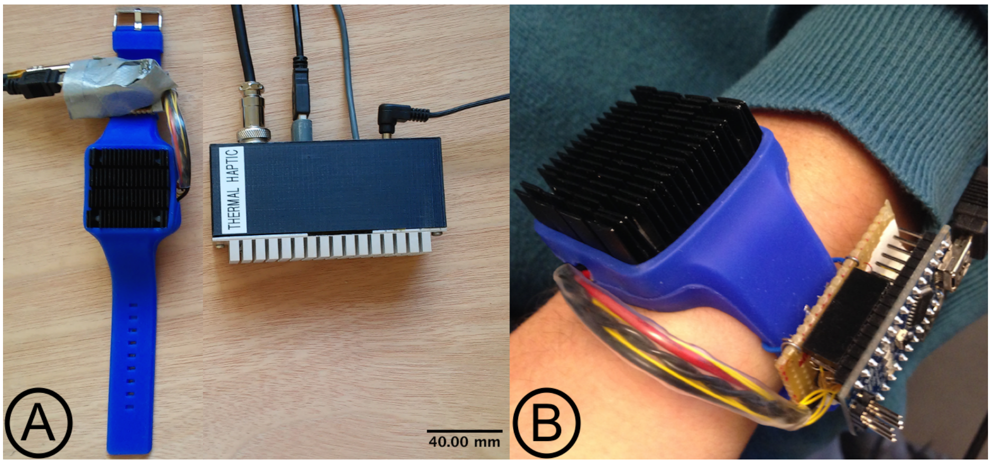


Fig. 3 A: Wearable thermal haptic notifier system. The wearable component is on the left and the power controller component is on the right, B: Thermal haptic notifier being worn.

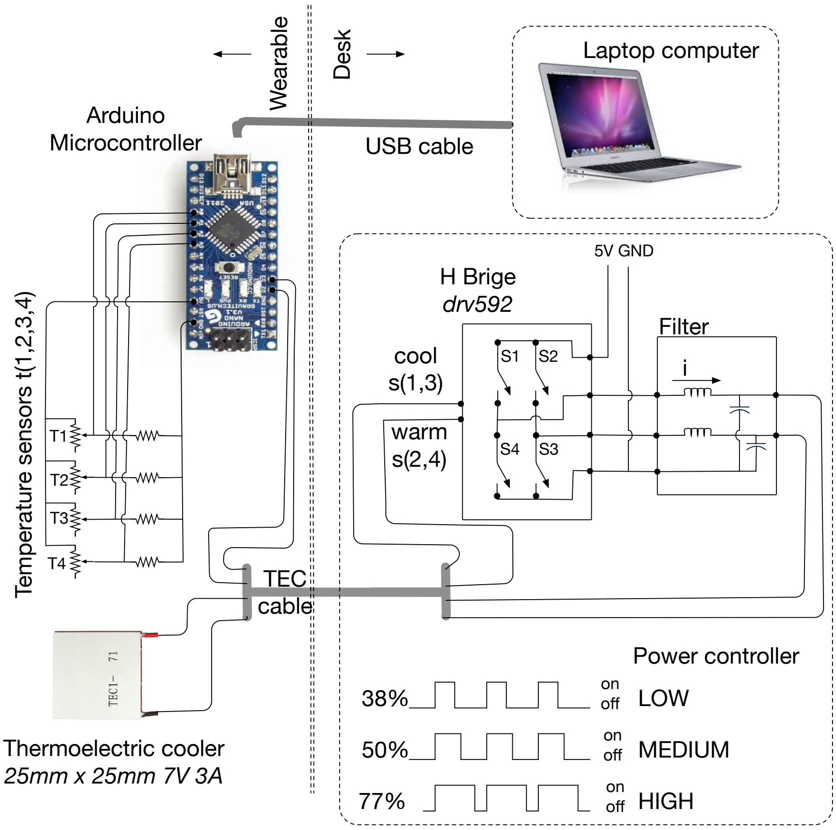


Fig. 4 System architecture diagram. The components of the wearable device are on the left side (Thermoelectric cooler, Arduino Nano, and 4 thermistor based temperature measurements. A simplified h-bridge is shown in the power controller block.

An Apple Macbook Air (2GHz i7, 8GB RAM, OS X 10.9.4) was used for the experiment. A Processing program generates thermal stimuli and logs keystrokes and temperature data.

6.2 Procedure

The experiment started with a short introduction to the task and to the use of the apparatus, but the main purpose of the experiment was not explained to avoid biasing the participants. The participants were told that they are participating in an experiment evaluating the thermal haptic wristband for notification purposes. After introduction the participants signed the consent form.

The experiment comprised three different sessions. The first session was training the participant to experience the thermal haptic stimuli. They were asked to wear the thermal haptic wristband on the non-dominant hand. This was a 5-minute training session where warming and cooling stimuli of different intensities were given along with a spoken cue from experimenter when the stimulus was presented. The objective of this is that the participant begins to understand the magnitude of the different stimulus types such that they are not startled or left unsure when sensing something during the experiment. After training session, they performed two 10-minute experimental blocks in which they were asked to respond to the thermal stimuli by pressing the space key. Only the findings from one of the experimental blocks is included in this article. The second experimental block contained a snake game and verbal communication with the participant for manipulation of cognitive load but was omitted from the study as this method was deemed inappropriate for our research questions after the experiment was concluded.

6.3 Design

The pilot experiment has a within-subjects design, each participant completed all conditions in one experimental session that lasted for approximately 45 minutes. In each condition, 18 stimuli were presented to the participants. The stimuli were presented in two different modes (cooling and heating), with 9 different intensities. The intensity of the stimuli was adjusted by 3 heating/cooling rates (slow, medium and fast), and 3 temperature changes (0.5°C, 1°C and 1.5°C).

7 Results

The data recorded in the experiment was in the form of a log file, every stimulus and keystroke was recorded. Post processing consisted of separating the stimuli and space-bar presses (participant senses thermal stimulus) from the rest of the data. If a participant pressed the space-bar within 10 seconds of the start of the stimulus, the stimulus was labeled as detected and as missed if no space-bar pressed occurred. Figure 5 shows the detection data for all of the participants. The variables are: direction (warm vs cool, between top and bottom sections), heating rates (left, middle and right plots) and temperature change (the vertical levels).

A repeated measure ANOVA is used to investigate the effect of each factor on number of detections. Post-hoc paired samples t-tests with a Bonferroni correction were used for pairwise comparisons ($\alpha = 0.05$).

Temperature changes: the number of detected notifications significantly varies with amount of temperature changes: $F(2, 14) = 5.354$, $p < .012$. The post-hoc analysis revealed that notifications with higher temperature change were detected more than notifications with lower temperature change.

Heating rates: the number of detected notifications significantly varies with the rate of temperature changes: $F(2,14) = 34.551$, $p < .0001$. The post-hoc analysis revealed that participants detected more notifications with faster changing rate compared to slower ones.

Direction: the number of detected notifications significantly varies with the direction: $F(1,14) = 146.689$, $p < .0001$. The post-hoc analysis revealed that participants detected more cooling notifications than heating notifications.

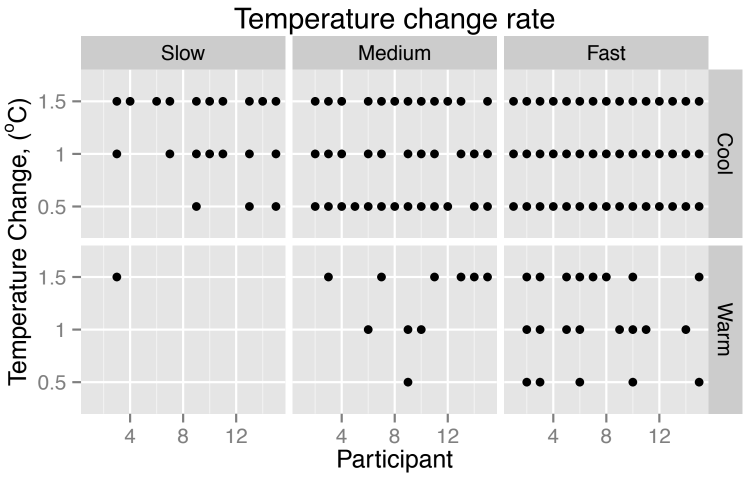


Fig. 5 Notification detection

8 Discussion & Conclusions

Trends were observed (Figure 5) where participants acknowledge more of the higher intensity stimuli than the lower intensity stimuli, and that cold stimuli were more effective than warming stimuli.

The results from our study indicate that for certain systems that generate potentially interruptive information, it might not be necessary to sense and process a large amount of contextual factors local to the receiver of a notification (the classical Context-Aware system design approach to graceful interruption) but rather to rely on previous knowledge about the specific wearers personal threshold for when a thermal stimuli transcends from being unconsciously perceived to consciously perceived (paths 3-4-9-10 and 3-4-6-8-10 in Fig. 1 respectively). In future studies, we will empirically investigate if sensitivity of human subjects to thermal stimuli varies with cognitive load. If we can find a sensitivity threshold for different cognitive load conditions, a context aware wearable device could rather rely on the human natural ability to discriminate between what is important and what is not, resulting in a mitigation of unwanted interruption.

In this article, we have presented the design and a first pilot experiment exploring the use of wrist-worn thermohaptic devices as mediators for graceful interruptions. We have grounded the proposed "self-mitigated interruption" approach in peripheral interaction theory and motivated it by the need to improve the situation for human agents in the emerging egocentric interaction paradigm where multiple devices (wearable, mobile, stationary) threatens to increasingly call for attention, on top of the focus demanded by current real world tasks. We have argued for why interrupting notifications from services should be handled in a centralised manner and exemplified a possible solution based on the concept of Wearable Personal Assistants. Further, we have argued for the use of wearable haptic actuators for notifications, in particular thermohaptic ones.

With respect to our empirical work, we can conclude that this initial exploratory experiment has validated our experimental setup and motivates us to perform a more focused experiment using our wrist-worn thermohaptic device in the future where we will reduce the number of factors to investigate, increase the number of participants, and calibrate for individual differences in stimuli sensitivity among participants.

With respect to theory development, we hope to have shown how existing peripheral interaction concepts resonate well with our own egocentric interaction vision, and that these conceptual frameworks have a huge potential in bringing (wearable) context-aware system developers and interaction designers together based on the huge overlap in interests.

Acknowledgements This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

References

1. Boehm-Davis, D.A., Remington, R.: Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. Accident Analysis & Prevention 41(5) (2009) 1124–1129
2. Adamczyk, P.D., Bailey, B.P.: If not now, when?: The effects of interruption at different moments within task execution. In: Proc. of the CHI '4, ACM (2004) 271–278
3. Teather, R.J., MacKenzie, I.S.: Effects of user distraction due to secondary calling and texting tasks. In: Proc. of the International Conference on Multimedia and Human Computer Interaction, International ASET, Inc (2013) 115.1–115.8
4. Grundgeiger, T., Sanderson, P.: Interruptions in healthcare: theoretical views. International journal of medical informatics 78(5) (2009) 293–307
5. Westbrook, J.I., Woods, A., Rob, M.I., Dunsmuir, W., Day, R.O.: Association of interruptions with an increased risk and severity of medication administration errors. Archives of Internal medicine 170(8) (2010) 683
6. Mark, G., Gudith, D., Klocke, U.: The cost of interrupted work: more speed and stress. In: Proc. of the CHI '8, ACM (2008) 107–110
7. O'Conaill, B., Frohlich, D.: Timespace in the workplace: Dealing with interruptions. In: Conference companion on Human factors in computing systems, ACM (1995) 262–263
8. Kern, N., Schiele, B.: Context-aware notification for wearable computing. In: Proceedings of the 7th International Symposium on Wearable Computing, IEEE (2003) 223-230
9. Hansson, R., Ljungstrand, P., Redstr¨om, J.: Subtle and public notification cues for mobile devices. In: Ubicomp 2001: Ubiquitous Computing, Springer (2001) 240–246

10. Salvucci, D.D., Taatgen, N.A., Borst, J.P.: Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. In: Proc. of the CHI '09, New York, NY, USA, ACM (2009) 1819–1828
11. Hausen, D.: Peripheral interaction - exploring the design space, PhD thesis (2014)
12. Wickens, C., McCarley, J.: Applied Attention Theory. CRC Press (2007)
13. Wickens, C.D.: Multiple resources and performance prediction. Theoretical Issues in Ergonomics Science 3(2) (2002) 159–177
14. Bakker, S.: Design for peripheral interaction, PhD thesis (2013)
15. Edge, D.: Tangible user interfaces for peripheral interaction. Technical Report UCAM-CL-TR-733, University of Cambridge, Computer Laboratory (December 2008)
16. Jalaliniya, S., Pederson, T.: Designing wearable personal assistants for surgeons: An egocentric approach. Pervasive Computing, IEEE 14(3) (July 2015) 22–31
17. Shallice, T., Burgess, P.: Supervisory control of action and thought selection. In: Attention: Selection, Awareness, and Control : a Tribute to Donald Broadbent. Oxford University Press (1993) 171–187
18. Richter, H.: Remote tactile feedback on interactive surfaces. PhD thesis, lmu (2013)
19. Caldwell, D.G., Tsagarakis, N., Wardle, A.: Mechano thermo and proprioceptor feedback for integrated haptic feedback. In: Robotics and Automation, 1997. Proc.., 1997 IEEE International Conference on. Volume 3., IEEE (1997) 2491–2496
20. Christensen, A.R.: Media, hot & cold— making sense of temperature in early psychophysics. International Journal of Communication 8 (2014) 5
21. Wettach, R., Behrens, C., Danielsson, A., Ness, T.: A thermal information display for mobile applications. In: Proc. of the 9th international conference on Human computer interaction with mobile devices and services, ACM (2007) 182–185
22. Jones, L.A., Berris, M.: The psychophysics of temperature perception and thermal-interface design. In: Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proc.. 10th Symposium on, IEEE (2002) 137–142
23. Kenshalo, D.R., Holmes, C.E., Wood, P.B.: Warm and cool thresholds as a function of rate of stimulus temperature change. Perception & Psychophysics 3(2) (1968) 81–84
24. Kenshalo, D.R., Decker, T., Hamilton, A.: Spatial summation on the forehead, forearm, and back produced by radiant and conducted heat. Journal of comparative and physiological psychology 63(3) (1967) 510
25. Wilson, G., Halvey, M., Brewster, S.A., Hughes, S.A.: Some like it hot: thermal feedback for mobile devices. In: Proc. of the CHI '11, ACM (2011) 2555–2564
26. Wilson, G., Brewster, S., Halvey, M., Hughes, S.: Temperatour, hot apps, cool widgets: thermal feedback for mobile devices. In: Proc. of the 14th international conference on Human-computer interaction with mobile devices and services companion, ACM (2012) 211–212
27. Wilson, G., Brewster, S., Halvey, M., Hughes, S.: Thermal feedback identification in a mobile environment. In: Haptic and Audio Interaction Design. Springer (2013) 10–19
28. Wilson,G.,Brewster,S.,Halvey,M.,Hughes,S.: Thermal icons: evaluating structured thermal feedback for mobile interaction. In: Proc. of the 14th international conference on Human-computer interaction with mobile devices and services, ACM (2012) 309–312
29. Halvey, M., Wilson, G., Vazquez-Alvarez, Y., Brewster, S.A., Hughes, S.A.: The effect of clothing on thermal feedback perception. In: Proc. of the 13th international conference on multimodal interfaces, ACM (2011) 217–220
30. Halvey, M., Wilson, G., Brewster, S.A., Hughes, S.A.: Perception of thermal stimuli for continuous interaction. In: CHI'13 Extended Abstracts on Human Factors in Computing Systems, ACM (2013) 1587–1592
31. Halvey, M., Wilson, G., Brewster, S., Hughes, S.: Baby it's cold outside: the influence of ambient temperature and humidity on thermal feedback. In: Proc. of the CHI '12, ACM (2012) 715–724
32. Arroyo, Ernesto, and Ted Selker. "Arbitrating multimodal outputs: Using ambient displays as interruptions." Human-Computer Interaction: Theory and Practice (Part II)-Proceedings of HCI International 2 (2003): 591-595.