TOWARDS DYNAMICALLY SWITCHING MULTIMODAL USER INTERFACES:

AN ARCHITECTURE FOR WEARABLE COMPUTERS

&

A FIRST EXPERIMENT

A thesis submitted in partial
fulfillment of the requirements for the degree of
Master of Science

by

Shahram Jalaliniya

January 2014
IT University of Copenhagen

Thesis Director
Dr. Thomas Pederson

ABSTRACT

Using mobile and wearable computers in a dynamic physical environment (e.g. while walking) implies extra cognitive and perceptual load, which can negatively affect the performance of human agents in real-world tasks. One of the strategies taken in previous work on wearable computing, context-aware computing, and multimodal interaction has been to deploy multimodal adaptation of the user interface based on the current physical context of the user. However, to the best of our knowledge, none of the existing solutions has thoroughly investigated the relationship between automatic and manual switching of interaction modalities. In this thesis, we propose an architecture for wearable computers intended to help mobile users handle the increased cognitive and perceptual load by facilitating modality and device switching. While implementing and assessing the full architecture is left for future work, we empirically investigated the optimal behavior of the adaptation component within the proposed architecture through a lab experiment. In the lab experiment, we simulated a typical mobile interaction scenario by designing a combination of physical and virtual tasks. The result of our experiment showed the fact that automatic adaptation of the UI when performing in a virtual task can significantly increase performance of the human agent in a physical task compared to the manual UI adaptation. However, some users preferred the manual adaptation due to the controllability challenge of automatic adaptation.

ACKNOWLEDGEMENTS

First, I would like to express my deepest gratitude to my supervisor Dr. Thomas Pederson for his insightful comments and engagement through the learning process of this thesis.

Furthermore I would like to thank my beloved family, Banafsheh and Sheida, for their kindness and support throughout the entire process of writing this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

As computing devices become smaller, people are increasingly carrying and using them while moving. However, most mobile systems do not support interaction in motion [3], and users need to stop to interact with their mobile devices. One of the main reasons, which makes user stop and interact with a mobile device is the fact that most of the mobile devices provide static channels for interaction with user. Touchscreen is the main channel of interaction between user and mobile devices. To perceive the information displayed on the touchscreen, users need to devote their visual attention to the mobile devices which means that they need to share their visual perception between the real world and the mobile devices. Also to provide input to a mobile device through touchscreen, at least one hand needs to be dedicated for holding and touching the mobile device. Apart from physical and perceptual challenges of interaction in motion, there are some other aspects associated with mobile interaction such as cognitive load. Even if mobile users do not need to dedicate their visual attention to their mobile devices, interaction with computing devices through other channels such as auditory modality can still raise cognitive problems. That is the reason why in many countries talking to the mobile phone while driving is forbidden.

In HCI community, several research areas have targeted to support mobile interaction. One of the main goals of developing wearable computers has been miniaturizing electronics and integrating computing devices into the clothes to overcome the physical constraints of mobile interaction [8]. Another complementary approach to mitigate the physical challenges of mobile interaction is to develop novel hands-free, eyes-free, and touch-less interaction techniques for mobile users.

Also, supporting mobile users to interact through several modalities such as speech, gesture, and gaze (multimodal interaction) in different situations can decrease the perception and cognition problems of mobile interaction. Another classic strategy to support mobile users is minimizing the need for explicit input through developing context-aware systems.

The main idea with this master thesis is to design an adaptive egocentric interaction [19] middleware to support mobile interaction. The egocentric interaction paradigm is a body centric approach to move from traditional device centric model of interaction. The adaptive egocentric interaction middleware is a wearable platform that allows higher-level applications to easily switch interaction between different modalities in different situations based on contextual cues. This enables the computer system to not block an important modality from being used in interaction with the real world by choosing system input and output modalities such that there is minimal overlap between modalities used for interaction with the system, and modalities used for interaction with the real world.

1.1 Research questions

Using mobile and wearable computers in a dynamic physical environment implies extra cognitive and perceptual load, which up until now, few systems have offered support in handling or reducing. One approach to reduce the negative effect of interacting with mobile/wearable devices on real-world tasks is multimodal adaptation of the user interface. But regarding the dynamicity of the user context and physical environment in mobile interaction scenarios, design and implementation of such mobile/wearable adaptive systems is challenging.

- ***Research question 1***: *How can mobile/wearable computers be designed to support mobile interaction through dynamic multimodal adaptation of user interface?*
- ***Research question 2:*** *given the recognized usability challenges of adaptive user interfaces (such as violating predictability, controllability, and transparency)[78], what adaptation strategy (manual, automatic, or something in between) is the best in order to provide the most effortless interaction channel to the users?*

1.2 Goal

- Design an architecture for wearable computers that help users handle the increased cognitive and perceptual load implied by using the system in a mobile and dynamically changing setting.
- To empirically gather requirements for the optimal behavior of the adaptation component within that architecture which is in charge of determining the best combination of input and output modalities to use at any given point in time in order to provide the most effortless interaction channel between the user and the system.

1.3 Approach

The study starts by reviewing literature on mobile interaction, wearable computing, multimodal interaction, and context-aware user interfaces to identify the interaction problem of using wearable devices in dynamic contexts and to learn about existing methods. Based on the findings from literature and design space analysis, we design an architecture for wearables to mitigate the interaction problem through multimodal adaptation of the user interface. In the next step, we design and conduct a lab experiment where the behavior of the adaptation component is investigated with the purpose to inform the implementation of that component and the rest of the architecture (as future work).

1.4 Organization of this Thesis

This chapter of this thesis was a brief introduction to the whole document explaining the main research problem, scope of the research project, and the research method. The second chapter reviews the main concepts and state of the art in the research areas of wearable computing, context-aware computing, and multimodal interaction. In the third chapter, we develop and describe an architecture for wearable computers to support dynamic multimodal adaptation of user interface based on context of use. The implementation of the experiment platform is explained in the fourth chapter. We reflect on the method and results of the experiment in the fifth chapter, and the sixth chapter concludes by discussing the main findings of the thesis and proposes future directions.

2 RELATED WORK

In this chapter, an overview of the key concepts and a review of the relevant previous studies are explained. As mentioned in section 1.2, this thesis targets designing an architecture for wearable computers to support mobile interaction in dynamic contexts. The most relevant research areas to the topic of the thesis are wearable computing, multimodal interaction, and context-aware computing (illustrated in Figure 1). The first step of this study is reviewing the literature in these areas and building a theoretical infrastructure based on the previous studies.

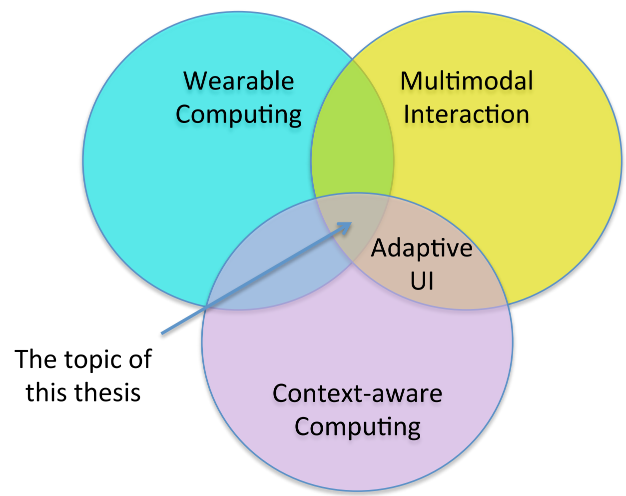

Figure 1. Position of the research topic

2.1 Key Concepts

In this section, the main relevant research areas are briefly described. The first concept described in the following sections is "mobile interaction" that is explaining the problem domain. The other concepts are "wearable computing", "context-aware computing", and "multimodal interaction" which are belonged to the solution area. Finally the concept of "egocentric interaction", as our theoretical approach in this study is explained. Reviewing this section helps readers of this thesis to have a common understanding of the main concepts.

2.1.1 Mobile Interaction

There is no generally accepted definition for "mobile interaction" in the HCI community. Taking a user-centric approach, mobile interaction could be defined as interaction between computers and mobile (or nomadic) users [1], while taking a device-centered view, mobile interaction could be described as interaction between users and mobile devices, which is more addressed in the Mobile HCI community [2]. There is also a third definition for mobile interaction, which is "interaction in motion" [3] or interaction in the go [4].

2.1.1.1 Interaction between computers & mobile users

Based on the user-centric concept of mobile interaction, which has been more investigated under the nomadic computing umbrella, the main requirements of nomadicity is defined as providing capabilities and services to the nomad as he/she moves from place to place in a transparent, integrated and convenient way [1]

2.1.1.2 Interaction with mobile devices

In the device-centric definition of mobile interaction, challenges arising from small screen, short battery life, network volatility, and limited memory and processing power have been addressed. [2]

2.1.1.3 Interaction in motion

The main focus of studies on interaction in motion is solving challenges of interaction with a mobile device when user is moving (walking, running, biking, driving or doing other forms of motion). Marshall and Tennent [3] have classified the main challenges of interaction in motion into four main categories:

- **Cognitive load:** humans are not good in multitasking which means people are able to attend to a limited number of things at the same time. That is even though people can see or hear when they are driving, walking, or biking; it is not easy to attend to both at the same time.
- **Physical Constraints:** because of the special figure of the human body it is hard to interact with touch-screen mobile devices in many forms of movement such as running or biking. When we are running, because of body movement it is hard to see what is on the screen, or when we are biking our hands are not free for interaction.
- **Terrain:** physical terrain can change on a short or long term basis because of changes in the light level, water level on the road, iciness, traffic and other obstacles. The dynamic nature of physical terrain requires user attention and sometimes physical effort to traverse.
- **Other people:** interaction in motion has also some social aspects for users. For example when users wants to interact with their mobile devices in a crowded area they need to take care of people passing by.

In this thesis, we used the third definition of “mobile interaction” which is interaction in motion or interaction on the go.
Also the concept of mobility could imply different levels of granularity. In the CSCW community, mobility falls into three main categories:

- *Micromobilty*: is about mobilizing physical objects (e.g. a paper document) for collaboration purposes [6]
- *Local mobility*: is defined as mobility within a building (i.e. in a hospital) in which people usually walk between different rooms or use elevator to move between different floors [5].
- *Remote mobility*: is defined as moving around different physical locations (e.g. construction sites) [6].

In this thesis, the granularity of mobility can be either *local* or *remote* mobility.

2.1.2 Wearable computing

Wearable computers are envisioned to be small and portable computing devices that can be easily worn by mobile users to support them without interrupting their activities in real world. Wearable computers are usually integrated into the user’s clothes or could be worn continuously like a wristwatch or glasses [7].
Steve Mann defines the concept of wearable computing as follows:
“*Wearable computing is the study or practice of inventing, designing, building, or using miniature body-borne computational and sensory devices. Wearable computers may be worn under, over, or in clothing, or may also be themselves clothes*” [8]
Wearable computers are supposed to run always in the background to facilitate users’ interaction with their environment and provide access to information anytime and anywhere. In other words wearable computers can act like humans’ prosthetic to extend humans’ mental or physical abilities. While the term human-computer interaction emphasizes the separation of computer and human, a wearable computer could be considered as the second brain of the wearer and its sensors as additional senses, which could be merged with user’s senses [9].
According to Starner [10], “wearable computing pursues an interface ideal of a continuously worn, intelligent assistant that augments memory, intellect, creativity, communication, and physical senses and abilities.”
Since users of wearable computers are able to move from place to place, the context of the use changes frequently. Therefore, one of the main attributes of a wearable computer is its context-awareness and ability to adapt to the context of use.

2.1.3 Context-aware computing

Through emerging mobile distributed systems, access to the information, communication, and computation is becoming ubiquitously available. Users of these mobile systems move from place to place and join or leave groups of people while using computing devices. The new context-aware mobile systems are able to detect changes in environment to adapt their user interface or behavior accordingly. [11]

"By context, we refer to any information that characterizes a situation related to the interaction between humans, applications, and the surrounding environ- men" [12]. Location and identity are two common pieces of context, which have been mostly used in different applications, but context is not just about location and identity. Context could include any kinds of knowledge about the environment in which the system is run. Identity of the user and other people than the user, location, time, user's activity, physiological data of the user (body temperature, heart rate, skin conductance, etc.), level of noise in the environment, light condition, and many other pieces of information could be part of context. [13]

From users' point of view, interaction with context-aware systems could be perceived in three different modes: [14]

1- *Personalization*: where the application let user set how the application should behave in different situations.
2- *Passive context-awareness:* where system does not change the interface or behavior of the application and let user specify how the application should change, if at all.
3- *Active context-awareness*: when system autonomously changes the user interface and behavior of the application based on the detected context.

From mobile interaction point of view, each mode of system adaption could be meaningful in different situations. For example, when a user is not able to use his/her eyes for interaction offering visual and auditory outputs as two different options is not relevant; therefore system can choose the non-visual output autonomously. On the contrary, when user has the possibility of interacting through both visual and auditory devices, system can ask user to select preferred modality.

2.1.4 Multimodal interaction

The term modality has been used in many different contexts and disciplines. For example, in psychology *sensory modalities* have been defined as vision, hearing, etc. but in HCI the term modality has been used differently. One of the first papers in HCI that defined the term modality as a way of interaction was Bolt's study on advantages of using a combination of speech and gesture for providing input to the system [15]. Human-computer interaction means "exchange of information with computer systems" [16]. When human exchanges information, the information should be physically instantiated in some way such as sound, light, etc. Tzovaras [16] defines a modality, as "a way of representing information in some physical medium. Thus, a modality is defined by its physical medium and its particular way of representation*."*

Information should be instantiated in one or more of the physical media listed in Table 1 if it needs to be communicated to humans.

	Information carrier	Perceptual sense	Medium
1	Light	Vision	Graphics
2	Sound	Hearing	Acoustics
3	Mechanical touch sensor contact	Touch	Haptic
4	Molecular smell sensor contact	Smell	Olfaction
5	Molecular taste sensor contact	Taste	Gustation

Table 1. Five physical media for information [16]

"*During interaction, the user produces input modalities to the system and the system produces output modalities to the user. A multimodal interactive system is a system that uses at least two different modalities for input and/or output. And a unimodal system is a system which uses the same single modality for input and output*."[16]

Multimodal interfaces can be advantageous by increasing the robustness of interface, preventing errors, facilitating error correction by users, and offering alternative communication channels to the user in different situations. [17]

Some researchers believe that multimodal interfaces could be constructed from unimodal modalities while some others define multimodality as creating entirely new ways of interaction. According to the first view, the unimodal modalities could be combined linearly, or in some cases we could expect to see non-linear effects. In the linear combination of modalities we can expect that different modalities add up their strengths. Modalities could be combined as complementary, redundant, or alternative. [16]

2.1.4.1 Modalities complementing each other

A good example of complementary combination of modalities is the "put that there" system [15]. In this system the user can point to the items on a large display and select or move them by vocal commands. In this example, the combination of gesture and speech gains strengths of both modalities without significant unexpected side effects. In fact sometimes combining modalities can lead to unexpected results. For example, in a graphical UI including graphs and complex textual data, replacing the visual modality with audio is not always satisfactory for user.

2.1.4.2 Modality redundancy

Sometimes because of the importance of some information it worth to represent the same information in different modalities i.e. using both visual and acoustic alarms in a factory for safety reasons. Also in these cases we do not expect significant implications.

2.1.4.3 Modality replacement (switching)

In some situations, one modality or a combination of more modalities is replaced by other modalities to represent the same information. For example, a written text on the display can be heard by blind users.

There are lots of examples for combining unimodal modalities without significant unexpected side effects. However, in many other cases, due to the complexity of modality choices, we need to evaluate the usability of any kind of novel combination of modalities before using the multimodal interface in real world applications. For instance, a simple modality replacement could be using voice interface for an email system. At the first glance, it seems the user can speak to the system to compose new emails or listen to the received emails, but having an overview of the received emails using audio modality takes a lot of time and seems unnatural. Therefore, in some cases we observe a non-linear effect after replacing or combining modalities, which makes switching between different modalities more complex.

2.1.5 Egocentric interaction

When we are designing for interaction in motion, the first design approach that comes to mind is a user-centric or even better a human body-centric approach. Because when user moves from place to place, different parts of the body get involve in real world activities and cannot be used for interaction with computing devices. While focus of the most classical HCI models is designing interface for computing devices, the egocentric interaction paradigm focuses on human body as the center of design.

The *egocentric interaction* [18] term versus *exocentric interaction* has been coined first by virtual reality community. In egocentric virtual reality systems the user interfaces from inside the environment while in exocentric systems the user interacts from outside (God's eye viewpoint).

2.1.5.1 Principles

The egocentric interaction concept has been developed later [19] as an interaction paradigm to support design of mixed-reality systems. The main principles of the egocentric interaction paradigm are as follows [20]:

1- **Situatedness:** humans have their own perspective of the surrounding environment. They are situated within a local place and with their body and can perceive or manipulate a particular set of artifacts.
2- **Attention to the complete local environment:** usually human agents attend to the whole local environment of interaction, even if they are interacting with limited objects or a single system.
3- **The proximity principle:** proximity plays a significant role in determining what can be done, and what agents want to do. The closer objects or people have a bigger chance to grab user's attention.
4- **Changeability of environment and agent-environment relationship:** paying attention to the users body movement and rearrangements and modifications of the environment.
5- **The physical-virtual equity principle:** paying equal attention to virtual and physical objects, circumstances, and agents.

2.1.5.2 Terminology

- *Human agent* instead of user: In the egocentric interaction paradigm, the human individual needs to be viewed and modeled as a *human agent* that can move in a mixed-reality environment, not as a user having a dialogue with compute systems.[20]
- *Action* and *perception* instead of input and output: the concepts of input and output devices are substituted with "*action*" and "*perception*" of human agent. [20]
- *Virtual Objects* and *Mediators* instead of Interactive Devices: In the egocentric interaction view, input and output devices can be redefined as *mediators* through which *virtual objects* are accessed [20].

2.2 Literature Review

As we mentioned before, the main goal of this thesis is to support interaction in motion via a wearable and multimodal system, which is able to detect different situations and switch to appropriate modalities. This section reviews previous studies on interaction with wearable computers, input and output modalities for wearable computers, adaptive user interfaces, and multimodal interaction for mobile users. The section of "Interaction with wearable devices" reviews the alternative design concepts to the desktop metaphor for wearable computers. By reviewing the new mobile interaction concepts, we will understand the characteristics of interaction in motion. Also review of the wearable input and output modalities gives us an overview of the different modalities that should be switched or combined in different situations to support mobile interaction.

2.2.1 Interaction with wearable devices

The main assumption of the traditional desktop metaphor is that users are sitting in front of stationary computers and can devote all of their perceptual resources to interaction with computing devices. On the contrary, users of wearable and mobile devices are usually moving, and it is not easy for them to devote completely their attention to the interaction with computing devices. Based on the previous studies on alternative interaction metaphors to the desktop metaphor, we extracted the main interaction concepts as follows:

- *Personal assistant* [21]: in this metaphor the wearable system plays the role of a personal assistant for wearer. A good wearable assistant system is unobtrusive and can predict and prepare the information might be needed by user.
- *Personal information space* [22]: another less explored design concept for wearable user interfaces is presenting information on a body-spatialised displays. In this approach users always have their information available and fixed relative to their body position. This information representation changes when users change their viewpoint orientation, but not as they change position.
- *Tele-presence (team collaboration)* [23]: one of the most explored application areas of wearable computers is remote collaboration and tele-presence to support geographically distributed team workers. In a typical tele-presence system a local user carries a wearable computer with a head-mounted display and a head-mounted camera. A typical setting for mobile tele-presence systems includes, on the "local" side (the location where someone needs support), a head-mounted display, a head-mounted camera that captures the field of view of the wearer,

and a small wearable computer connected wirelessly to a remote computer [24, 25].

- *Microinteraction* [26]: is defined as an interaction with a device that takes less than four seconds. Microinteractions are interesting for wearable and mobile devices because they minimize interruptions. They allow users to have a fast interaction with computing devices; so that the user can quickly turn back to the task at hand. It seems this interaction concept is drawing more attention by emerging new generation of unobtrusive wearable devices with limited amount of processing power and battery life such as Google Glass[1] and Vuzix Smart Glass[2].

From the above mentioned interaction concepts, the main characteristics of the wearable systems can be defined as *predictive*, *egocentric*, *collaborative*, and *non-interruptive*.

2.2.1.1 Input modalities

Text input

By increasing availability of wireless networks, nowadays most of the mobile and wearable devices are continuously connected to the Internet and can use online speech recognition services as alternative to keyboards for text input. Moreover, the accuracy and performance of the speech recognition systems has been significantly improved. While speech recognition is still a secondary alternative for text input on mobile devices, some of the new wearable devices such as Google Glass are using speech as the primary text input method. However, in many applications such as messaging or email, keyboard is still the best device for text input. One of the most well-explored wearable keyboards is Twiddler (Figure 2) [27, 28, 29], which provides a chord-based one-handed text entry. Twiddler has 12 keys and one joystick that could be used to control mouse pointer. By pressing a combination of different keys at the same time, user can type 101characters. Twiddler is recognized as a wearable keyboard because it fits neatly in the wearer hand and does not fall when user opens the hand. It is also possible to type eyes-free, which is really important in mobile interaction. Studies have shown that users can reach the speed of 47 words per minute after 400 minutes of training [29]. While a very skilled typist can type 120 words per minute [30], and the maximum speed of the speech to text entry has been reported 107 WPM [31] but correction takes over three time as long as entry time. Some other chording devices have also been proposed such as Frogpad3, which is a half-sized keyboard or data glove [32] in which the keys are placed on the fingers.

---

[1] http://www.google.com/glass
[2] http://www.vuzix.com/consumer/products_m100.html
[3] www.frogpad.com

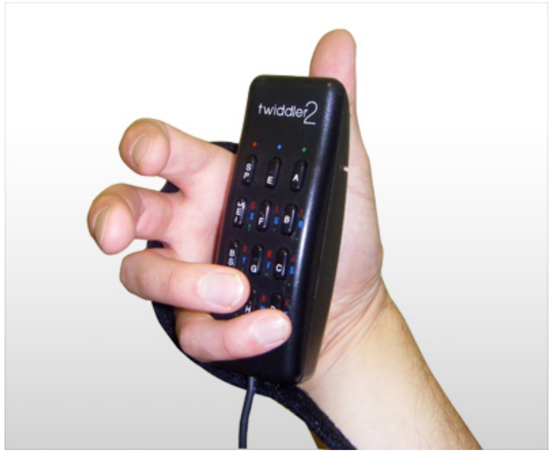

Figure 2. The Twiddler chording keyboard

Command input

Besides text input, in many applications users need other ways of interaction to control the application. One of the most intuitive modalities to send commands to the system is speech; however, the speech recognition is challenging in noisy environments. Another well-researched modalities to dispatch command to the wearable systems is through body gesture. So far several data gloves have been developed to detect hand gestures of the users [33, 34, 32]. In addition to the data gloves, body gestures could be detected either with a camera-based approach [35] or with body-worn sensors [36, 37]. In [35] a wearable Gesture Pendant has been developed which is able to detect hand gestures of the wearer with a camera. Gesture Wrist [36] uses a combination of accelerometer and capacitive sensors to detect hand gestures of the wearer, and in [37] a wristband style inertial sensor with accelerometer, gyroscope, and magnetometer has been used to detect six different hand gestures. In [38] electromyogram signals from wearable sensors have been used to detect hand gestures. Mardanbeigi et al. [39] have developed a head gesture detection method using a wearable gaze tracker. One of the main challenges of gesture-based interaction techniques is lack of a reliable mechanism to start and stop gesture detection, which can lead to unintended commands to the system.

Implicit input (context)

"*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.*"[61]

To provide implicit input to wearable computers, the sensors should continuously acquire the data from the context of use, and the context recognition algorithms should interpret the data to a higher level of abstraction, which is understandable by other applications. A typical architecture of a context-aware system is shown in Figure 2.

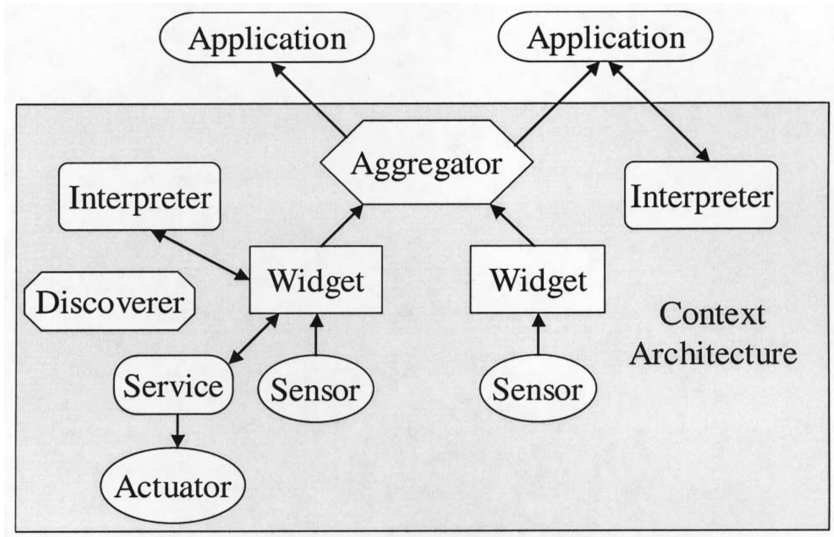

Figure 3. A typical configuration of context-aware system [12]

The implicit input can be highly valuable for interaction in motion since they limit need for explicit inputs by users. In addition, the contextual data could be used to adapt the input devices to the current situation. For example if the environment is noisy and the wearable system works on the speech control mode, the interaction mode can be switched to another modality. And finally, the context could be used to reduce the selection of information that could be useful for user in the current situation [40].

Other input modalities

Aside from recognized input modalities there are other modalities to provide input to the wearable systems. For example, Finger Mouse and Ring Mouse are two commercial forms of tangible wearable pointing devices enabling a mobile user to point on a screen like an ordinary mouse. Another example of tangible input modality is earPod [41], which provides eyes-free menu selection input to wearable and mobile devices.

2.2.1.2 Visual output modalities

Head-mounted displays

Head-mounted displays (HMD) 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 virtual world 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 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 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 HMD). In the other type of see-through HMDs (video see-through), a head-mounted camera streams the real-world image, which is combined electronically with computer-generated image. In Figure 4 different types of head-mounted displays are shown.

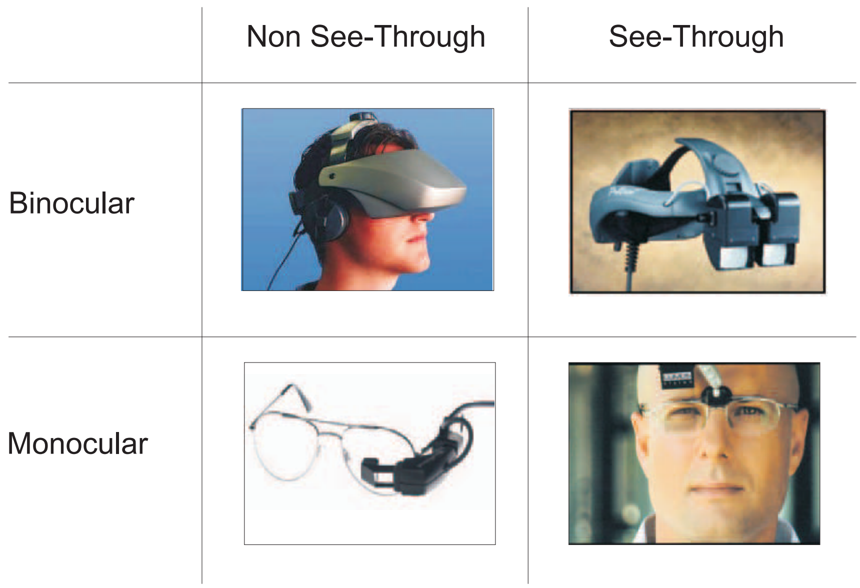

Figure 4. Different types of head-mounted displays [73]

In general, monocular HMDs are easier to use for wearer compared to the binocular ones due to their less weight [42]. They also offer the free eye 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 wearer of the monocular HMDs [43]. 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 [44, 45].

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, which are usually kept in 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. [46]

Wearable projectors

By emergence of pocketsize Pico projectors, the vision of augmenting the physical world with interactive projection came true [49]; however, the concept of wearable projection has been studied earlier using bulky projectors [47]. Interactive projection In the Sixth Sense project by MIT Media Lab [48] a gestural interface using a wearable projector and a camera was developed and evaluated (Figure 5). This system is able to see what the user is seeing and project information onto surrounding surfaces or physical objects. Omni Touch [49] system is another study on wearable projection by Microsoft Research Group. The Omni Touch 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 Omni Touch 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.

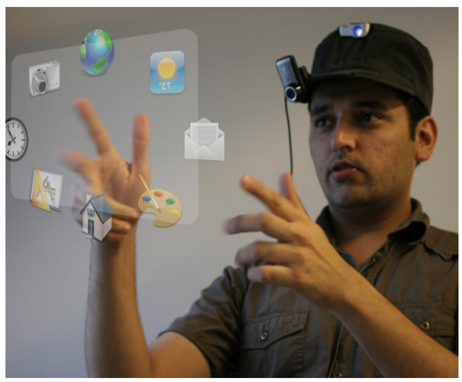

Figure 5. Sixth Sense: a wearable gestural interface [48]

One of the main advantages of wearable projectors is the possibility of sharing information in collaborative settings [50]; however, there are some limitations for projection technology such as contrast challenges in bright environments. In addition, the quality of the projected image depends on the texture and color of the background surface.

Wearable laser pointers

Wearable motorized laser pointer is another technology to superimpose information (e.g. sketches, text, point, etc.) onto the physical objects and surfaces around human agents. Stationary laser pointers have been used for augmented reality applications as an alternative technology to HMDs [51]. In [52], a combination of shoulder-mounted laser pointer and HMD has been evaluated to guide a local worker by a remote expert in a tele-guidance scenario. In this study, the shoulder-mounted laser pointer was used by a remote guide to direct the attention of the wearer to particular objects in local site. In another study [24] a helmet-mounted laser pointer was evaluated as an alternative to HMD in a tele-guidance scenario. The wearable laser pointer is able to superimpose polygons onto the physical environment.

Figure 6. The helmet- mounted motorized laser pointer [24]

In general, wearable laser pointers like wearable projectors have the advantage of intuitive augmenting physical environment compared to HMDs or other mobile devices. In projection-based augmented reality, the computer generated content superimposed directly onto the physical objects. But in HMD-based augmented reality systems, user needs to see the computer generated graphics and the real world image through the HMD. In fact the indirect view of the real world causes some problems such as eye fatigue and focusing problems for user of the HMD (more detail information is provided in the HMD section). However, existing motor-controlled wearable laser pointers are not unobtrusive enough (Figure 6) and suffer from lack of stability of the laser point [24]. Instability of the wearable laser point means that the projected content moves by moving the wearer's head. In remote collaboration scenarios, the instability of the pointer decreases the accuracy of pointing to the local side objects. Also in augmented reality applications, instability of the laser point increases the registration problem.

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 [53]. The first prototype of the contact lens display comprises a single pixel led, a miniaturized IC, and an antenna for electricity induction (Figure 7). This single pixel prototype has been tested on a live rabbit successfully.

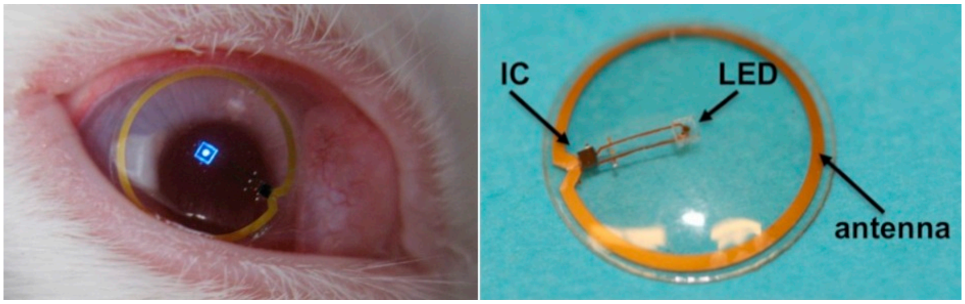

Figure 7 The first prototype of the contact lens display [53]

2.2.1.3 Auditory output modalities

Audio-based interfaces can offer eyes-free and in some cases hands-free input modality to the user which can be very valuable for interaction in motion. Generally, audio can be used for system output in two different ways [54]:

1- *Speech auditory interface,* which can use a recorded human voice like in an audio tour system or computer-synthesized voice such as automatic computer telephone system.
2- *Non-speech auditory interfaces* can use a combination of music and different sounds to communicate more complex information.

The main advantages of auditory output modality are as follows [55]:

- Sound can be heard all around the user independent of the position of its source.
- In a wearable system using sound in interface can reduce visual load.
- Sound can grab user's attention, and it is difficult to avoid hearing auditory stimuli compared to visual cues due to the omni directional property of the sound. User can easily avoid seeing something by changing the direction of the eye, but sound can be heard in all directions. Therefore sound can be used to grab user's attention for important messages.

Despite of the usefulness of the auditory output for mobile interaction, there are some limitations on using sound. The first weakness of sound is that it is transient which means once it was presented it is gone. This could be challenging for systems, in which users need to recall some of the information later. The other problem is the annoyance of the auditory feedback for users in some situations [55]. For example, using intensity, as a cue in sound can be annoying for the user, or using sound as output modality in collaborative settings generate noise for nearby colleagues.

2.2.1.4 Haptic output modalities

"A haptic device is a system generating an output which can be perceived haptically " [56]. In other words, "haptic devices generate a feedback to the skin and muscles, including a sense of touch, weight, and rigidity" [57]. One of the first attempts to develop a wearable tactile display was the rabbit system [58], in which three small speakers on the back of the wearer generated directional tactile patterns. The rabbit display was designed based on a perceptual illusion called the cutaneous rabbit, which implied the wearer something was crawling up their spine. "ActiveBelt" [59] is a wearable tactile-based navigation system providing user multiple directional information with the help of eight vibrators in the belt of user. "TactiGlove" [60] is another vibrotactile 3D navigation system integrated into a glove, which helps users find invisible or digitally annotated objects in 3D space (Figure 8).

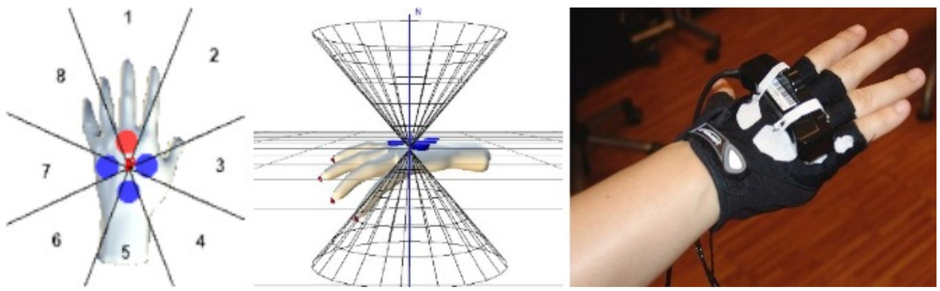

Figure 8. TactiGlove a vibrotactile 3D navigation system [60]

One of the important benefits of haptic output is integratability of haptic devices into the textile and clothes. The other advantage of the haptic output modality for mobile interaction is capability of eyes-free and silence communication with human agents; however, only limited information can be transmitted through haptic devices. In fact, the variety of the information that can be communicated through haptic output modality depends directly on capabilities and limitations of the human's touch sense. Humans can sense only thermal changes, mechanical forces, light, chemical substances, motion and vibration on the skin. Moreover, touch receptors are not distributed evenly over the human's body. Therefore, there is limited information that can be coded in the form of haptic output. Usually combination of haptic devices with visual or auditory displays forms a good compromise.

2.2.1.5 Implicit output modalities

Implicit output is a much less-explored concepts within the HCI community than explicit output. Schmidt [61] defines the implicit output as "the output of a system that is not directly related to an explicit input and which is seamlessly integrated with the environment and the task of the user." Based on this definition at least two kinds of modalities can be categorized as implicit output modalities.

Ambient displays

Ambient displays enable users to be aware of their background information at the periphery of their perception [62]. The concept of ambient displays is defined and developed in the ubiquitous computing research area based on Weiser's vision: " the most profound technologies are those that disappear "[63]. However, ambient displays in the environment can be a part of mobile interaction session to provide implicit output at the periphery of the human agent's perception.

Subliminal

Based on the definition of implicit output, directing user's attention by providing subliminal stimuli can be considered as implicit output modality. "Stimuli are rendered subliminal if they are attended to by the brain, but not consciously perceived"[64]. Previous studies [65] showed the fact that providing subliminal cue to participants in a target selection task increases their performance. Since implicit output is not explicitly observable by human agent, providing implicit output might not increase cognitive load of a mobile user in mobile interaction. Even though it is still an open question whether perceiving subliminal cues increases cognitive load or not.

2.2.2 Software architecture for wearable computers

Since one of the main research questions of this thesis is how to design a dynamic adaptive wearable system, reviewing the existing software platforms for wearable systems gives us an overview of the basic components of a wearable system. It also helps us choose an appropriate architecture to design the adaptive platform.

This section reviews the architecture of three software platforms developed for wearable computers: 1- Egocentric interaction middleware [71], 2- WUI-toolkit [73], 3- NETMAN software architecture [74]. There are many more studies on software platform for wearable computers with different names: middleware, software architecture, tool-kit, etc. [66, 67, 68, 69], but we chose these three because they belong to three different generations of the wearable computers. The main focus of NETMAN software platform, as one of the first architectures for wearable systems, is managing computing resources. The WUI-toolkit, as a more advanced wearable platform, has been used for developing context-aware and adaptive wearable UI, and finally, the Egocentric interaction middleware, as one of the recently developed platforms for wearable systems, incorporates the smart objects as part of the interaction session.

2.2.2.1 Egocentric interaction middleware

The Egocentric interaction middleware is a personal activity centric platform for wearable systems, aiming at supporting human agent interaction with smart objects and other applications (Figure 9). Smart objects are ordinary physical object with additional computing capabilities. The main components of the Egocentric interaction middleware are described in the following sections. [70, 71, 72]

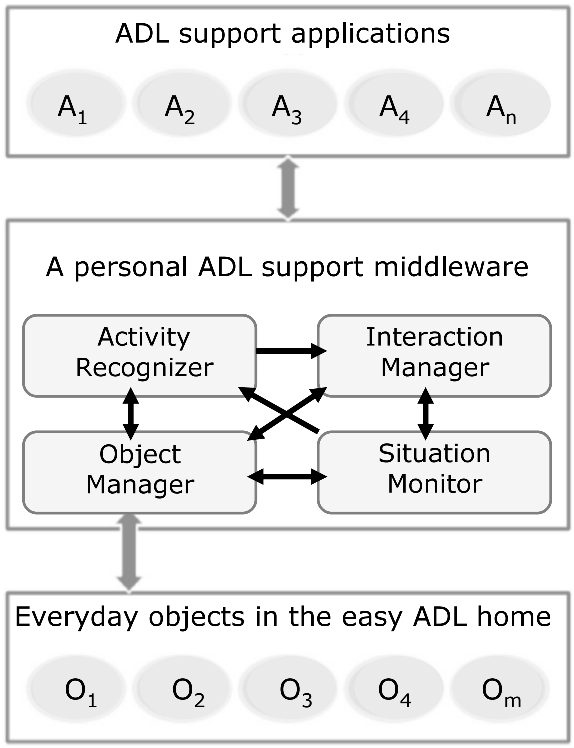

Figure 9. Architecture of the Egocentric interaction middleware [72]

Interaction manager

The interaction manager facilitates interaction with physical and virtual objects in different situations by making decision about where, when, and how virtual objects should be displayed so that they can be perceived by human agent. Another responsibility of the interaction manager component is facilitating multimodal interaction by keeping track of the current state of the mediators and their proximity to the human agent. The interaction manager is also responsible for facilitating interaction initiation by either human agent or applications.

Activity recognizer

The activity recognizer contains activity models and activity recognition algorithms. The data from the situation manager is used to recognize current activity of the human agent. In the current implementation [70] of the middleware, an offline supervised learning approach has been implemented to recognize activities. After recognizing activity, the activity recognizer exchanges the result with other applications to provide appropriate services or information to the human agent.

Object manager

The object manager component is responsible for managing smart objects. It detects smart objects and communicates with them via wireless connection. The object manager obtains more information about the smart object by connecting to the manufacturer database. After detecting the smart objects and mediators (input and output devices), the object manager keeps track of location of the physical objects.

Situation monitor

The main responsibility of the situation monitor component is keeping track of the location and state of the smart objects, mediators, and human agent. Situation monitor exchanges this information with other components to recognize activities and etc.

2.2.2.2 WUI-Toolkit

The WUI-Toolkit [73] is a software platform for adaptive wearable UI. Figure 10 shows the basic components of the architecture.

The UI of applications on the wearable computer is modeled by designers through the Abstract UI API. The abstract description of the UI can be converted into the Abstract User Interface Description (AUIDL) notation. The Presentation Component renders the UI model in the AUIDL format. In order to generate the UI, the Presentation Component should map abstract entities to the UI components. If there are multiple options during mapping process the Presentation Component decides about appropriate UI components.

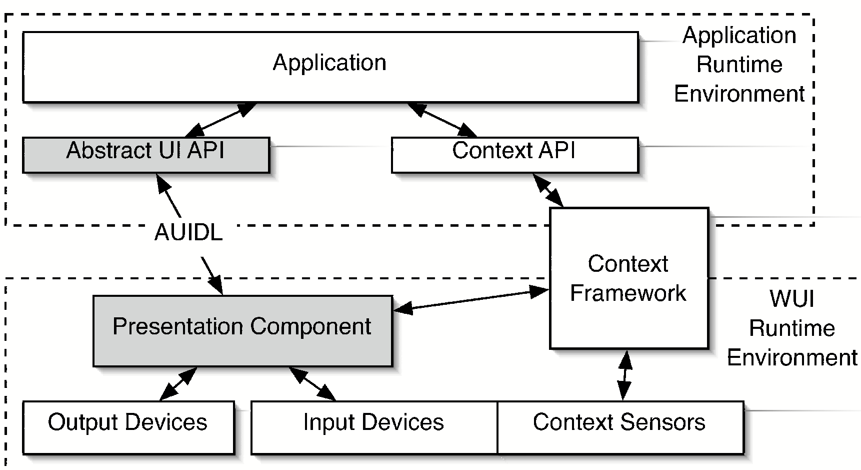

Figure 10. Architecture of the WUI-Toolkit [73]

Whilst the context changes, the Context Framework detects the change by interpreting the data from Context Sensors and informs the Presentation Component to update the UI according to the new state of the context.

2.2.2.3 NETMAN wearable computer software architecture

One of the earliest attempts to develop a software platform for wearable computers was NETMAN wearable computer software architecture [74]. Figure 11. NETMAN wearable computer software architecture [74]Figure 11 shows the schematic view of the NETMAN software platform.

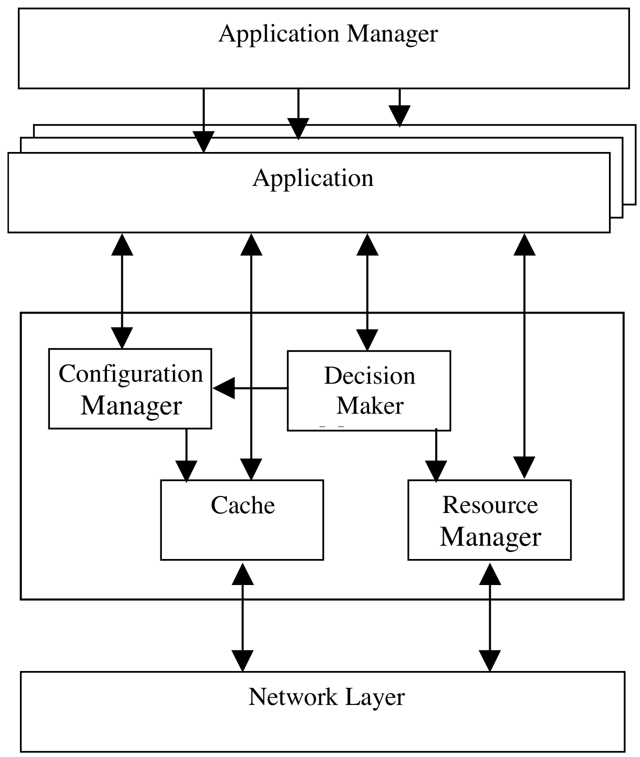

Figure 11. NETMAN wearable computer software architecture [74]

The highest level of the architecture is Application Manager that provides a graphical UI to the user to interact with applications. The next layer of the architecture represents the applications that can be loaded or unloaded dynamically. The third layer is actual middleware and consists of four components. The Configuration Manager component monitors hardware and software installed on wearable computer. The Resource Manager keeps track of state of resources such as network bandwidth, disk space, and etc. The decision maker component uses a probabilistic approach to manage resources by taking decisions like whether a task should be run locally or remotely on a server. The Cache is a local data storage that can be used by other applications, and finally the network layer provides a network interface for applications.

2.2.3 Multimodality in wearable/mobile systems

As we mentioned in section 2.1.3, in mobile interaction the context of use changes frequently. For example, in a noisy environment the speech input modality might not be a perfect option, or when a user is crossing the street probably the HMD is not an appropriate output device. This means that multimodality can be a valuable property for wearable and mobile systems to support user in different situations by offering several interaction channels to the user.

According to the definition of multimodal systems in section 2.1.4, most of the wearable and mobile devices are multimodal because they are able to communicate with users through several modalities. However, a few of them provide a combination of different modalities for a graceful interaction. For example, in [75] a combination of pen and speech has been developed for interaction with PDAs. In another study [76], head gesture and audio feedback has been used to control mobile devices. However, to the best of our knowledge there is no study in the literature on automatically switching between different modalities for wearable and mobile devices.

Multimodal human-computer interaction is a wide and multi-disciplinary subject and crosses different research areas such as computer vision, psychology, artificial intelligence, and many others. Since the main focus of this thesis is to design a platform for UI adaptation in different situations, we have reviewed the concept of adaptive user interfaces in the following section.

2.2.3.1 Adaptive user interfaces

"An adaptive user interface is an interactive software system that improves its ability to interact with a user based on partial experience with that user" [77]. The concept of adaptive user interfaces has been mainly developed for graphical user interfaces; however, some studies extend the adaptation to the multimodal interfaces [78].

The main goal of the adaptation of a user interface is improving usability of the system via *re-modeling* or *re-distribution* of the user interface [16]:

- UI re-modeling means any kind of re-organization of the UI elements that is perceivable to the user. Re-modeling could be insertion of new UI components (i.e. displaying more information when there is more space on the screen) or deletion of some existing UI components (i.e. removing unnecessary UI components from a PDA screen).
- UI re-distribution denotes the reallocation of the UI components to different interaction resources in a ubiquitous computing environment in which multiple interconnected devices or smart objects could be used as interaction resources.

Adaptations may be triggered by different kinds of causes, which can be classified into three main categories:

- *User related context*: any kind of voluntary or involuntary actions of users can be a reason for adaptation. For instance, when a human agent is walking, part of his/her visual perception is occupied so that if system wants to initiate an interaction, HMD might not be the best output device.
- *Environment of interaction*: changing social or physical aspects of the environment could trigger the adaptation. For example, the noise or light condition of the environment can raise some difficulties for interaction through speech or HMDs.
- *Computing platform*: availability of the computing devices or smart objects in a situation can also cause the adaptation. For instance, when a larger display is available, the adaptive system can upgrade the UI by displaying more information elements.

From an interaction point of view, adaptation can be triggered either manually by user (*adaptable*), or automatically (*adaptive*). The adaptable systems are also called personalizable or customizable systems. According to this definition, most of the PC software is adaptable since they let users to change the UI of system. Bezold and Minker [78] proposed a general structure for adaptive interactive systems, which is illustrated in Figure 12. In this model adaptation is an iterative process including two main stages: user modeling and performing adaptations. In the user modeling phase, the system monitors user behavior through collecting data from sensors and interpreting them to update the user model in run-time. The outcome of the user modeling could be i.e prediction of the user's next action or estimating user's proficiency. The next step is preparing and executing adaptations based on user model. The adaptation process relies on four different models, which are described in the following sections.

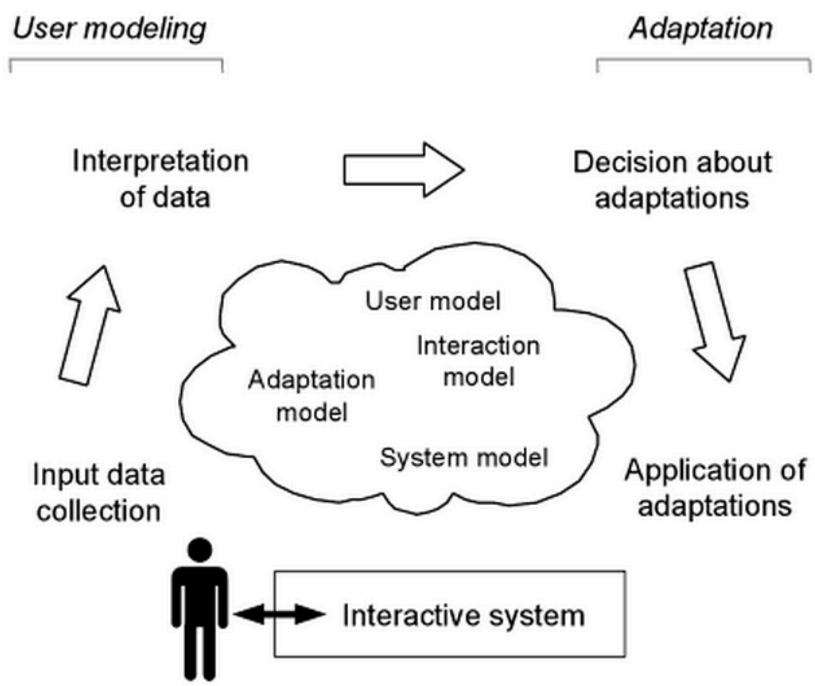

Figure 12. A model of adaptive interactive systems [78, p8]

The system model

The system model represents different aspects of an interactive system. The system model is mostly static and determines the changes generated by adaptation. The description of UI components and elements are part of this model. In a graphical UI, UI components are text labels, buttons, lists, images, and etc., which can be represented in a hierarchical model. While for voice interfaces the UI elements could be recorded or synthesized speech, language models, and speech recognition algorithms.

The user model

User model represents an abstraction of user characteristics from adaptation point of view. The user-modeling component of an adaptive system preforms learning, inference, and decision-making. The result of this computation updates the user model. The user model includes user preferences, history of interaction, and prediction of next steps. Some of the user characteristics also could be a part of user model. For example, level of user proficiency could be detected by the system to adapt the UI accordingly. Usually the user model comprises a wide range of data types from simple flags to complex algorithms. Inputs to the user model can be explicit through forms and questionnaires or implicit from i.e. physiological sensors or system internal events. However, some of the user characteristics such as level of proficiency can be changed over time and should be updated frequently.

The interaction model

The interaction model defines the actions of a user and possible relationships between these actions. A user action is an atomic unit of the interaction. The user action can be described by a single observation or a sequence of several observations in the system. For example in a multimodal messaging system, "composing a new message" can be defined as a *user action* which can be observed by pushing the "compose button" in the UI, or receiving a speech command (e.g. "write a message"). The interaction model includes all available user actions correspond to the system functionalities and the ways of observing actions by the system. In fact an interactive system needs to recognize user actions through different modalities in run-time.

The adaptation model

The adaptation model describes a set of adaptations and the context in which the adaptation should happen. The adaptation model includes two main parts:

- A *declarative* description of prerequisites and effects of the adaptation, which connects the adaptation to the user model.
- A *functional* part defines the changes that should be applied to different parts of the UI.

2.3 Summary

In this chapter, we reviewed relevant concepts and technologies to mobile interaction in three different research areas: wearable computing, context-aware computing, and multimodal interaction.

In wearable computing literature the need for *adaptive multimodal user interface* is discussed in several studies (see section 2.1.2); however, we have not found any practical solution for automatically adapting UI in wearable computers.

In context-aware computing research area, the need for *updating system behavior* by changing context of use has been discussed in several studies, and three different modes have been defined for updating system behavior: 1- *Personalization*: manual updating, 2- *Passive context-awareness*: user should define the preferred changes in a particular contexts and system changes the behavior when detects that context 3- *Active context-awareness*: system autonomously changes the behavior (more details in section 2.1.3).

In multimodal interaction field, the concepts of adaptive UI (automatically adaptation of UI) and adaptable UI (manually customizable UI) have been defined as two different modes of UI adaptation. A model-driven approach has been used in several studies to design and develop adaptive multimodal interactive systems.

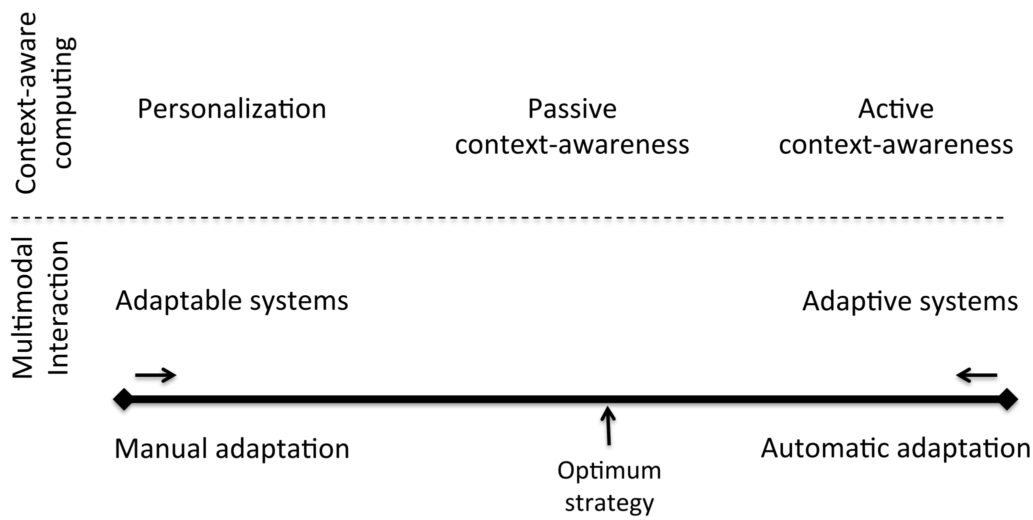

Figure 13. The system adaptation continuum

To sum up, we can conclude that there is a continuum between completely automatic adaptations of the UI to UIs that are set up completely manually by users themselves. Given the fact that there is always an uncertainty in results of the context recognition, the optimum adaption strategy could be defined somewhere between the two extremes (Figure 13).

3 DESIGN

The main focus of this chapter is design a software platform for dynamic UI adaptation of wearable systems. In the first step, we analyzed the design space based on the concepts and topics identified by the work presented in the literature review. In the next part, the architecture of the Egocentric interaction middleware is extended to support dynamic UI adaptation.

3.1 Design space analysis

Figure 14 illustrates the design space of adaptive multimodal wearable system to support mobile interaction. As shown in the Figure 14, the design space is characterized by following dimensions:

- *Interaction initiator* (section 2.2.1): each side of the interaction user or system can initiate interaction between system and user.
- *Context of use* [16, 78]: includes the main causes of adaption which could be user behavior or characteristics, physical environment, or state of computing platform.
- *Adaptation strategy* (section 2.3): the decision of UI adaption could be taken by user (manual), system (automatic), or a combination of user and system.
- *UI Adaptation means* [16]: the means used for adaption could be remodeling the current UI, redistributing UI elements between different devices, or reconfiguration of the whole UI.
- *Modality combination* [16]: modalities could be combined in different ways to adapt UI: replacing one modality with another one, adding a redundant modality to the UI, or using a new modality as complementary modality in adapted UI.

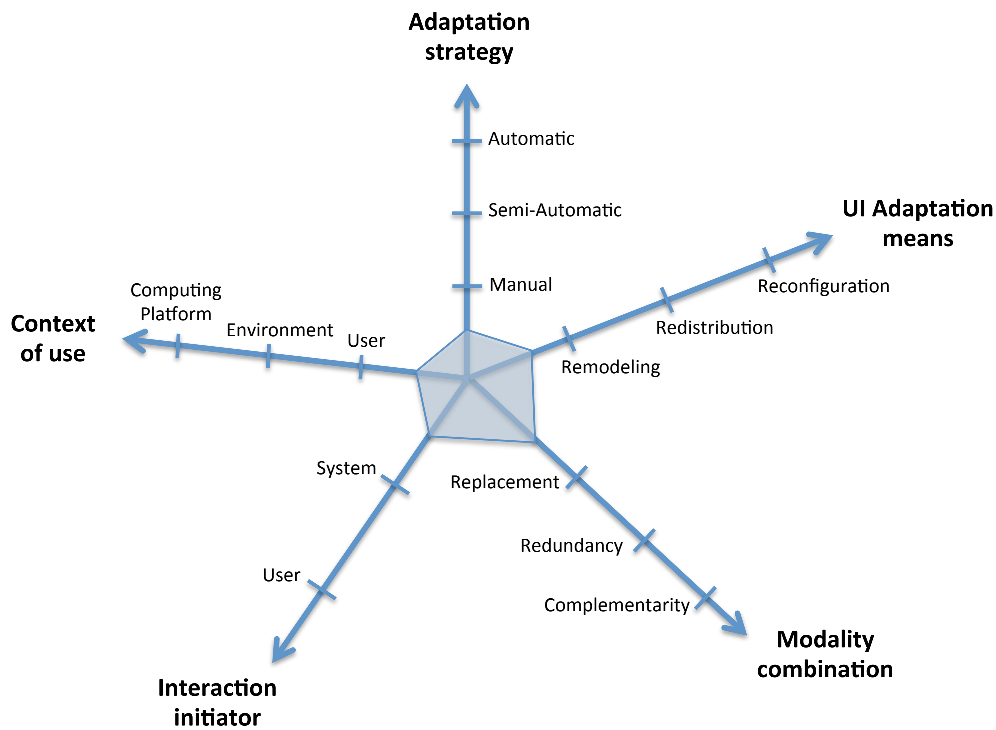

Figure 14. Design space of adaptive multimodal wearable interface

In the following sections each aspect of the design space is elaborated.

3.1.1 Interaction initiator

3.1.1.1 System

According to the definition of wearable computers, a wearable system should be able to predict and prepare the information might be needed by users (section 2.2.1). In such a system, some of the interaction sessions should be started proactively by the system. Usually, the initiator of an interaction chooses the modality and time of initiating an interaction session; this means that a wearable system should be able to decide about the modality and the exact time of interaction. In this kind of interaction sessions, system uses the context as implicit inputs, which means users do not need to input data explicitly. In mobile interaction scenarios, providing explicit input to the system is challenging for users, and if the system can make decisions on the basis of implicitly gathered input, it is valuable. Also the time of interaction is an important factor for mobile users. For example, when user is crossing a conjunction the wearable system should postpone the interaction session to a proper situation.

3.1.1.2 User

Even if a predictive wearable system can proactively provide information for human agent, in many situations human agent initiates the interaction session. Apart from these kinds of interaction sessions, in many situations system is not able to predict user's needed information due to the lack of context data or inaccuracy of the context recognition algorithms. In these cases, the human agent initiates the interaction session. As mentioned in the previous section, usually the initiator of an interaction session chooses the interaction modality; however, if during the interaction, context of use changes the wearable system can adapt the UI accordingly. In such situations, system should be able to recognize the action of the user and predict the next steps based on predefined action models in the system. Also the activity recognition and context recognition module of the wearable system can learn from the manual interaction to improve their prediction.

3.1.2 Context of use

3.1.2.1 User

User context is one of the most important causes of the UI adaptation. For example, when a user switches the activity, the perceptual requirements of the new activity might interfere with the current modalities. As an illustration, imagine a user is driving and at the same time is providing input to a system through speech modality. If the user starts talking to his/her phone the speech input cannot be a proper modality anymore. User context includes user characteristics, user behavior, and user emotions. User characteristics are more static than behavior and emotions. Some of the characteristics that affect the UI adaptation are personal preferences, visual or hearing impairments, age, and physical or mental disabilities [78]. Some of these characteristics can be defined explicitly for a wearable system, and some of them such as user proficiency can be extracted automatically in run-time. Recognizing user behavior is usually more challenging and needs sophisticated algorithms. The implicit input from wearable sensors and internal events should be collected and interpreted with the help of machine learning techniques to change the UI or modality according to the activity and state of the user. For instance, when a mobile user is walking, the HMD might not be a proper output device for long period of time because of the fact that walking needs human visual perception and interaction through HMD can interfere with walking activity. The emotions of the user can also be important to adapt the UI. For example, when a user is angry, the wearable system can postpone some of the cognitive demanding tasks. The system can also learn what kinds of tasks affect users' emotional state. This knowledge can be used to schedule users' tasks. The emotional state of the user can be interpreted based on user's voice or facial expressions.

3.1.2.2 Computing platform

In desktop applications, information about computing devices is probably the most available context data. But in a mobile scenario in which the availability of different computing resources changes frequently, it is not always easy to have an updated profile of all available computing devices. According to the egocentric interaction paradigm, computing devices (mediators) could be available or unavailable in different situations. Also the smart objects can be resources (e.g. tangible UI elements) for interaction. Detecting mediators and smart objects and integrating them into the whole interaction setting could be challenging without standard protocols to define properties and behavior of each mediator or smart object. Beside the mediators and smart objects, other computing resources such as network connection status or battery life could be important to adapt the UI or select a modality. Similar to the user context, some parts of the computing platform context are more static and can be defined explicitly for system (e.g. profile of standard mediators such as HMD) while some other data (e.g. new mediators and smart objects profiles) can be dynamic. When the wearable system adapts UI, the profile of the mediators can be used to choose appropriate mediator and exploit the device resources optimally. For example, if large amount of data needs to be displayed a mediator with bigger screen should be selected as output device.

3.1.2.3 Environment

Environment context includes a wide range of data from location to the physical state of the environment (temperature/noise/light condition, etc.). In the egocentric interaction paradigm, proximity of the objects to the user is an important factor to determine a situation. The proximity of the smart objects can also be used to recognize the current activity of user. For choosing a proper output device, proximity of the mediators to the human agent shows whether the displayed data on the output device can be perceived by human agent or not. Other important aspect of the environment is social context. As mentioned in section 2.1.1.3, presence of other people can affect mobile interaction due to the physical constraints, privacy or security issues. The social context should be detected by the wearable system to adapt the UI accordingly. For example, when user is talking to someone, system can detect the face of other person and postpone system interactions to a proper time or switch to the silent mode (e.g. using haptic modality for notification). Another potential use of the social context can be adapting the interaction modality based on preferences of the majority of people located in the same location. This needs a proximity-based crowdsourcing platform to share and use the interaction preferences among a group of users. For example, when a mobile human agent is listening to a lecture in a classroom, most of the devices would be on silent mode. This information can be used by a wearable system to adapt the UI accordingly and avoid embarrassment.

3.1.3 Adaptation strategy

3.1.3.1 Manual

As mentioned in section 2.3, the UI adaptation can be completely done by user. Today in most of the mobile devices user should manually change the modality, and just a few devices provide location-aware notification services. Due to the uncertainty of the context recognition models, it should be always possible to change the UI manually by human agent. The wearable system can record the history of manual adaptation to improve next predictions.

3.1.3.2 Automatic

In the other side of the continuum is fully automatic adaptation by the system. The main goal of a wearable personal assistant system is adapting UI automatically without decreasing usability of the UI. An important aspect of the automatic adaptation is informing the human agent of changing the UI. For example when the adaptive system replaces one modality with other one, it can be confusing for user to switch to the new UI. Especially when several output and input devices are available in the setting, finding the active one can be challenging.

3.1.3.3 Semi-automatic

The adaptation strategy could be adjusted based on the certainty of the system about context recognition and user preferences. The optimum strategy to change the UI can be somewhere between two extremes. For example, for the small adaptation of the UI (e.g. changing the size of a button in UI) system can take the decision autonomously but for replacing a modality, it might be better to keep the human in the loop of decision taking.

3.1.4 UI adaptation means

3.1.4.1 Remodeling

As mentioned before, UI re-modeling is any kind of re-organization of the UI elements that is perceivable to the user including insertion of new UI components or deletion of some existing UI components. In a wearable adaptive system, remodeling concept can be defined as UI adaptation on one device. For example, when a mobile user is interacting with wearable system in a very bright place, the color and size of the content on HMD can be adjusted to the light of environment.

3.1.4.2 Redistribution

UI redistribution denotes the reallocation of the UI components to different interaction resources. In an adaptive wearable UI, the redistribution of UI can be defined as reallocation of the UI components to different predefined devices including currently body-worn devices and also devices embedded in the environment. The profile of the predefined devices exists in the system; therefore, the UI model can be regenerated based on static profiles and adaptation patterns.

3.1.4.3 Reconfiguration

The reconfiguration concept can be specially defined for egocentric adaptive UIs. In the egocentric interaction paradigm, mediators and smart objects can be added to the interaction environment or removed from the interaction. As mentioned before, smart objects can be spontaneously used as tangible elements of the UI. Also new output devices in the perception space of a mobile human agent can display information as a mediator. In such scenarios, first of all the profile of the new smart objects and mediators need to be known or discoverable by the wearable system. This profile contains the properties and functionalities of the new device. In order to use new devices in an interaction setting we need to define an interaction model. The interaction model describes actions a user performs and possible relations between actions [78]. The actions are logical steps to represent the atomic possible interactions (input and output). In UI reconfiguration, the interaction model can be defined or updated by human agent in runtime. For instance, a human agent can assign the role of slider component in graphical UI to a smart object. In reconfiguration scenarios, human agent needs to define a new interaction model for the system. In the next step, the system regenerates the UI based on the new interaction model. The result of these kinds of adaptation might be unexpected from the usability point of view.

3.1.5 Modality combination

3.1.5.1 Replacement

Sometimes in a mobile interaction scenario, one modality is better replaced with another modality. For example, when a human agent is using audio-based modalities in a noisy environment, the performance of the speech recognition could be affected by ambient noise. In this case, the system should replace the input or output modality with other suitable modalities.

3.1.5.2 Redundancy

In particular situations, due to the importance of the information or other reasons the wearable system might need to display the same information through more than one modality. For example, an acoustic alarm can be added to a visual alarm to improve security.

3.1.5.3 Complementarity

Combining different modalities to use the complementary strength of each modality can improve usability of the UI. For example in "put that there" system [15] a combination of gesture and audio input modalities provided a more graceful interaction. The complementary use of modalities can provide hands-free [37] or eyes-free [81] interaction techniques to a mobile user. In an adaptive wearable UI, modalities can also be combined to increase the accuracy of inputs when the system detects a high rate of user mistakes in runtime.

3.2 Adaptive egocentric interaction manager

Based on the design space analysis, there are several aspects in design and development of adaptive multimodal UI for wearable systems. On the one hand, a mobile human agent needs to use several applications such as calendar, navigation system, and messaging system. On the other hand, most of the context data can be captured and interpreted independent from applications. In fact the context data can be shared among different applications. Therefore, the most efficient approach is designing a middleware for wearable system to collect the context data from sensors and interpret it to a higher level of abstraction for other applications. Also most of the adaptation mechanisms are user-dependent not application-dependent. Thus the adaptation module can also be a part of middleware. However, some of the limitations for UI adaptation rooted into the applications. For instance, the audio might not be an ideal modality for drawing a sketch. These kinds of limitations can be defined in the system profile of each application to be used during UI adaptation.

From the literature review, we found the architecture of the Egocentric interaction middleware as the best start point for designing an adaptive multimodal middleware for wearable systems. Because of the fact that the current version of the Egocentric interaction middleware supports several aspects of an adaptive multimodal UI such as multimodality, activity recognition, body-centric, and situatedness.

In this section, we extended the architecture of the Egocentric interaction middleware (shown in Figure 9) to support adaptive multimodality. In the Egocentric interaction middleware, the adaptation functionality is assigned to the interaction manager

component. Therefore, we propose a conceptual design for interaction manager component to support multimodal adaptation.
Figure 15 illustrates a schematic view of the adaptive Egocentric interaction middleware. In the following sections, each component of the interaction manage module and relationships between components are described.

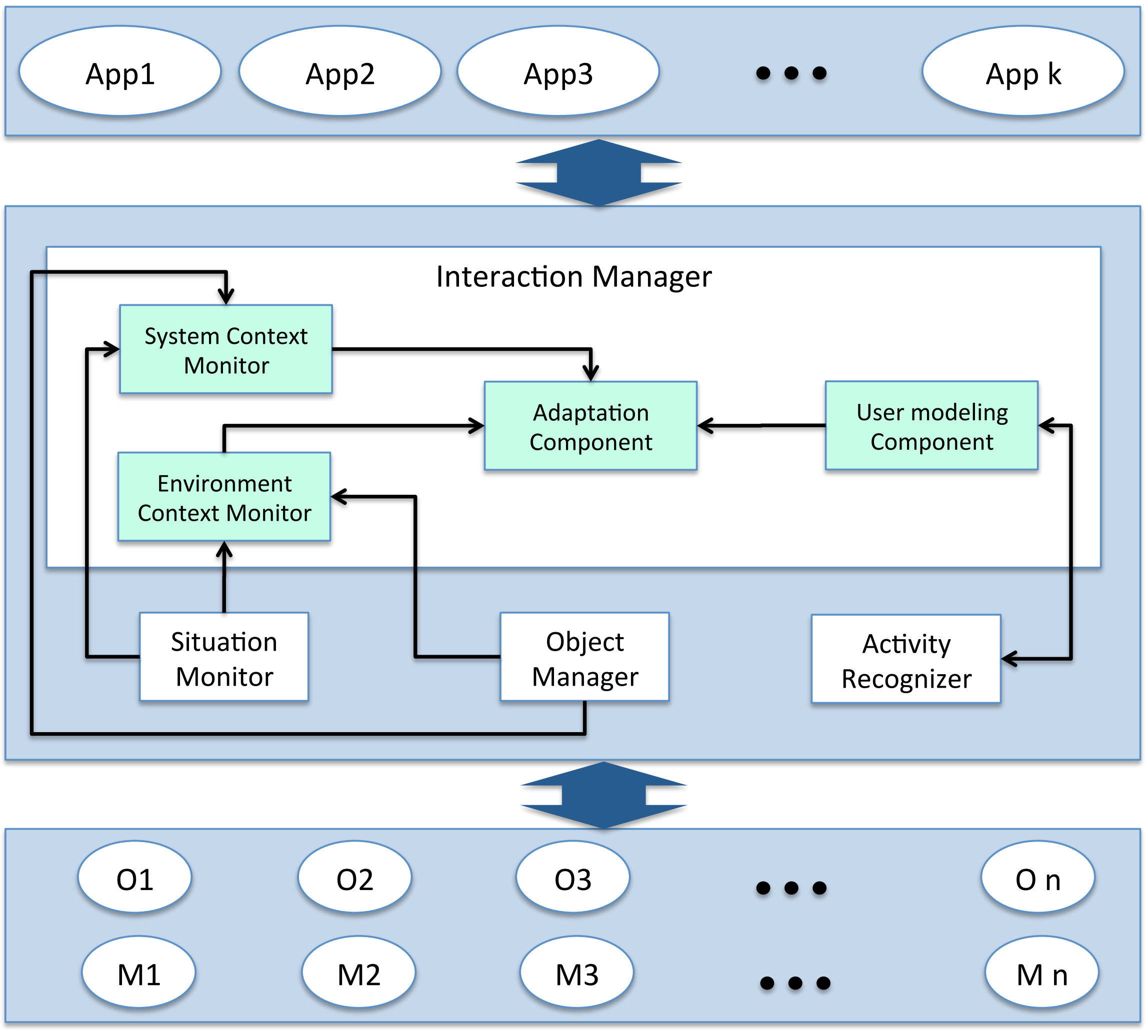

Figure 15. The architecture of adaptive Egocentric interaction middleware

3.2.1 User modeling component

User modeling component forms the backbone of user modeling process. In the user modeling process, the collected data from sensors and other system observations is processed to infer user behavior.
The user model is one the most important resources of user modeling component (see Figure 15). The *user model* describes user behavior in a sequence of events or actions

[78]. The user-modeling component analyzes the sequence of events to recognize user actions.

Another important part of the user-modeling component is *task model. "*A task defines the activity required to accomplish a specific goal, and an activity is defined by means of a hierarchical arrangement of user actions" [78]. By actions we mean explicit input the human agent provides to system. These actions can be related sequentially or alternatively in different tasks. The task model can be updated in runtime based on a new interaction model (described in section 3.1.4.3). The task model is used in runtime to predict most likely next user action to adapt UI accordingly. Different machine learning approaches (supervised or unsupervised) and models (Markov, Neural Network) can be used to predict user actions [78].

User characteristics such as age, visual or hearing impairments, and preferences are the static parts of the user-modeling component. However, some of the static information such as preferences and proficiency can change over the time (concept drift phenomena), and system need to monitor and learn these characteristics of human agent. Some parts of the user model can be entered explicitly, while some other data can be captured implicitly in runtime by system.

The activity recognizer component of the Egocentric interaction middleware can get benefit of the user model as an input to improve activity recognition. In fact, the user model represents the user actions in the system, which can be used by activity recognizer component. For example when a user is searching for a recipe in the system, the most likely next activity can be cooking. Also the output of the activity recognizer can be used by user modeling component to predict user action or adapt the UI according to the high-level goals and activities of human agent.

3.2.2 Environment context monitor

The environment context monitor includes the context model and context recognition algorithms to interpret the collected data from different sensors to a higher level of abstraction. For example, the signal from a wearable microphone can be used to analyze ambient sound and recognize the place of human agent.

The environment context monitor acts as a context service provider to other applications (App1 to Appk in Figure 15). All of the important context data such as social context, proximity of objects and mediators, and physical conditions of the environment (e.g. noise level, temperature, light, etc.) are defined and modeled as context entities. Whenever the value of a context element changes the entity listener service informs the adaptation component to adapt the UI accordingly.

Since the sensed data is inherently imprecise and uncertain the output of the context monitor is also uncertain; therefore, the quality of the interpretation is also estimated and

sent to the adaptation module. The certainty parameter of the context data can be used to take decision about adaptation. For instance, if the level of certainty is not significant about the current activity of the user, system can follow a semi-automatic or manual strategy to adapt the UI.
The environment context monitor component receives information about the proximity of objects and mediators and state of the situation from situation monitor and object manager.

3.2.3 System context monitor

The system context monitor contains the profile of all applications (UI models), computing devices (mediators and smart objects), and other resources such as network and battery. The UI models are the most important information for UI adaptation. The UI model includes different UI components in different modalities. For instance, in a graphical UI, text boxes, lists, and buttons should be defined as elements of the interface. The UI model describes also the sequence of events in UI. For example, pressing a button by user triggers other events in the system. The adaptation component uses this model to render the new adapted UI in runtime.
In adaptive multimodal systems with a fixed set of computing devices, the system model is static [78], but in mobile interaction with egocentric approach new smart objects, or mediators can be added or removed from interaction session. This means that the profile of the computing devices and associated UIs are dynamically updated. The state of the smart objects and mediators is updated by situation monitor and object manager components.

3.2.4 Adaptation component

The main goal of adaptation UI is improve the usability of the system. Therefore, to adapt the UI, usability aspects should be considered. At the first glance, the adaptation of UI seems to violate usability principles such as predictability [79]. But the adaptations can be defined in a way that they comply with usability guidelines. Defining adaptation patterns according to the usability principles can be a solution to this problem. According to [78], the main usability fundamentals in designing adaptive systems are as follows:

- *Predictability*: if user can anticipate the reaction of the system to their input, the learnability and user speed would be improved. So that the adaptation patterns should design in a way that help user to anticipate the system reaction. Also, if during the interaction human agent chooses a particular modality, the system should immediately adapt to the selected modality.

- *Transparency*: helps users understand the inner workings and current interface of the system. When UI adapts automatically, it is crucial to represent explicitly the logic of new interface to the human agent. Also it is important that users understand the logic behind the decision to adapt the UI or switch the modality.
- *Controllability*: necessitates that the human agent should be able to control the interaction. The adaptation patterns should be designed in such a way so as to let human agent change the adaptation decisions in all steps of adaptation.
- *Unobtrusiveness*: the adaptation should not distract human agents during interaction; however, at the same time the adaptation should be noticeable to the human agent.

The adaptation patterns are the most important information stored in the adaptation component. These patterns are the general rules of adaptation and should be designed according to the requirements of human agent in mobile interaction. These rules can be formulated in different procedural or declarative notations. The adaptation patterns also include the triggers of adaptation. By changing the context and triggering the adaptation pattern, the current UI changes according to the adaptation rules in runtime.

4 IMPLEMENTATION

4.1 Scope of implementation

According to the previous studies on adaptive multimodal interactive systems [78], by using a model-driven approach implementation of the proposed architecture is technically feasible; however, the implementation of the complete adaptive Egocentric interaction middleware (see Figure 15) is out of the scope of this master thesis. The implementation part of this thesis focuses on building the software and hardware platform needed for evaluating behavior of the adaptation component, which is part of the egocentric interaction middleware. More specifically, we decided to implement an experimental apparatus for investigating the influence of human agent's performance depending on where in the adaptivity continuum (page 27) the system engaged in interaction with the human agent.

As mentioned in the previous chapter, several human factors (to some degree covered by usability principles) should be considered in design and development of adaptive systems. Predictability, transparency, controllability, and unobtrusiveness are some of the most important addressed principles in the literature. The most challenging aspect in all of these principles is automatic adaptation. By contrast, in the manual adaptation the human agent decides about the adaptation; therefore, the result of the adaptation would be predictable and transparent to the human agent. In addition, keeping human in the loop implies that human agent is controlling the system. It is only with respect to the unobtrusiveness principle, it is an open question whether manual adaptation is more distractive than automatic adaptation or not.

For mobile interaction, unobtrusiveness is a crucial aspect of user interface. Since the mobility increases cognitive and motoric load of humans, and manual adaptation demands extra actions for the human agent to decide about the adaptation. This extra action can affect the performance of the human agent in physical-world task.

To investigate the effects of automatic or manual adaptation on the performance of human agents, an experiment was conducted. To conduct the experiment, we needed to design a mobile interaction scenario in which the human agent has to perform a physical task and at the same time interact with a computing device. Usually the real physical tasks are hard to control and measure in the laboratory. That is the reason why in HCI community for evaluation of the mobile and wearable systems, usually we use a simulator of real tasks. For example, some studies [80, 81] have used the car-driving simulator as an abstraction of physical tasks because it is easy to measure the performance of user in such tasks. In this thesis, we implemented Hotwire task, which has been used in previous studies for evaluation of wearable user interfaces [82]. We also developed a multimodal wearable system for presenting information about virtual tasks to the participants. Head gesture and speech input are used as input modalities, while displaying on the HMD and synthesizing sound were chosen as output modalities. Moreover, a monitoring system has been developed to store all of the quantitative data and measure performance of the participants. In the following sections we describe the design and implementation of each system briefly.

4.2 Hotwire apparatus

According to [82], a simulated task for wearable system should be a physical task and generates some motoric load on the human agent. The simulated task needs to be easy to learn to minimize the human agent's error during the experiment, and finally it should be adaptable to different kinds of physical task. The Hotwire apparatus fulfills all of these characteristics. The main concept of Hotwire adopted from a children's game. The toy version of Hotwire is used to train motoric abilities of children. The game setup consists of a bent metallic wire, which is fixed on a piece of wood, and a hand-held tool with a metallic ring (Figure 16). The goal of the game is passing the ring from one end of the bent wire to the other side without touching the wire.

Figure 16. The commercial version of Hotwire toy

The manual characteristics of the Hotwire task is explained in [73]. Visual attention, physical demands, and cognitive demands of the Hotwire task can be adjusted by changing either diameter, length, and shape of the bent wire or diameter and weight of the hand-held part.

In this study, we made the Hotwire platform from a piece of copper pipe with the length of 2m and 1cm diameter. The copper pipe is bent and mounted on a wooden plate with the size of 10*55*3cm (Figure 17). The hand-held part of the device comprises of a wooden stick and a metallic ring with the size of 3cm. The total weight of the hand-held is about 15 gr.

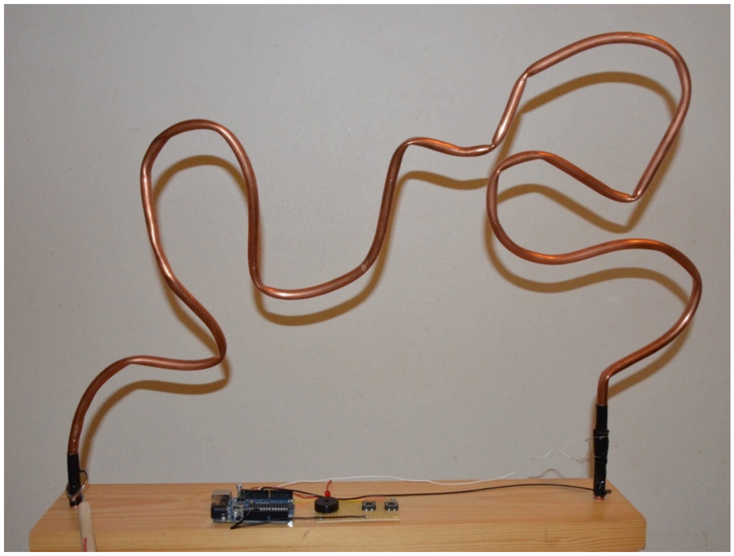

Figure 17. The prototype of the Hotwire apparatus

4.2.1 Technical Setup of the Apparatus

The technical setup of the Hotwire apparatus consists of the Hotwire platform and a microcontroller (Arduino Uno) connected to a computer via serial connection to transfer the quantitative data of the user performance. The user performance in Hotwire task is defined based on number of errors (touching the wire) and task completion time. In order to measure the task completion time the Hotwire apparatus captures the start and finish events. By pressing the start button, microcontroller sends the start signal to the monitoring system. When user reaches to the end of the wire, by touching the end point, the microcontroller sends the finish signal to the monitoring system. During the task whenever user touches the wire, an error signal is sent to the monitoring system. A schematic view of the technical setup of the Hotwire apparatus is illustrated in Figure 18.

4.2.2 How to perform Hotwire task

To start the Hotwire task, human agent should press the start button manually, and to finish the task the end point should be touched by the hand-held stick. The user should try to pass the wire from start point to the end point without touching the wire within the shortest possible time. The performance of the user is calculated automatically based on the total number of errors (touches) and task completion time.

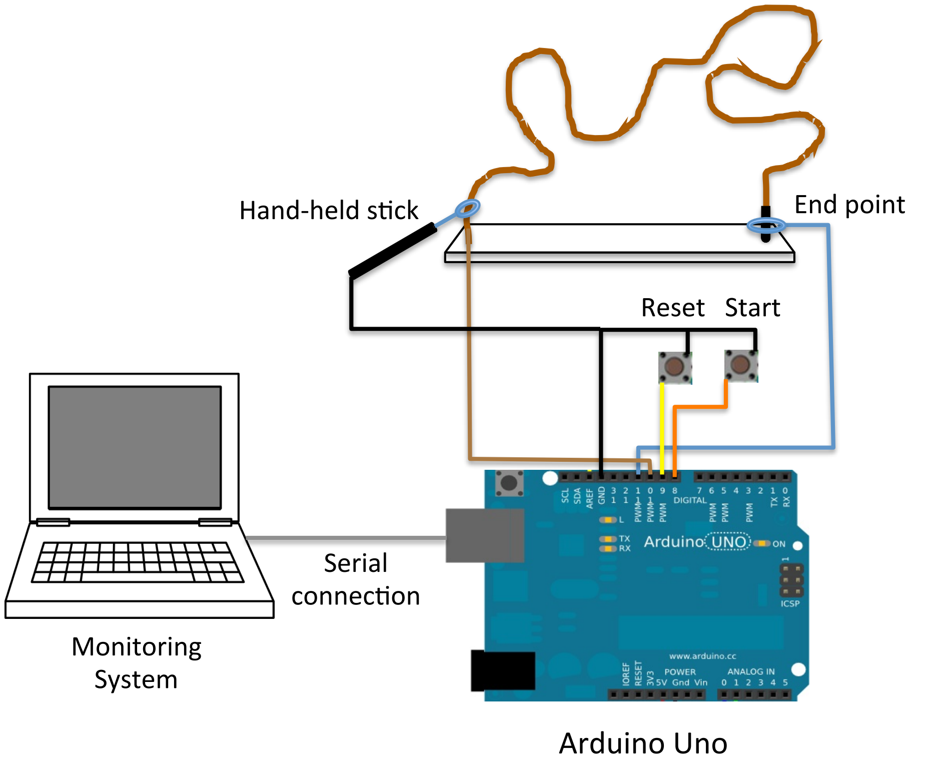

Figure 18. Schematic view of the Hotwire apparatus

4.3 Virtual task generator

As mentioned in section 4.1, for the purpose of our experiment, human agent should interact with the wearable system in parallel with the physical task. To generate the virtual task, first we developed a match selection task [80], but our first evaluation showed that we could not generate the same kind of interaction in both visual and aural modalities. Therefore, we changed the task to a simple calculation task. In the calculation task, the human agent is asked to answer a simple mathematical question. The task is defined as calculating two basic operations (+ or -) between two numbers in the range of 1 to 9. The calculation task can be displayed and answered visually on the HMD or vocally via headset. In the visual modality the task is displayed on the HMD in a multiple choices format with two options. The human agent can choose the correct answer between left and right by performing the head gesture to the left or right side. In the audio-based modality, a synthesized speech asks the question, and the human agent is supposed to answer the question vocally. During the audio-based interaction, the HMD shows a black background to avoid distraction.

4.4 Wearable system

The wearable system consists of the following hardware components:

- A monocular non-see-through HMD (MicroOptical SV-9, 640x480 pixels) mounted on a safety goggles
- An ordinary headset, which is augmented by a wireless head motion sensor
- And an ordinary Macbook Pro 15” with 4GB RAM

The software platform of the wearable system comprises the virtual task generator, the speech recognition, and the gesture recognition modules. The software platform is developed in the Processing[1] environment due to the simplicity and fast prototyping capabilities of the Processing environment. For speech synthesize, we used the built-in text to speech function of the Mac OS by sending the say command to the terminal from the Processing program.

4.4.1 Speech recognition

There are two approaches to develop speech recognition applications: offline and online (or server-based). For developing an offline speech recognition system we need to install the speech recognition engine locally. For mobile and wearable system with limited processing capabilities the online approach is preferred since the performance of the online methods has been evaluated as good as offline methods [83].
To develop the speech recognition module of the wearable prototype, the Google speech recognition service has been used. To call the speech recognition service we used stt library in the Processing code. To capture the voice of human agent, after playing the vocal question, system starts recording the sound until receiving a complete sentence or reaching the timeout limit, which was set to 7 seconds from beginning of the question.

4.4.2 Head gesture recognition

Another input modality developed for wearable system was body gestures. We had three different options for gesture input modalities: hand gestures, foot gestures and head gestures. Since performing foot gestures decreases the balance of the user, and in Hotwire task user needs to keep the balance, foot gestures approach was rejected. In the first iteration of the implementation, we tried both hand and head gestures to provide input to the system while performing Hotwire task. The first evaluation showed the fact that performing hand gestures while doing the Hotwire task decreases physical balance of the

[1] www.processing .org

user and increases the number of errors in Hotwire task. Therefore we used the head gestures as input modality for the wearable system.

4.4.2.1 Modeling gestures

According to the experimental design (described in the evaluation chapter), for dispatching input commands to the system a set of four well-distinguishable gestures have been defined: Up, Down, Right, and Left. All head gestures are illustrated in (Figure 19). Our head gesture recognition approach is adopted from hand gesture recognition method in [37].

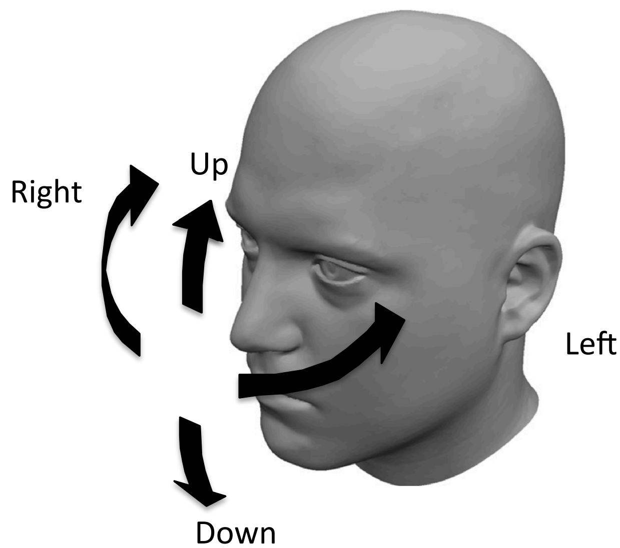

Figure 19. Head gestures

To detect the head gestures, a head motion sensor module is developed and placed on the headset (Figure 20). The sensor module comprises a microcontroller (Arduino Pro Mini 3v), a Bluetooth communication module, and a 6 degree of freedom inertial sensor with acceleration and rotation sensing capabilities (ITG3200/ADXL345 from Sparkfun).

Figure 20. The head motion sensor module fixed on the headset

4.4.2.2 Learning gestures

We had two different options for training our system to detect head gestures: a person independent learner or training system for each user. The second approach would increase the accuracy of the system at the expense of recalibrating system for each user, which is time consuming. We followed the person independent strategy because our hand gestures are simple and easy to learn for user. For each gesture a training set of 20 gestures is recorded and labeled. This training data is used as the input for a learner component, which is trained for each gesture. Then the trained learner component is used for the online gesture recognition.

4.4.2.3 Gesture recognition

In order to recognize the four gestures, we followed a two-stage strategy with a set of single-output artificial neural networks. In the first phase, we discriminate between the idle state and any other head gesture with the help of a single neural network. When a gesture was detected, a further set of 4 neural networks was used to recognize the exact gesture. The final recognized gesture was the one for which the network output score was closest to the training data. All neural networks developed based on default Matlab settings with a single hidden layer of size 4.

Features were computed from the head motion sensor 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 20hz with about 2% error due to missing packets while transmitting the data via the Bluetooth connection. The features used for head gesture recognition 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() is the variance measure taken over acc (accelerometer) and gyr (gyroscope) signals in the sample window. dGyr signals is defined as 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.

Matlab is used for processing the acquired data and interfacing with the wearable system through a TCP-IP connection. In order to connect the head motion sensor to the gesture recognition system, we used a Bluetooth connection.

5 EVALUATION

5.1 Experiment design

As mentioned in section 4.1, the scope of the implementation in this master thesis is limited to study the behavior of the adaptation component through investigating effect of automatic or manual UI adaptation on the performance of human agent. To study the effects of adaptation strategy, we designed and conducted an experiment, in which a human agent performs a physical task (the Hotwire task explained earlier) and at the same time, in parallel, is asked to perform a virtual task (performing simple addition and subtraction calculations) by receiving information about the virtual task, and by driving it forward, using two different interaction modality (combinations). The hypotheses of the experiment are explained in the following sections.

5.1.1 Effect of interaction modality on user performance in Hotwire task

Previous studies [81] on multimodal interaction have shown that in dual-task situations – e.g. situations when a human agent needs to attend to both a physical (real-world) task and a virtual (digital-world) task – when the physical task demands physical attention, it is significantly distractive to also receive information about the virtual task in the visual modality. In such cases, the auditory modality has been proven to be less distractive. Based on these findings, our first hypothesis is defined as follows:

- ***Hypothesis 1****: The performance of a human agent in the Hotwire task is higher when interaction with the system regarding a parallel virtual task is done using auditory modality only (speech for both input and output) compared to a combination of visual output and head-gesture input.*

5.1.2 Cost of switching between modalities

Prior research by cognition community has showed that switching modalities has cognitive costs for human's brain and affects user performance [84]. Based on this finding we expect that in our experiment, switching between on the one hand audio-only, and on the other hand the combined visual output/gesture input modalities, for driving the virtual task forward in parallel with performing the physical Hotwire task would affect the user performance negatively, if compared to sticking to only one of the modality (combinations) throughout the whole task. Therefore, the second hypothesis of the experiment is formulated as follows:

- ***Hypothesis 2****: Changing the modality of interaction for the virtual task will negatively affect the performance of the physical task performed in parallel.*

5.1.3 Automatic or manual UI adaptation

In the real world, when a mobile user interacts with his/her wearable or mobile device, the surrounding context (e.g. noise, light condition, etc.) changes frequently. One can imagine that if given the possibility, the changed context might make the human agent want to stop interacting with the system through the current modality (combination) and select a different one to continue the interaction session in a more comfortable way. To simulate the above mentioned scenario, we arranged so that the audio modality for input and output could be made less desirable by generating disturbing noise (the sound of a passing train) while for the visual output modality we lowered the contrast of the information visualized on the head-mounted display to simulate a sudden increase in ambient light (see Figure 21). When the noise starts, in one condition the system switches the modality automatically, while in the other condition the human agent can change the modality manually. Since one of the most important goals of the automatic adaptation is decreasing the obtrusiveness of the UI [78], our last hypothesis is defined as follows:

- ***Hypothesis 3****: The human agent performance in Hotwire task is higher when the system switches automatically between modalities compared to the condition, in which modalities are switched manually by human agent.*

We intend to verify these hypotheses by measuring human agents' performance in Hotwire task, as an abstract variation of real-world physical tasks, while interacting with a wearable system at the same time via different modalities.

5.2 Method

5.2.1 Participants

8 participants (mean age=30, 2 females) were recruited among local university students to participate in the experiment. Most of the participants were highly skilled computer users ($\bar{x}$= 4.37, σ= 0.91 where the range was 1 to 5), and all of them had perfect visual acuity.

5.2.2 Apparatus

As mentioned in the section 4.2, the Hotwire apparatus has been designed and developed to simulate physical tasks in the real world. The Hotwire apparatus was connected to an ordinary Macbook via serial connection to communicate with the monitoring system.
As mentioned in the section 4.4, the wearable system consisted of a HMD (MicroOptical SV-9, 640x480 pixels) mounted on a safety goggle and an augmented headset with a head motion sensor. The wearable system was connected to the same MacBook as Hotwire apparatus. By starting the Hotwire task, the wearable system starts asking simple math

questions every 7 seconds through visual or audio modalities. When a participant answered a question on the HMD, a black screen was displayed on the HMD until the next question appeared. The questions were presented randomly within subjects in run-time.

Participants were able to provide input to the wearable system through speech modality and head gestures. Four different head gestures were assigned to different commands: the *right* and *left* gestures were defined to choose the correct answer displayed on the HMD (see Figure 21), while the *up* and *down* gestures were assigned to switch the current modality to visual and auditory modalities. The *up* gesture was selected for switching to the visual modality because the eyes are located higher than mouth, and we thought it might be easier to memorize by participants.

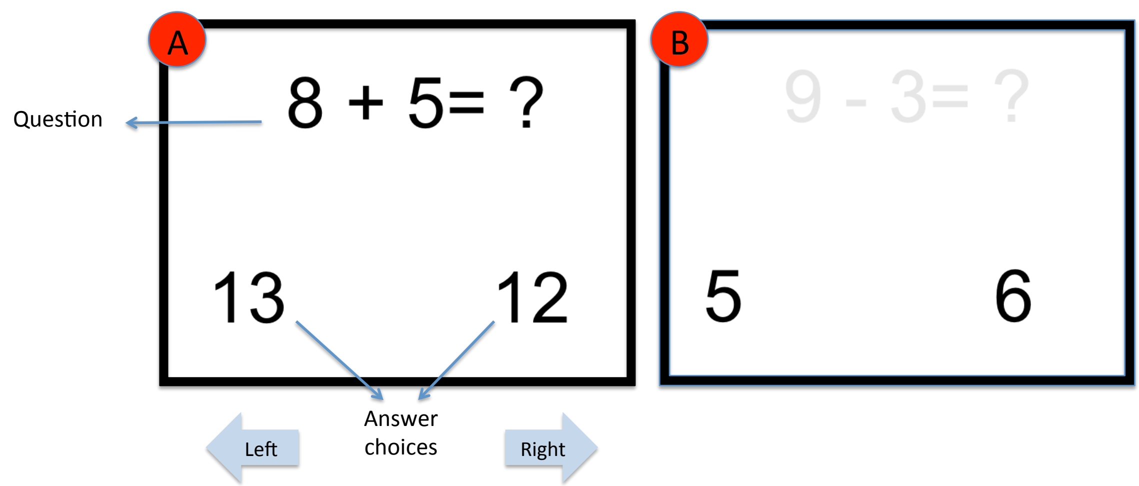

Figure 21. A: Snapshot of the presented information for virtual task on the HMD, B: Snapshot of the lowered contrast

Based on our limited evaluation of implemented speech recognition and gesture detection methods, the accuracy of our speech recognition method using an ordinary microphone and in a quiet room is about 85% while the accuracy of our head gesture detection system is about 78%. Since there is no way to find a perfect speech or gesture recognition system, we used the Wizard-of-Oz approach to control the independent variables during the experiment. That is a wireless ring style mouse was used by the researcher to translate the participants' gestures and speech and send the translated command to the system. The participants were told that the system is working based on their input. Figure 22 shows one the participants wearing the HMD and the headset and performing the Hotwire task.

Figure 22. A participant with the wearable system performing the Hotwire task

5.2.3 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 devices for all conditions; however, they did not need to use all devices in all conditions. After participants were prepared for the experiment, they were asked to use the system until they felt comfortable. This usually took 2-3 minutes. Next, each participant was asked to complete the Hotwire task in five different conditions. In Hotwire task they needed to pass the hand-held stick through the bent wire form start point to the end point as fast as possible without touching the wire. The conditions in which the Hotwire task was completed were as follows:

Condition 1: Hotwire task

Performing the Hotwire task without any interaction with wearable system.

Condition 2: Hotwire task + calculation task on HMD through gestures

Performing the Hotwire task while answering the simple math questions displayed on the HMD through head gestures (*right* and *left*).

Condition 3: Hotwire task + calculation task through audio modality

Performing the Hotwire task while listening to the simple math questions on headset and answering verbally to the questions.

Condition 4: Hotwire task + calculation task + automatic switching modality

In this condition, the participants performed Hotwire task while answering simple math questions. The questions started on audio modality. After the third question the noise (train sound) was played. For the next question, system switched automatically to the visual modality. When the participants saw the questions on the HMD they could answer through head gestures. After the second question on the HMD the color of the graphics on the HMD changed to a bright color to simulate the difficult light condition for the HMD. For the next question, system switched automatically to the audio modality. The questions were presented in a predefined order until the Hotwire task finished.

Condition 5: Hotwire task + calculation task + manual switching modality

In this condition, the participants performed Hotwire task while answering simple math questions. The questions started on audio modality similar to the 4th condition. After the third question the noise (train sound) was played. In this condition the participant had to switch to visual modality by performing the *up* gesture. When the participants saw the questions on the HMD they could answer through head gestures. After the second question on the HMD the color of the graphics on the HMD changed to a bright color similar to the previous condition. This time, the participant had to switch to the auditory modality by performing the *down* gesture. The questions were presented in the same order as the 4th condition until the Hotwire task finished.

To investigate the learning effect of the task, after finishing the 5th condition, we asked the participants to repeat the first condition again. After the tasks were completed for all conditions, the participant was asked to complete a short questionnaire (Appendix I) with 5-point likert scale questions polling their experiences completing the task and using the system.

5.2.4 Design

The experiment was an 8×5 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 $\times$ 5 conditions = 40 trials.

5.3 Results

5.3.1 User performance

We measured the task completion time and the number of errors (touching the wire) for each condition. Since the user performance is a combination of task completion time and errors, we defined the performance as follows:

Task completion time $= t_{start} - t_{end}$

Total error $= \sum_{i=t_{start}}^{t_{end}} contact\ (i)$

Performance $= Task\ completion\ time + Total\ error \times 4.72$

To calculate the performance, we used the regression coefficient between error and time, which is equal to 4.72. Figure 23shows the average user performance in each condition.

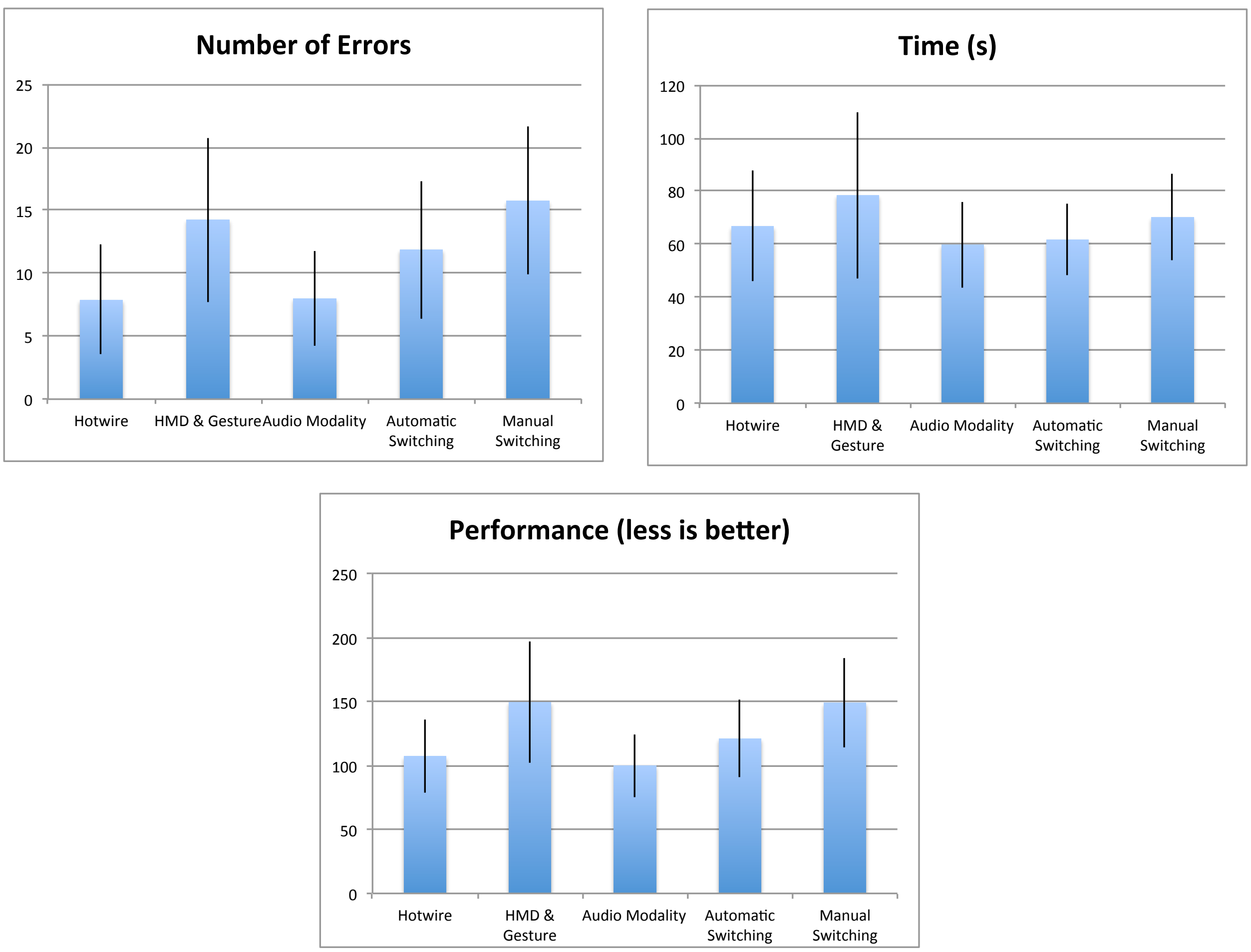

Figure 23. Average of user performance in each condition

To verify the *Hypothesis 1* (page 48), we used the t-test to compare user performance, task completion time, and error between conditions 2 and 3. The statistical test showed that there is a significant difference in user performance ($t(14) = 2.61$, $p = .01 < .05$) and error (t(14) = 2.34, p =.017 < .05) between conditions 2 and 3; however, no significant difference was indicated in task completion time between conditions 2 and 3 (t(14) = 1.485, p = .079 > .05 with confidence interval of 95%). In other words, the performance of user in Hotwire task is significantly higher when interacts with the system through auditory modality compared to the condition in which the HMD is used as output device and head gestures as input modality.

To verify the *Hypothesis 2* (page 48), the user performance in conditions 4 and 5 were compared with user performance in conditions 2 and 3. The two by two t-test (see Table 2 in which the significant differences are bolded) showed that the user performance in condition 5 is significantly less than the user performance in condition 3 (t(14) = 3.59, p =.001 < 0.05, with 95% confidence interval). No other significant difference was indicated. To put is simply, the cognitive cost of switching between modalities was not detected.

Performance	Condition 1	Condition 2	Condition 3	Condition 4	Condition5
Condition 1		**t(14) = 2.19, p = .022 < .05**	t(14) = .49 p = .31 > .05	t(14) = .98, p = .17 > .05	**t(14) = 2.95 , p = .005 < .05**
Condition 2	**t(14) = 2.19, p = .022 < .05**		**t(14) = 2,61 p = . 01< .05**	t(14) = 1.45, p = .08 > .05	t(14) = .06, p = . 47 > .05
Condition 3	t(14) = .49 p = .31 > .05	**t(14) = 2,61 p = . 01< .05**		t(14) = 1.52, p = . 07 > .05	**t(14) =3.59 , p = .001 < .05**
Condition 4	t(14) = .98, p = .17 > .05	t(14) = 1.45, p = .08 > .05	t(14) = 1.52, p = . 07 > .05		**t(14) = 1.96, p = .03 < .05**
Condition 5	**t(14) = 2.95 , p = .005 < .05**	t(14) = .06, p = . 47 > .05	**t(14) =3.59 , p = .001 < .05**	**t(14) = 1.96, p = .03 < .05**	

Table 2. Performance comparision between different conditions

The *Hypothesis 3* (page 49) is about comparison of user performance between automatic and manual switching conditions (comparing between conditions 4 and 5). The result of t-test indicated a significant difference in user performance (t(14) = 1.96, p = .03 < .05) between conditions 4 and 5; however, the task completion time was not significantly different in conditions 4 and 5 or the difference was not detected (t(14) = 1.13, p = .13 > .05). Also there is a tendency to lower error in automatically switching modalities compared to the manual switching condition (t(14) = 1.62, p = .06 > .05, confidence interval 95%). In other words, the automatic switching between modalities has a positive effect on the user performance in Hotwire task compared to the manual switching condition.

5.3.2 Questionnaire

5 out 8 participants preferred automatically switching modality to the manual condition, as they argued that the automatic switching was easier and decreases the extra interaction with the system, while the other group wanted to have control over the system. The result of the questionnaires (Table 3) shows that the first condition was selected as the easiest task by most of the participants ($\bar{x} = 2$, $\sigma = 0.75$), while the manual switching was indicated as the most difficult one ($\bar{x} = 4.25$, $\sigma = 0.70$).

Question: How easy/difficult did you find the task? (1 easy to 5 difficult)	**Min**	**$\bar{x}$**	**Max**	**σ**
Hotwire task alone	1	2	3	0.75
Hotwire + HMD & Gesture	2	3.6	4	0.74
Hotwire + Audio modality	2	2.87	3	0.35
Hotwire + Automatically switching	3	4.12	5	0.64
Hotwire + Manually switching	3	4.25	5	0.70

Table 3. The questionnaire result

5.4 Discussion

The result of the performance comparison between different conditions verified our first hypothesis about higher performance of the user in Hotwire task while interacting through auditory modality. The main reason for this finding is the fact that the Hotwire task needs visual attention and physical stability –like many real-world tasks– while reading the question on the HMD interferes with visual focus on Hotwire task. This means the participants need to stop the Hotwire task to look at the HMD, which increases the task completion time. Aside from that, moving the head by the participant to preform the gestures decreases the physical stability of the participant and increases the error rate. According to Table 2, our second hypothesis is rejected. Given previous studies by cognition science community, switching modality from visual to audio and vice versa decreases human's performance. We expected to observe a higher performance in unimodal conditions (2 and 3) compared to the multimodal ones (conditions 4 and 5). However, the statistical tests have not proved any significant difference between unimodal and multimodal conditions except between unimodal auditory (condition 2) and multimodal with manual switching (condition 5). The considerable difference between conditions 2 and 5 is observed due to the significant difference between interacting through auditory and visual modalities while performing the Hotwire task. Since in the 5th condition, in addition to providing extra input for switching modality, the participant needed to interact through the combination of HMD and gesture modalities, which is much harder than auditory modality. In other words, the cost of switching between modalities was hidden by the significant difference between auditory and visual modalities.

The statistical tests verified our last hypothesis about superiority of the automatic to the manual switching modalities. The main reason of the higher performance can be unobtrusiveness of automatic switching which did not need an extra command to change the modality. However, the result of the completed questionnaires indicated that some of the participants (3 out 8) preferred the manual switching since the preferred to be able to control the interaction. In condition 5 (manual switching), two of the participants did not switch to the HMD. Even though answering to the questions was not easy with the background noise ($\bar{x}$ = 3.25, σ = 0.88). They argued that they preferred the background noise to distracting HMD. Which means different human agents has different preferences for adaptation strategy; even if, the performance could be lower in the preferred modality.

6 CONCLUSIONS

6.1 Summary

The first goal of this thesis (page 2) is design an architecture for wearables to support mobile interaction through multimodal adaptation of the UI in different situations. The first research question was how dynamic adaptation of UI can be implemented? To answer this question, we reviewed the literature in three main research areas:

1- Wearable computing: because wearable computers are envisioned to support mobile users and mobile interaction

2- Context-aware computing: since one of the main aspects of UI adaptation is context-awareness

3- Multimodal interaction: because in mobile interaction, there is no perfect modality for all situations, and an important aspect of the adaptation is switching from an obtrusive modality (in a particular situation) to a proper one

The literature of wearable computing helped us understand the main interaction concepts and roles of computing devices to support mobile users. By reviewing the literature of multimodal interaction and adaptive systems, we learned how a model driven approach could be used for dynamic UI adaptation, and finally the context-aware computing helped us to identify different aspects of the context as the main cause of adaptation. Based on the findings from literature, the main aspects of dynamic multimodal UI adaptation for mobile interaction were extracted and used to analyze the design space. The design space analysis helped us to transfer the knowledge we learned from literature to design a middleware for dynamic multimodal adaptation. The adaptation middleware is conceptually designed as a part of egocentric interaction architecture. One of the important findings from literature is the usability challenge of UI adaptation. While the main goal of adaptation is decreasing the obtrusiveness on the UI, other usability principles such as predictability, controllability, and transparency can be violated through the adaptation process. This led us to investigate the behavior of the adaptation component within the proposed architecture by conducting an experiment to investigate the effects and practical implications of multimodal UI adaptation on user performance in mobile interaction scenarios. We simulated a typical mobile interaction scenario by designing a combination of physical and virtual task. The participants performed a physical task while interacting with a wearable system through different modalities. The result of our experiment showed the fact that automatic adaptation of the UI can increase the performance of user in mobile interaction scenarios while some users proffered the manual adaptation. The result of our user study also indicates the controllability challenge of UI adaptation.

The main contribution of this master thesis is clarifying the design space of multimodal UI adaptation for wearable computers in mobile interaction scenarios. Another important contribution is revealing some of the usability challenges of automatic adaptation, which can inform the next steps of implementing an overall working architecture.

6.2 Future Work

The next steps to arrive to an overall working architecture is designing the adaptation patterns (described in section 3.2.4) for mobile interaction scenarios through more experimental studies, and also conducting user studies in real applications, which helps us extract general rules for multimodal UI adaptation of wearable systems in mobile interaction scenarios.

REFERENCES

1. Kleinrock, L. Nomadic computing an opportunity. SIGCOMM Comput. Commun. Rev. 25, 1, pp. 36-40, 1995.

2. Roth, J. Patterns of mobile interaction. *Personal and Ubiquitous Computing*, 6(4), pp. 282–289, 2002.

3. Marshall, J., Tennent, P. Mobile interaction does not exist. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13).* ACM, New York, NY, USA, pp. 2069-2078, 2013.

4. Billinghurst, M., Interaction on the Go, retrived on 10th Dec 2013, http://www.ep.liu.se/ecp/011/001/ecp011001b.pdf

5. V. Bellotti, S. Bly. Walking away from the desktop computer: Distributed collaboration and mobility in a product design team. In K. Ehrlich and C. Schmandt, editors, In *Proceedings of CSCW'96*, ACM, ACM Press, pp. 209–218, 1996.

6. Luff,P., Heath, C. Mobility in collaboration. In *Proceedings of the CSCW '98*. ACM, New York, NY, USA, pp. 305-314, 1998.

7. Surie, D. Egocentric interaction for ambient intelligence. *Diss*. Umeå University, 2012.

8. Mann, S. Smart Clothing: The Shift to Wearable Computing. In *Communications of the ACM*, 39 (8), pp. 23-24, 1996.

9. Mann, Steve, and Hal Niedzviecki. *Cyborg: digital destiny and human possibility in the age of the wearable computer*, Doubleday Canada, 2001.

10. Starner, T. The Challenges of Wearable Computing: Part 1. *IEEE Micro*, 21, pp. 44- 52, 2001.

11. Schilit, B., Adams, N., & Want, R. Context-aware computing applications. In *Proceedings of WMCSA '94, First Workshop on Mobile Computing Systems and Applications,* IEEE, pp. 85-90, *1994.*

12 Dey, A. K., Abowd, G. D., Salber, D. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-computer interaction*, 16(2), pp. 97-166, 2001.

13. Abowd,G., Mynatt, E. Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)*. 7, pp. 29-58, 2000.

14. Barkhuus, L., & Dey, A. Is context-aware computing taking control away from the user? Three levels of interactivity examined. In *UbiComp 2003*. Springer Berlin Heidelberg, pp. 149-156, 2003.

15. Richard A. Bolt. “Put-that-there”: Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques (SIGGRAPH '80)*. ACM, New York, NY, USA, pp. 262-270, 1980.

16. Tzovaras, Dimitrios, ed. *Multimodal User Interfaces: From Signals to Interaction*. Springer, 2008.

17. Alejandro Jaimes and Nicu Sebe. Multimodal human-computer interaction: A survey. *Comput. Vis. Image Underst.* 108, 1-2, pp. 116-134, 2007.

18. Poupyrev, I., Weghorst, S., Billinghurst, M., and Ichi- kawa, T. Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. *Computer Graphics Forum*, 17(3). pp. 41–52, 1998.

19. Pederson, T. Egocentric Interaction, *Workshop on What is the Next Generation of Human-Computer Interaction*. 2006.

20. Pederson, T., Janlert, L. E., & Surie, D. Towards a model for egocentric interaction with physical and virtual objects. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pp. 755-758, 2010.

21. Clark, A.F., Newman, N.; Isaakidis, A.; Pagonis, J. What do we want from a wearable user interface?, *Eurowearable*, IEEE, pp. 23-27, 2003.

22. Billinghurst, M., Bowskill, J., Dyer, N., Morphett, J. An Evaluation of Wearable Information Spaces, in *Proc. IEEE Virtual Reality and Artificial Intelligence Symp.* (VRAIS 98), IEEE CS Press, Los Alamitos, Calif., pp. 20-27, 1998.

23. Smailagic, A., & Siewiorek, D. Application design for wearable and context-aware computers. *Pervasive Computing*, IEEE, 1(4), pp. 20-29, 2002.

24. Jalaliniya, S., Pederosn, T., Houben, S. Wearable Laser Pointer as good as Head-Mounted Display for Tele-Guidance Applications? , *Augmented Human*, ACM, 2014. (submitted)

25. Alem,L., Haung, W.,Tecchia, F. Supporting the Changing Roles of Maintenance Operators in Mining: A Human Factors Perspective, *The Ergonomics Open Journal*, 4, (Suppl 2-M4), pp. 81-92, 2011.

26. Ashbrook, D. Enabling Mobile Microinteractions. PhD Thesis, 2010.

27 . Clawson, J., Lyons, K., Starner, T., Clarkson, E. The impacts of limited visual feedback on mobile text entry for the twiddler and mini-qwerty keyboards. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, pp. 170–177, 2005.

28 . Lyons, L., Gane, B., Starner T., Catrambone, R. Improving novice performance on the twiddler one-handed chording keyboard. In *Proceedings of the International Forum on Applied Wearable Computing*, IEEE Computer Society, pp. 145–159, 2005.

29 . Lyons, K., Plaisted, D., & Starner, T. Expert chording text entry on the twiddler one-handed keyboard. In Wearable Computers, ISWC 2004. *Eighth International Symposium on Wearable Computers (ISWC'04)*, Washington, DC, USA, IEEE Computer Society, pp. 94-101, 2004.

30 Ayres, R. U. *On the Reappraisal of Microeconomics: Economic Growth and Change in a Material Change*. K. Martinás (Ed.). Edward Elgar Publishing, p. 41 , 2005.

31 Karat, C. M., Halverson, C., Horn, D., & Karat, J. Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, pp. 568-57, 1999.

32 R. Rosenberg and M. Slater. The chording glove: a glove-based text input device. *IEEE Transactions on Systems*, Man, and Cybernetics, Part C, 29(2), pp. 186–191, 1999.

33 Witt, H., Leibrandt, R., Kemnade, A., & Kenn, H. Scipio: A miniaturized building block for wearable interaction devices. In *Proceedings of the 3rd International Forum on Applied Wearable Computing (IFAWC)*, pp. 1-5, 2006

34 Wusheng, C., Tianmiao, W., & Lei, H. (2003, March). Design of data glove and arm type haptic interface. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems*. HAPTICS 2003. Proceedings. 11th Symposium on, IEEE. pp. 422-427, 2003

35 Starner, T., Auxier, J., Ashbrook, D., & Gandy, M. The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. In *Proceedings of the 4th IEEE International Symposium on Wearable Computers, ISWC '00,* IEEE, pp. 87-94, 2000.

36 . Rekimoto, J. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Proceeding 5th International Symposium on Wearable Computers (ISWC 2001)*, pp. 21-27, 2001.

37. Jalaliniya, S., Smith, J., Sousa, M., Büthe, L., & Pederson, T. Touch-less interaction with medical images using hand & foot gestures. In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, ACM, pp. 1265-1274, 2013.

38. Kim, J., Mastnik, S., & André, E. EMG-based hand gesture recognition for realtime biosignal interfacing. In *Proceedings of the 13th international conference on Intelligent user interfaces*, ACM, pp. 30-39, 2008.

39 . Mardanbegi, D., Hansen, D. W., & Pederson, T. Eye-based head gestures. In Proceedings of the Symposium on Eye Tracking Research and Applications ACM, pp. 139-146, 2012.

40 . Schmidt, A. Implicit human computer interaction through context. *Personal Technologies,* 4(2-3), pp. 191-199, 2000.

41 Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., Baudisch, P. Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, pp. 1395-1404, 2007.

42 Baber, C. Wearable computers: A human factors review. *International Journal on Human Computer Interaction*, 13(2), pp. 123–145, 2001.

43 NRC. *Tactical display for soldiers: Human factors considerations*. National Academy Press, Washington, DC 1997.

44 Laramee, R. S., & Ware, C. Rivalry and interference with a head-mounted display. ACM, *Transactions on Computer-Human Interaction*, 9(3), pp. 238-251, 2002.

45 Rolland, J. P., Fuchs, H. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3), pp. 287-309, 2000.

46 Narayanaswami, C., & Raghunath, M. T. Application design for a smart watch with a high resolution display. In *Proceeding of The Fourth International Symposium on Wearable Computers*, IEEE, pp. 7-14, 2000.

47 Karitsuka, T., & Sato, K. A wearable mixed reality with an on-board projector. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, p. 321, 2003.

48 Mistry, P., Maes, P., and Chang, L. WUW - wear Ur world: a wearable gestural interface. In *Proceeding CHI '09 Extended Abstract*, pp. 4111–4116, 2009.

49 Harrison, C., Benko, H., & Wilson, A. D. OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, ACM, pp. 441-450, 2011.

50 McFarlane, D. C., & Wilder, S. M. Interactive dirt: increasing mobile work performance with a wearable projector-camera system. In *Proceedings of the 11th international conference on Ubiquitous computing*, ACM, pp. 205-214, 2009.

51 Schwerdtfeger, B., Pustka, D., Hofhauser, A., Klinker, G. Using laser projectors for augmented reality. In *Proc. of VRST '08.* ACM, New York, NY, USA, 134-137, 2008.

52 Sakata, N., Kurata, T., and Kuzuoka, H. Visual assist with a laser pointer and wearable display for remote collaboration. In *Proc. 2nd international conference on Collaboration Technologies*, 2006.

53 Lingley, A. R., Ali, M., Liao, Y., Mirjalili, R., Klonner, M., Sopanen, M., Parviz, B. A. A single-pixel wireless contact lens display. *Journal of Micromechanics and Microengineering*, 21(12), 125014, 2011.

54 Rocchesso, D., Bresin, R., Fernstrom, M. Sounding objects. *IEEE MultiMedia*, 10(2), pp. 42–52, 2003.

55 Brewster, S. A. Non-speech auditory output. *Human Computer Interaction Handbook*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 220-239, 2002.

56 Kern, T. A. *Engineering haptic devices*. Springer, p. 22, 2009.

57 Iwata, H. Haptic interfaces. In Julie A. Jacko and Andrew Sears, editors, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, chapter 10, Lawrence Erlbaum Associates, pp. 206–219, 2003.

58 Tan, H. Z., Pentland, A. Tactual displays for wearable computing. *Personal Technologies*, 1(4), pp. 225-230, 1997.

59 Tsukada, K., & Yasumura, M. Activebelt: Belt-type wearable tactile display for directional navigation. In *UbiComp 2004: Ubiquitous Computing*. Springer Berlin Heidelberg, pp. 384-399, 2004.

60 Stevenson, G., Riener, A., Ferscha, A. "TactiGlove"–A Guidance System to Effectively Find Hidden Spots in 3D Space. In *Mobile Computing, Applications, and Services*, Springer Berlin Heidelberg, pp. 80-99, 2013.

61 Schmidt, A. Interactive Context-Aware Systems Interacting with Ambient Intelligence, *Ambient Intelligence*, IOS Press pp. 159-178, 2005.

62 Ishii, H., & Ullmer, B. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, ACM, pp. 234-241, 1997.

63 Weiser, M. The computer for the 21st century. *Scientific american*, 265(3), pp. 94-104. 1991.

64 Brooks, S. J., Savov, V., Allzen, E., Benedict, C., Fredriksson, R., Schiöth, H. B. 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*, 59(3), pp. 2962-2973, 2012.

65 Chou, W.-L., & Yeh, S.-L. Subliminal spatial cues capture attention and strengthen between-object link. *Consciousness and Cognition*, 20, pp. 1265–1271, 2011.

66 Lehikoinen, J., Holopainen, J., Salmimaa, M., Aldrovandi, A. MEX: a distributed software architecture for wearable computers. In *Proceeding of The Third International Symposium on Wearable Computers*, Digest of Papers, IEEE, pp. 52-57, 1999.

67 Song, Y., Moon, S., Shim, G., Park, D. Mu-ware: a middleware framework for wearable computer and ubiquitous computing environment. In *Pervasive Computing and Communications Workshops*. IEEE, pp. 455-460. 2007.

68 Muro, M., Amoretti, M., Zanichelli, F., Conte, G. Towards a flexible middleware for context-aware pervasive and wearable systems. *Medical & biological engineering & computing,* 50(11), pp. 1127-1136, 2012.

69 Bardram, J. E. The Java Context Awareness Framework (JCAF)–a service infrastructure and programming framework for context-aware applications. In *Pervasive Computing*, Springer Berlin Heidelberg. pp. 98-115, 2005.

70 . Surie, D., Janlert, L. E., Pederson, T., Roy, D. Egocentric interaction as a tool for designing ambient ecologies—The case of the easy ADL ecology. *Pervasive and Mobile Computing*, 8(4), pp. 597-613, 2012.

71. Pederson, T., L-E. Janlert, Surie, D. A Situative Space model for mobile mixed-reality Computing. *Pervasive Computing, IEEE* 10.4, pp. 73-83, 2011.

72 . Surie, D., Pederson, T., & Janlert, L. E. The easy ADL home: A physical-virtual approach to domestic living. *Journal of Ambient Intelligence and Smart Environments*, 2(3), pp. 287-310, 2010.

73 . Witt, H. User Interfaces for Wearable Computers, PhD thesis, p. 177, 2007.

74 . Fickas, S., Korteum, G., Segall, Z. Software Organization for Dynamic and Adaptable Wearable Systems. 1st International Symposium on Wearable Computers, pp. 56-63, 1997.

75 S. Dusan, G.J. Gadbois, J. Flanagan, Multimodal interaction on PDA's integrating speech and pen inputs,In *Proceeding of Eurospeech*, 2003.

76 Brewster, S., Lumsden, J., Bell, M., Hall, M., Tasker, S. Multimodal eyes-free interaction techniques for wearable devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 473-480, 2003.

77 P. Langley. User modeling in adaptive interfaces. In *Proceedings of the 7th International Conference on User Modeling*, UMAP'99, Secaucus, NJ, USA, pp. 357-370,1999.

78 Bezold, M., & Minker, W. *Adaptive multimodal interactive systems*. Springer, 2011.

79 B. Shneiderman, Direct manipulation for comprehensible, predictable, and controllable user interface, *International Conference on Intelligent User Interfaces*, 1997.

80 . McFarlane, D. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. Human-Computer Interaction 17, 1, pp. 63-139, 2002.

81 . Zhao, S., Brumby, D. P., Chignell, M., Salvucci, D., Goyal, S. Shared Input Multimodal Mobile Interfaces: Interaction Modality Effects on Menu Selection in Single-Task and Dual-Task Environments, *Interacting with Computers*, 2013.

82 . Witt, H., Drugge, M. Hotwire: An apparatus for simulating primary tasks in wearable computing. In *CHI'06 extended abstracts on Human factors in computing systems*, ACM, pp. 1535-1540, 2006.

83 . Barrena, S., Klotz, L., Landes, V., Page, A., & Sun, Y. Designing Android applications with both online and offline voice control of household devices. In *Bioengineering Conference (NEBEC), 38th Annual Northeast*, pp. 319-320, IEEE, 2012.

84. Sandhu, R., Dyson, B. Modality and task switching interactions using bi-modal and bivalent stimuli*, Brain and Cognition Journal*, 2013.

APPENDIX I: QUESTIONNAIRE

Age				
Sex				
Computer Literacy				
Novice 1	Beginner 2	Average 3	Professional 4	Expert 5
How often do you use computing devices in motion (walking, running, biking, ...)?				
Never 1	Rarely 2	Sometimes 3	Frequently 4	Mostly 5

	How easy/difficult did you find the tasks?				
Task	Very easy 1	Easy 2	Average 3	Hard 4	Very hard 5
1-Hotwire task alone					
2-Hotwire + Visual questions (HMD)					
3-Hotwire + Audio-based questions (Headset)					
4-Hotwire task + switching automatically between HMD & Audio					
5- Hotwire task + switching manually between HMD & Audio					

In tasks 4 & 5, it was hard to perform the task when there was noise?				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

In tasks 4 & 5, it was hard to perform the task when there was visual problem?				
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Which approach did you prefer to switch between HMD and Audio-based interactions: automatically switching between HMD and Audio (4) or manually switching (5)?	
Manual switching ☐	Automatically switching ☐

Why do you prefer the above approach to switch between HMD & Audio?

Additional comments: