

AUGMENTED REALITY BASED USER INTERFACE FOR MOBILE APPLICATIONS AND SERVICES

**PETER
ANTONIAC**

Faculty of Science,
Department of Information
Processing Science,
Infotech Oulu,
University of Oulu

OULU 2005



PETER ANTONIAC

**AUGMENTED REALITY BASED USER
INTERFACE FOR MOBILE
APPLICATIONS AND SERVICES**

Academic Dissertation to be presented with the assent of
the Faculty of Science, University of Oulu, for public
discussion in Auditorium IT115, Linnanmaa, on June 17th,
2005, at 12 noon.

OULUN YLIOPISTO, OULU 2005

Copyright © 2005
University of Oulu, 2005

Supervised by
Professor Petri J. Pulli

Reviewed by
Associate Professor Yoshiro Ban
Professor Tapio Takala

ISBN 951-42-7695-7 (nid.)
ISBN 951-42-7696-5 (PDF) <http://herkules.oulu.fi/isbn9514276965/>
ISSN 0355-3191 <http://herkules.oulu.fi/issn03553191/>

OULU UNIVERSITY PRESS
OULU 2005

Antoniac, Peter, Augmented reality based user interface for mobile applications and services

Faculty of Science, Department of Information Processing Science, University of Oulu, P.O.Box 3000, FIN-90014 University of Oulu, Finland, Infotech Oulu, University of Oulu, P.O.Box 4500, FIN-90014 University of Oulu, Finland

2005

Oulu, Finland

Abstract

Traditional design of user interfaces for mobile phones is limited to a small interaction that provides only the necessary means to place phone calls or to write short messages. Such narrow activities supported via current terminals suppress users from moving towards mobile and ubiquitous computing environments of the future. Unfortunately, the next generation of user interfaces for mobile terminals seems to apply the same design patterns as commonly used for desktop computers. Whereas the desktop environment has enough resources to implement such design, capabilities of the mobile terminals fall under constraints dictated by mobility, like the size and weight. Additionally, to make mobile terminals available for everyone, users should be able to operate them with minimal or no preparation, while users of desktop computers will require certain degree of training.

This research looks into how to improve the user interface of future mobile devices by using a more human-centred design. One possible solution is to combine the Augmented Reality technique with image recognition in such a way that it will allow the user to access a "virtualized interface". Such an interface is feasible since the user of an Augmented Reality system is able to see synthetic objects overlaying the real world. Overlaying the user's sight and using the image recognition process, the user interacts with the system using a combination of virtual buttons and hand gestures.

The major contribution of this work is the definition of the user's gestures that makes it possible for human-computer interaction with such Augmented Reality based User Interfaces. Another important contribution is the evaluation on how mobile applications and services work with this kind of user interface and whether the technology is available to support it.

Keywords: applications and services, augmented reality, deviceless interface, gesture recognition, human-computer interaction, mobile, mobile device, sign interpretation, user interface

Abbreviations

2D	Two Dimension
3D	Three Dimension
3DOF	Three Degrees of Freedom
6DOF	Six Degrees of Freedom
1G	First Generation
AR	Augmented Reality
AWT	Abstract Window Toolkit
CAD	Computer Aided Design
CD-ROM	Compact Disk - Read Only Memory
CPU	Central Processing Unit
CRT	Cathode-Ray Tube
CSCW	Computer Supported Collaborative Work
DGPS	Differential Global Positioning System
DMA	Direct Memory Access
DOV	Depth of View
DSP	Digital Signal Processing
FOV	Field of View
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications
GUI	Graphical User Interface
HCI	Human Computer Interface
HMD	Head-Mounted Display
HOE	Holographic Optical Element
HSI	Hue-Saturation-Intensity
HSV	Hue Saturation Value
IA	Information Appliance
IEEE	Institute of Electrical and Electronics Engineers
I/O	Input/Output
IWAR	International Workshop on Augmented Reality
JNI	Java Native Interface

LAN	Local Area Network
LCD	Liquid-Crystal Display
LED	Light Emitting Diode
LiIon	Lithium Ion (battery)
LOD	Level of Detail
LOE	Level of Error
MAR	Mobile Augmented Reality
MARS	Mobile Augmented Reality System
MARISIL	Mobile Augmented Reality Interface Sign Interpretation Language
MIT	Massachusetts Institute of Technology
mPARD	A mobile Passive Augmented Reality Device
MR	Mixed Reality
NAIST	Nara Institute of Science and Technology
NTE	Near-the-Eye
OLED	Organic Light Emitting Diodes
OS	Operating System
PAN	Personal Area Network
PAULA	Personal Access and User Interface for Multi-modal Broadband Telecommunication
PCMCIA	Peripheral Component Microchannel Interconnect Architecture
PDA	Personal Digital Appliance
POTS	Plain Old Telephone System
QFD	Quality Function Deployment
RGB	Red-Green-Blue
RPC	Remote Procedure Call
SA	Selective Availability (in GPS accuracy)
SDK	Software Development Kit
UI	User Interface
USB	Universal Serial Bus
VE	Virtual Environment
VoIP	Voice over Internet Protocol
VM	Virtual Machine (as in Java VM)
VR	Virtual Reality
VTT	State Research Center of Finland
WAN	Wide Area Network
WAP	Wireless Application Protocol
WIMP	Windows, Icons, Menus, Pointing devices

Preface

In my earlier education to become a master of engineering in theory of systems and computers science, one of my research colleagues from the laboratory suggested that I had potential to become a real scientist and that I should follow this inclination. At that time, my personal goal was to complete my courses and to earn the engineering degree. Even so, the words remained in my subconscious.

Before relating the story behind the work for this thesis, I would like to introduce the motto that governed my life during the research period: “It is characteristic for the whole world of ideas that they do not come as memories from the past but expectations for the future” – William Stern. I request the reader to look into the expectation of the future rather than the past or present.

My work as a scientist started as a research assistant at the Digital Signal Processing Laboratory in the “Politehnica” University of Bucharest, Romania. That was my first contact with researchers and first assignment as a researcher. The work on this thesis can be traced back to the years 1997-1998 when, as a research scientist for the European Project, ESPRIT No. 25946 (CE-Net Project), I was designing and building a user interface for the “Who is Who” database. The work comprised of collecting data and building a friendly and lightweight web interface to access it. The experience gained from this project guided me towards designing the small interface for browsing the data that was later useful in another research project called Cyphone (Personal Virtual Services Based on Picocellular Networks, a project funded by the Technology Development Center of Finland - TEKES) at the University of Oulu.

In Cyphone, the research problem was to design a small device with a small screen size that would be able to display a large amount of data in a comprehensible way. At the commencement the research, the platform for development was the wireless application protocol (WAP) phone and the interface was based on the wireless mark-up language (WML). The constraints identified for the research platform were: computing power of the device and limited physical size of the screen. The computing power was to be solved in the future. The size of the screen, even with a higher resolution, was the main concern. To achieve a better representation of the data in a small space, the best solution was to have fewer characters on the display and the best solution was to represent the

information within a table, resulting in a high amount of information per displayed page. The table could also allow easy reordering of information and easy browsing.

Later work in the Personal Access and User Interface for Multi-modal Broadband Telecommunication (Paula) Project, funded by Academy of Finland, focused on mobile devices and how to use them for other purposes than voice communication and to take advantage of broadband communication. The project aim was to look into the future use of mobile phones and unearth new applications to take advantage of the better communication bandwidth that was being forecast. The groups, at University of Oulu and the State Research Center of Finland (VTT), were focused on *Navigation* and *Meetings*. The task that I was involved with was to deliver easy implementation of the user interface for the navigation system of a future “media-phone” device. When studying what was the best design solution to be implemented for the future, the size of the display was identified as the main constraint. Later, after deeper analysis of the problem, I proposed the use of augmented reality and sign language as the base design for the user interface. By applying this technique, the display’s size could be extended to all the surrounding area of the user, and the user would be able to interact in a very ergonomic way.

The idea was embraced and encouraged by Professor Petri Pulli. A patent (see Appendix) was applied for and I started work to write the scenario and produce a multimedia demo in order to reveal the technique in a more exhaustive form. The purpose of the movie was to emphasise, in a visual way, the manner in which this type of devices could work. Professor Pulli and myself presented the demo movie to several audiences around Europe, Japan and United States of America. We received enough positive feedback to continue the research and to start creating a prototype. Later, after constructing the prototype, my research focused on how to apply it and extend it to work with more applications and services. This thesis discusses in depth the consequences of this research on user interfaces based on augmented reality, how to implement them and what applications and services could be used with them.

Infotech Oulu (Virgin and Interact groups), the Academy of Finland and the Nokia Foundation, financially supported this work. For the research years and the work invested in this thesis I hereby acknowledge publicly my humble appreciation to all my colleagues, friends and family that helped, encouraged and stimulated me to pursue and finish this work. An incomplete list of people follows in the acknowledgement’s section of this manuscript.

Acknowledgments

I wish to thank the following people for their help and advice:

- My advisor and mentor, Prof. Petri Pulli;
- Preliminary examiner, Prof. Tapio Takala, Dept. of Computer Science, Helsinki University of Technology, Espoo, Finland;
- Preliminary examiner, Associate Prof. Yoshiro Ban, Information Science and Technology Center, Kobe University, Kobe, Japan;
- Prof. Dr. Jouni Similä, for his help and understanding during the time of research and writing the thesis;
- Prof. Emer. Pentti Kerola, for his friendly and constant support on the research problems encountered;
- Prof. Kari Kuutti, for his advice and ideas on the human-computer interaction issues and for guiding me into the field;
- Prof. Timo Jokela, for his clear mind over the usability issues;
- Prof. Dr. Aurelian Stanescu, for his constant concern on the status of this thesis;
- Dr. Hiroshi Sasaki, for sharing his ideas and effort involved in the cooperation and development of the software for the prototypes and without him most of the work would not have been realized;
- Seamus Hickey, for his advice, reviews and feed-back given during the period of research and writing of this manuscript;
- Dr. Tony Manninen, for his help and advice concerning interactions and other related topics;
- Dan Bendas, for his contribution, advice and friendship during the research and development stage of the work;
- Isabella Ion, for her help and ideas concerning the design and graphics included in this manuscript;
- Dr. Tomohiro Kuroda, for his close look at my work and his sharp comments;
- Dr. Tuomo Tuikka, for his faith in my work and for supporting my research with valid information from his past experience;
- Marko Salmela, for his persuasion and passion over the use of Java;
- Mikko Kerttula, for his friendship and ideas for applications;
- Tino Pyssysalo, for his help and contribution at the start of the work;

- Jukka-Pekka Metsävainio, for the design and futuristic visions;
- Marc Pallot, for his friendship and support during the research;
- Jukka Ahola, for his encouragement and suggestions, as well as support even though it was across the Atlantic;
- Prof. Mikko Siponen, for his kind help and patience during the editorial work;
- Jacob Matthan for his kind help with the proof reading;
- All my colleagues from Dept. of Information Processing Science;
- My fiancée, Meri Ruuskanen, for encouraging, understanding, making suggestions and assistance during the writing of the thesis; and
- My mother, for understanding the situation and supporting my dream during the duration of the studies even though being so far away.

Table of Contents

Abstract	
Abbreviations	
Preface	
Acknowledgments	
Table of Contents	
1 Introduction.....	15
1.1 Research Question	17
1.2 Research Approach	18
1.3 Survey of Earlier Work	21
1.3.1 Augmented Reality History.....	25
1.3.2 Virtual User Interfaces	34
1.3.3 Mobile Augmented Reality Interfaces	35
1.4 Definitions	36
1.5 Publications and Dissemination of the Research Results.....	38
1.6 Thesis Contributions	39
1.7 Thesis Outline	39
2 Mobility and Mobility Requirements for User Interfaces	41
2.1 Mobility of People	42
2.2 Mobility of Infrastructure.....	44
2.3 Mobility of Information	45
2.4 Mobility and User Interfaces.....	47
2.5 Mobility Requirements for User Interfaces.....	48
2.6 Summary	49
3 Prior Mobile Systems and Their User Interfaces	51
3.1 Devices Made for Mobility	52
3.1.1 Wearable Computers	52
3.1.2 Mobile Devices	55
3.1.3 Ubiquitous computers	56
3.2 Applications for mobile augmented reality systems	58
3.2.1 Support and Maintenance.....	58
3.2.2 Manufacturing	59

3.2.3 Games and Entertainment	60
3.2.4 Sport and recreation	61
3.2.5 Medicine	61
3.2.6 Tourism	62
3.2.7 Architecture.....	63
3.2.8 Collaboration.....	63
3.2.9 Military	64
3.2.10 Business and brokerage.....	65
3.3 Challenges and Test-beds.....	65
4 Experiments and Design Problems	67
4.1 Dense Displays.....	67
4.1.1 Tabular Interfaces for World Wide Web.....	68
4.1.2 Tabular Interfaces for Mobile Phones	70
4.2 Extended Displays	71
4.2.1 Augmented View.....	71
4.2.2 MARISIL and HandSmart Introduction.....	73
4.2.3 MARISIL Scenario and Movie Creation.....	74
4.3 Summary.....	76
5 Interaction Specifications for MARISIL.....	78
5.1 Foreword to Input Gestures.....	80
5.2 Core Input Gestures	81
5.3 Basic Input Gestures	84
5.4 Advanced Input Gestures	87
5.5 Summary.....	89
6 Evaluating the MARISIL	90
6.1 Features Analysis	90
6.1.1 Example 1: Making a Phone Call.....	92
6.1.2 Example 2: Taking pictures.....	94
6.1.3 Example 3: Recording a video	96
6.1.4 Example 4: Writing a text	97
6.2 Summary.....	100
7 Evaluating the New Applications	101
7.1 Mobile Devices	101
7.1.1 Mobile Phones.....	102
7.1.2 Personal Data Assistants	103
7.2 User Interface Appliance	104
7.3 Mixed Reality.....	107
7.4 Virtual conference	109
7.5 Virtual Prototyping.....	111
7.6 Trusted Interface	111
7.7 Tangible User Interfaces	112
7.8 Mobile Engineering	114
7.9 Summary.....	115
8 Building of the HandSmart Prototypes	116
8.1 Hardware Architecture.....	116
8.1.1 Displays.....	117

8.1.2 Video Cameras	120
8.1.3 Video Cards.....	121
8.1.4 Network.....	122
8.1.5 Calibration Sensors	123
8.1.6 HandSmart Prototypes	124
8.2 Software Available.....	125
8.2.1 jLibDC1394 Library	126
8.2.2 jLibDC1394 Architecture.....	128
8.2.3 Finger Tracking and Hand Recognition Process	130
8.2.4 MARISIL Module.....	134
8.3 Summary.....	135
9 Evaluation of the HandSmart Implementation and some Implications.....	136
9.1 Hardware.....	136
9.2 Software	138
9.3 Implications	139
9.3.1 Security and privacy.....	139
9.3.2 Ergonomics	140
9.3.3 Social Aspects	141
9.3.4 Mobile Services and Applications.....	141
9.4 Perspectives and Conclusion.....	142
10 Conclusion	143
11 Future Work	147
References	
Appendix	

1 Introduction

Past decades in mobile communication systems have shown impressive evolutionary changes in terms of technology, but utilization has remained immutable: from the first-generation (1G) in the 1970s until the later second-generation (2G) in the 90s, the main purpose had been to serve classical voice communication. Mobile devices evolved from dumb mobile phone terminals which delivered only voice services into more complicated Internet text terminals (Fig. 1) and advanced media-phones (including pictures, video broadcasting/viewing). These devices have become small enough to be hand-held. They were usually provided with a keyboard for user input and a display for showing the results of the inputs. However, users do not always want to carry mobile devices in their hands, nor is it always easy to input data into them.

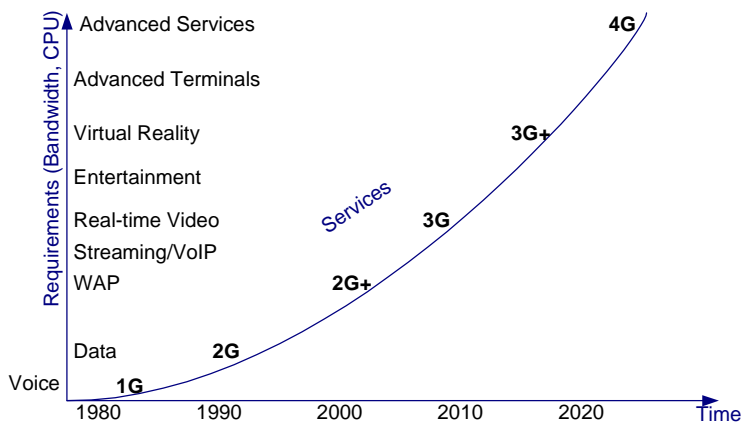


Fig. 1. Evolution of mobile terminal device applications and their requirements (an adaptation from (Latva-Aho 2002)).

In the future, because of technological advances, there will be a migration of applications from desktop computers to mobile devices. As seen in Fig. 1, more advanced services

and applications, as in Virtual Reality (VR) or entertainment, will penetrate further into the mobile sector. However, the penetration of this visual media will depend greatly on the availability and refinement of the interface. Even if current advances in display technology (e.g., colour display, pixel resolution as high as 640x320) fulfil the need of General Packet Radio Service (GPRS) terminals, imaging and video streaming are already pushing for more display and interaction. The limiting factor is not the technology as much as the physical size of current displays. After all, there is a logical contradiction in the statement of greater portability with a larger display.

As a result, the present work contributes to recent developments in the field of mobile devices, specifically for personal mobile ones. The major contribution has been in finding a novel user interface (UI) for this type of device that could extend the physical boundaries of the display and enhance the UI functionality. Special emphasis has been on forthcoming applications and services from mobile devices, as well as on how to make them available for new UI types.

The main emphasis of the research was to make human activities more personal, free in space and time and, therefore, the focus was on mobile personal communication and its applications. This particular research area does not, any longer, only explore the network and wireless communication field but also the devices used and their applications. Moreover, the research on wireless communication devices highlights the usability, availability and ergonomics and therefore, this thesis emphasizes the new designs of UIs for such devices.

The argumentation for this work anchors on appraising that, in forthcoming decades, mobile communication development should bring more bandwidth to devices that would also be more powerful. New devices would require new applications and services to exploit increased capacities and advances in infrastructure and terminals. Future personal devices that would utilize such infrastructures would benefit from new designs of UIs. The designs should confront the problems encountered today in UIs for mobile devices: small screen size, short battery life, poor interaction and low processing power.

It can be forecast that battery life will increase – as more research is conducted in the area (e.g., polysilicon transistors and organic displays that should decrease the requirement of energy). The processing power of small devices should also increase, since technological pace (Moore's law¹) has supplied chip architects with better technologies.

What remains unsolved and has become the subject of this dissertation, is the poor interaction between the user and the mobile device and the support for larger displays. By improving the human-mobile device interaction, usability of the devices could be increased and, therefore, future mobile users could be provided with an instrument to enhance their personal life with more mobility for the applications and services they may want to access.

¹ An observation was made in 1965 by Gordon Moore, co-founder of Intel, that the number of transistors per square inch on integrated circuits had doubled every year since the integrated circuit was invented. Moore predicted that this trend would continue for the foreseeable future. In subsequent years, the pace slowed down slightly, but data density has doubled approximately every 18 months, and this is the current definition of Moore's Law, which Moore has blessed. Most experts, including Moore, expect Moore's Law to hold for at least another two decades.

Problems in using a physical display could probably be solved if, instead of physical, the display was “virtual”. This could be achieved by using the so-called “Near-The-Eye” (NTE) display, also known as a head-mounted display (HMD). Using this technique, the user’s eyes would perceive a small image displayed near the eyes as a bigger image at a greater distance. Unfortunately, when using a NTE display or HMD, the user’s view of the display could float and move as the head or body moves. This could cause a problem for some who have reported dizziness or discomfort while using such displays. In order to compensate this disadvantage, some image processing technique could be used to “anchor” the image of the display to a physical object which would provide the user with a more stable view over the interface.

Making the display virtual could bring more flexibility over the interface as the space to display information could increase. As HMD devices are present in fields like VR and Augmented Reality (AR), the applications available from these fields could improve the functionality of mobile devices. This could contribute to the goal of this work, specifically to expand the applications and services available for mobile personal communication devices.

In order to have a better understanding of the topic, major research areas were surveyed during the work as well as during the writing of the dissertation. The fields closest to the main research area were identified as: *Virtual Reality*, *Augmented Reality*, *Mobile Computers and Devices*, *User Interface Appliances*, *Human Computer Interaction*, *Sign Language*, *Gesture-Recognition*, *Computer Vision/Image Processing*, *Ubiquitous Computing* and *Virtual Prototyping*. A special thoroughness was given to *Augmented Reality* and *Computer Vision*, since they were subjects that were not covered at the start of the work. Later in the research, the attention shifted towards *User Interface Appliances* as a more advanced topic for research.

1.1 Research Question

Mobile phones, historically speaking, were researched primarily by the military and militarised services (and public services such as the police and fire-fighters) as strategic communication devices. Subsequently, their adoption was for the needs of the travelling business community. In the 1980s, the penetration of mobile phones reached mass consumers. Today, they are with almost everybody. This mass adoption introduced new usability requirements on their interfaces (Nielsen 1993).

In the middle of 1990s, mobile phones were joined by many other portable devices, (Personal Digital Assistants (PDAs), Mobile Computers, Communicators), which were more advanced communication devices. These “mobile computing platforms” were able to include other applications (fax, web browsing, video, calendar, to-do’s and other organizer specific tasks) but they also became increasingly more complicated to use. Additionally, their small displays and keyboards posed a great challenge for future designs of their interface (Väänänen-Vainio-Mattila & Ruuska 2000). The research question was, therefore, *whether making the display virtual would provide the next*

generation of mobile information appliances (IAs) with a better display size and better interaction, hence allowing the design and development of new applications and services.

This research was not strictly limited only at finding the answer to just the above question, but attempted to also answer three questions which are encompassed by it, namely:

1. *When would the technology be able to provide such devices with answers to handle the proposed specifications?*
2. *Are there prevalent applications available with respect to the new approach?*
3. *Are there new services or applications that could take advantage of such devices?*

1.2 Research Approach

This work was initiated knowing that the field of mobile computing, in later developments, would require new interaction paradigms (Bertelsen & Nielsen 2000, Klinker *et al.* 2000). If the answers to the research questions proved to be positive, it would mean that by virtualizing the display, the new method could lead to the next generation of mobile devices with a UI that would not fall within the physical restrictions of the present ones. Moreover, having a UI that is fully virtualized, could also enhance developers and UI designers with more options regarding implementation and human-computer interaction.

The implications of this research are large: virtualizing the display could allow the next generation of mobile devices to break the physical barriers of designing the UI and it could create a new dynamic for their usage. Virtualised mobile devices could add more scalability and flexibility in the design of the UI (Rodden *et al.* 1998). This could provide the manufacturers with a platform capable, not only of handling the older basic applications that are available today in mobile phones, but also a new and more advanced class of applications that could extend their targeted markets to the larger markets of mobile information appliances¹ (Norman 1998). The devices could also come with a higher level of personalisation and customisation. The users could then be able to enjoy accessing a device that is easy to use, tangible and pervasive, and something that blends into the surrounding environment without being obtrusive.

Another benefit from this research, if the answer to the research question is affirmative, could be the increase in the number of services and applications that could be deployable for this class of devices. Mass production of such a system could bring the price down to the level of today's communication devices (such as the Nokia

¹ Information appliance (IA) refers to any device that is capable of processing information, signals, graphics, animation, video and audio; and the exchange of such information with another IA device. Typical devices in this category are smartphones, smartcards, PDAs, portable PCs, and so on. Digital cameras, the older cellular phones (only voice), set-top boxes, and TVs are not IAs unless they have the capability of communication and information processing functions. Information appliances may overlap in definition and are sometimes referred to as mobile devices, smart devices, wireless devices, wearable devices, Internet appliances, web appliances, handhelds, handheld devices or smart handheld devices.

Communicator) while their role could expand from being a communicating device to the more advanced class of an information device. This, in combination with a more advanced network infrastructure, could create a new way of providing information services to the masses, as the Internet is now doing. In addition, since the present desktop computer paradigm, by its nature, is fixed and non-contextual, mobile devices could be bound to a large contextual variation (Johnson 1998). Based on the nature of the context, a new set of applications could be created along the line of applications common to desktop computers.

The research was started with a focus to provide an answer to the question of extending the display size of current mobile devices through ways of making the device virtual (“virtualize”). Attempts to virtualize the input or output devices had been done earlier, starting with virtualizing the keyboard (Kalawsky 1993a). However, they were mostly concerned with the virtualisation as the outcome and not with mobility or enhancement of the device that was being made virtual. Later, the focus shifted towards providing future mobile devices with a larger range of applications, so the ultimate goal was to increase the applicability and usability of mobile devices. As a result, the work was split between the research for providing the feasibility of the devices and that for supporting the interaction mechanisms for the user.

In order to demonstrate that virtualizing of the display would provide better usability and increase the range of applications, several prototypes were analyzed and built. Additional requirements were extracted during each step of the development stage of the prototypes. By stressing the utility of the new artefact, the research led to the identification of a novel device. Because of the innovative aspect of the work and the artefact building, the research method used was an innovative building approach (Järvinen 2001).

The creation process of the artefact was done in iterative steps due to the novel aspect of the research and the difficulty of forecasting or imagining the target state of the final artefact. Each step or version of the artefact provided new insights and new requirements for the subsequent building processes. The specification and implementation processes that led to a new version are presented in Fig. 2. The prototypes varied through various devices implementations to future specifications extracted from scenario analyses and the movie production. Each stage in the implementation contributed a new set of requirements that were translated into new specifications that were later part of a newer production process.

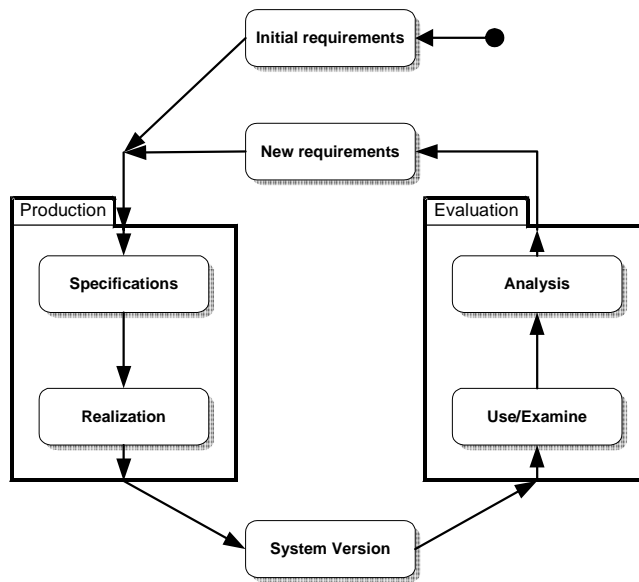


Fig. 2. Iterative requirement gathering and incremental implementation process for constructing an artefact.

The evaluation of new versions was done based on the comparison between the old and new target, but keeping in mind the usability and utility attributes of the newly created artefact. The scientific contribution came from the novelty of the system and through the argumentation that the usability and utility were improved (March & Smith 1995). The evaluation criteria were a set of requirements extracted from the work with the artefacts.

The focus of the evaluation was on the feasibility rather than efficiency, since, like most artefacts of this nature, efficiency comes with use (i.e., the keyboard was and still is slower to use than hand writing for some people). As a result, the evaluation of the efficiency of the system was left for further research. This was also because the suggested system was not yet deployable outside of the laboratory and hence, it could not be tested for efficiency by the public in general daily usage circumstances. The prototype designed to test the feasibility was far from being portable due to some manufacturing limitations of the off-the-shelf components used (i.e., batteries, cables, port connectors, etc.). In the future, such evaluation should be more feasible, once the implementation and mass-production of such classes of systems are closer to realisation.

Instead, a comparison between current systems and the proposed one and how this system was capable of removing some of the restrictions imposed by the current design of UIs for mobile phones was available through each version of the system. The study is a requirements analysis of the tasks that a user could perform using the proposed version of the system.

1.3 Survey of Earlier Work

An important field of study that is closely related to the subject of this thesis is VR. The hypothesis is that making the display of the device virtual should enable future mobile devices with larger screens leading to VR. Re-adapting Ivan Sutherland's definition of VR (Sutherland 1965), virtualizing the display of the mobile device means presenting the display to the user's senses such that it is hard to distinguish between the real display and a virtual one. Such a definition implies that the user's eyes should see the artificial display of the device at a higher, or at least equal, resolution as his or her eyes are capable of distinguishing. In reality this task is currently impossible.

A set of devices has been contributing to the process of immersion that could lead a user into experiencing the virtual world. These devices include displays, haptic interfaces, audio devices and scent generators. Even if the technology advances far enough to provide a sense of immersion, it would be up to the user to accept the imperfections of the virtual world and perceive it as real. This redefined VR is known as Immersion-Interaction-Imagination or I³ (Burdea & Coiffet 1994).

To immerse the user view into a three dimensional (3D) world implies that both eyes should be exposed to stereo images. Stereoscopic images can be tracked back in history to Euclid or Leonardo da Vinci (Sexton & Surman 1999), the latter realizing that in order to capture the reality in a painting, the artist should consider producing two paintings (one for each eye). The same idea works for audio as in stereo or surround audio systems.

Various devices can provide a user with stereovision. Head-mounted displays are by far the most popular. Another interesting approach is CAVE (Cruz-Neira *et al.* 1993). In CAVE, the user can experience the 3D world through a pair of glasses that are synchronized with projections from the walls. Because of the wall projections, the user's eyes are focused on the walls and not on the close range HMD. Another benefit of the CAVE environment is related to the motion of the user and the head. The user can move in CAVE and enjoy the view without having the disturbing lag caused by the delays in presenting the proper view corresponding to the changes in the direction of the glance of the user. The big disadvantage in adopting the CAVE system compared to the one proposed here is that it is, in the main, not mobile.

The more popular approach to stereovision is to use a HMD. Most surveys describe Ivan Sutherland's system (Sutherland 1968) as the first to have produced the HMD. In reality, Comeau and Bryan were the first to describe a system developed by Philco Corporation for surveillance that used a HMD (Kalawsky 1993b). Unfortunately, when using a HMD, the user is able to see the 3D world, but when moving or the head is turned (parallax motion), the image will remain the same, causing nausea and dizziness. To correct this problem, special tracking sensors need to be used on the HMD that would compensate for the head movement. Unfortunately, a set of sensors is required to be available in the room to provide the tracking data, making the mobility of the system almost impossible. Some researchers managed to relax the implementation by using special fiducial markers that were easy to install (Naimark & Foxlin 2002). A compromised solution would be to install the tracking sensors on the body of the user and access only the relative coordinates of the head to the body (Foxlin & Harrington 2000). Even by using accelerometers and other visual approaches (Simon *et al.* 2000), the user

of VR systems is not capable of seeing anything but the virtual world and hence cannot interact and carry out activities other than those available in that world.

A more relaxed definition of VR, where immersion is not a requirement, includes systems in which the artificial images are viewed on the display of a monitor (as in Desktop VR) or on a semitransparent mirror (as a virtual push-button keyboard (Kalawsky 1993a)). With the introduction of the term Mixed Reality (MR) and the Milgram's taxonomy of virtual displays (Milgram & Kishino 1994), these systems could be better categorized as AR systems.

To augment means to add or to amplify (to grow) (Company 1994). Based on this, AR means that it is a way to enhance or amplify a user's sense of the real world with synthetic information. The most common augmentations are the ones of the view of the user, but other augmentations could be of the audio (i.e., noise reduction systems in airplanes). Augmented reality has been included as part of VR (or a subfield of VR). A better categorization is given by Milgram and Kishino (Milgram & Kishino 1994). They defined it as part of MR, a more proper term to describe the blending of the two worlds, virtual and real. Azuma used their taxonomy and representation of the "virtuality continuum" (Fig. 3) in his survey of AR (Azuma 1997). In the figure, the virtual and the real world stand at the two edges. The VR operates in the Virtual Environment (VE). Between the two worlds is the MR, where the virtual and real are blended. Augmented reality is closest to the real world, having the virtual objects overlaid on the real environment, while Augmented Virtuality has the real objects present in the VE.

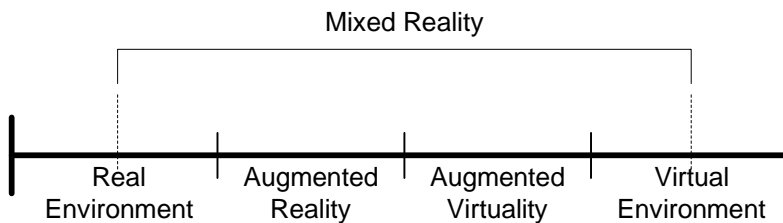


Fig. 3. Milgram's simplified representation of a "virtuality continuum" in which AR is part of the MR and adapted by Azuma (Azuma *et al.* 2001).

By having the surrounding environment as real, AR is much better suited to the requirements of mobility for a system. The user is capable of operating with virtual objects while seeing the real environment around. This characteristic alone suggests that the user is capable of interacting with the virtual world within the real one. This characteristic is only present within AR, where artificial images superimpose the real world.

The field of applications for AR is very wide. To mention just few, it is useful in medicine, manufacturing, engineering, entertainment and military (Azuma 1997). The synthetically generated information can be manipulated easily and dynamically. This should allow the design of the UI to be less constrained by the real and solid constructs. A UI constructed with computer generated objects to fit a user's need should be more flexible and adjustable than a physical one. This means that interaction could become more intuitive and natural in the MR environment. Moreover, the normal UI

characteristics, otherwise constrained by its material nature, should develop more freely, aided by the new virtual nature provided within an AR environment. One example concerning the size can be found when looking at the display limitations of a classical desktop computer or that of a mobile device, the limitations caused by the incapability to increase the size without replacing/upgrading the device. These limitations are not relevant in the virtual world as the only limitation is the user's field of view (FOV) or the resolution of the NTE displays and HMDs¹.

A user of AR could interact using a special set-up in which a video camera grabs the surrounding of the user. Such a system would be capable of tracking movements and interpreting them later as inputs for the system. An improvement to the approach would be to make it mobile. The literature describes this as being a "Mobile Augmented Reality System" (Julier *et al.* 2000, Höllerer *et al.* 2001) or MARS. This thesis describes the operation of such a mobile platform in which a video camera is also available to allow the image processing to take place. Such a system consists of the following parts:

1. *An NTE display*

Visual information is the most important aspect of the MARS. The augmented information would be added to the sensory data perceived by a user's eyes. The NTE display should be as light as possible in order to be ergonomically worn.

2. *Wireless system*

The wireless network should, if available, allow the system to access various resources including computation of data remotely. In some references (Julier *et al.* 2000) the MARS could also include a Global Positioning System (GPS).

3. *Mobile computing system*

An adequate Central Processing Unit (CPU) with a high-powered graphics chip would be required in order to achieve sufficient computing capabilities to handle the applications for such an appliance. Usually, the main computational power used would be to update the see-through head-worn display screen and, therefore, the need for a fast and low power consumption graphic chip would be essential. The unit should also be as light and small as possible.

4. *Power unit*

Sometimes referred to as the battery pack, it is indispensable for a mobile system and has to be light and have enough power to enable the user of a MARS with enough autonomy to work.

The MARS architecture, including the digital video camera for the image processing software, is presented in Fig. 4.

¹ The Head Mounted/worn Display (HMD) is regarded here as subclass of NTE displays. The NTE displays could be monocular or binocular and composed from a display(s) that is situated closer to user's eye(s). Compared to an HMD, the NTE display could be a microdisplay that is embedded in the eye-glasses (Fig. 8 on page 31). The definition of NTE displays does not include the class of displays that are kept close to the eye by the user's hands but only displays that are worn on the head (i.e., eye glasses).

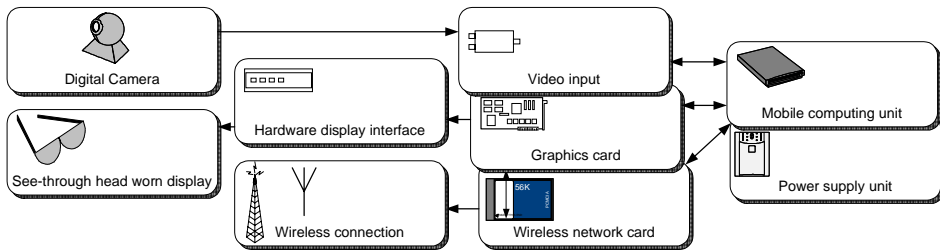


Fig. 4. MARS with image processing capabilities.

The interaction of an AR based UI is less abstract than a classical UI or Graphic User Interface (GUI). Real objects can turn, for example, into input and output devices just as classical computer interfaces do. Moreover, when overlaid with a virtual image, normal objects could become a source of interaction and used in such a real-synthetic environment. The objects would be real, but they could change the state of the virtual ones – therefore, they could become a source of intuitive interaction. This type of interface (referred to in the literature as the Tangible Augmented Reality User Interface (Kato *et al.* 2000)) would be able to operate physical objects as synthetic ones. An example would be when a user plays with a real object (a plate/paddle) and picks up, moves, drops or removes the virtual objects (tables) in a Computer Aided Design (CAD) application. Another example would be when AR is used as an interaction tool for virtual keyboard applications (Taylor 1999) or for virtual prototype designs (Halttunen & Tuikka 2000, Tuikka & Kuutti 2001).

The survey revealed the unavailability of the connection between AR based UIs and the MARS. In the past, there has been substantial research effort involved in collaborative AR (Butz *et al.* 1999, Höllerer *et al.* 2001), AR desktop applications (Dempski 1999, Kato *et al.* 2000, Sato *et al.* 2000), entertainment (Dodsworth 1998) and personal navigation filtering (Julier *et al.* 2000). However, very few (Sasaki *et al.* 2000) have tackled the problem of UIs for the MARSs. The most advanced approach was developed in Nara, Japan, at the Nara Institute of Science and Technology (NAIST). Even in Hiroshi Sasaki's approach, the interaction was limited to a small set of menus (five, corresponding to each finger on a hand), restricting, therefore, the relevance of the UI to a small set of applications.

As it is less obtrusive and allows the user to interact with a virtual interface, AR could become the natural choice for implementing a virtual display for mobile devices. The following sections look at the past work available on this field. A history of AR, as well as the description of the I/O devices and some discussions, completes the picture about the subject and how it could contribute to the realization of the research problems.

1.3.1 Augmented Reality History

Historically, traces of AR systems can be found dating back to the year 1897, when the psychologist George Malcolm Stratton, in his report (Stratton 1896), referred to “upside-down” glasses. These glasses, more like goggles, consisted of two lenses with equal focal length distanced from the eyes at two focal lengths. Using the goggles, the eyes were able to see the rays of light from the top at the bottom and vice-versa. Stratton reported that after wearing the goggles for a period of several days, he was able to adapt and function quite normally with them on.

Even if Stratton’s approach can be considered as primitive and there was no augmentation of the real world, his system has some resemblance with Ivan Sutherland’s system (Sutherland 1968) of the late 1960s. Sutherland and his colleagues designed a HMD with which a user could see computer-generated images mixed with real objects (Fig. 5).

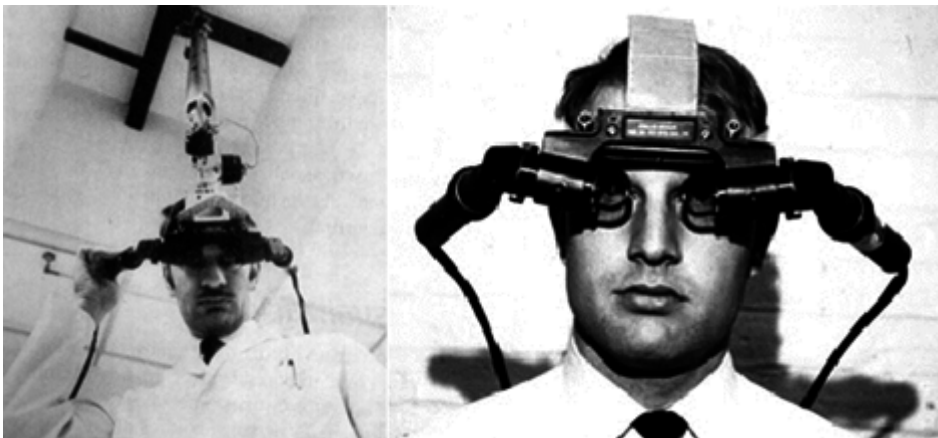


Fig. 5. Ivan Sutherland and his pioneering research on HMDs (picture available from Kalawski (Kalawsky 1993c) and Internet¹).

Later, the work continued and the army was able to create an application that enhanced the vision of a pilot with more information. This new potential of applications increased the development and interest in the field. The US Air Force Super Cockpit project developed displays for pilots to see more information overlaid on the helmet (Kalawsky 1993c). Their work and the findings of the project were made available publicly, and this became the core of several civil research projects.

However, the term “Augmented Reality” was first cited in the summer of 1990 by a team of researchers from Boeing. The team members were Thomas Caudell (co-editor of the book “Fundamentals of Wearable Computers and Augmented (sic) Reality”) and a colleague, David Mizell. They had the task to find an alternative to the expensive diagrams and marking devices used to guide workers on the factory floor (i.e., Boeing factory, the cabling department). The outcome was a head-mounted apparatus designed to

¹ <http://mitpress2.mit.edu/e-journals/Leonardo/isast/spec.projects/osthoff/osthoff1.html>

display instructions about the cabling of the airplane through a high-tech “eyeware”, and then project them on some panels so the workers could follow them. The apparatus was supposed to be an alternative to the manual reconfiguring of each plywood board/markings device. While using it, the user was able to see the instructions virtually. This permitted the system to alter quickly and efficiently, through a computer system, the instructions for the worker (or the layout). In 1992, they published their results in a conference paper (Caudell & Mizell 1992) linking their name to the name of “Augmented Reality”. Thomas Caudell left Boeing and pursued a career to research aspects closer to AR.

In recent years, AR had received much attention from scientists and hence, progressed considerably. Another corner stone in disseminating the idea was during the 1997 when Dr. Ronald Azuma published the “Survey of Augmented Reality” (Azuma 1997), which was the basis of the research and work described in this thesis. In 1998, the first International Workshop on Augmented Reality (IWAR '98) was sponsored by IEEE.

The research in AR has multidisciplinary implications. Hence, the applications are also widely dispersed in various fields. For example, AR enhances wearable computers with better interfaces and enables ubiquitous computing to be more omnipresent.

The research in wearable computers contributed to a better understanding of AR and applications. Starner and Mann worked together and wrote the first paper linking the two fields (Starner *et al.* 1995). They described various applications for wearable computers in which enhancements of AR methods were present.

Commercially, handheld devices have been available on the market since the 90's. Supporting the idea of a miniature or palmtop computer, the Psion I is already history while the new design and light UI boosted the popularity of the Palm Pilot devices (1996).

One important field in which AR has contributed substantially is ubiquitous computing. Ubiquitous means being available everywhere at the same time. Researchers define such devices as being everywhere and enhancing humans with a better understanding of their surroundings. The ubiquitous computer was defined for the first time by Mark Weiser (Weiser 1993) as a computer that is present everywhere. The subject of ubiquitous computers has evolved since then and several researchers have been looking into applications. Some researchers found AR as an approach to define future UIs for this field (Newman & Clark 1999). The primary objective, as stated by Mark Weiser, was to provide hundreds of wireless computing devices per person per office, of all scales (from a 1 cm display to a wall-sized one). This requires more work and development in many areas like operating systems, UIs, networks, wireless, display devices, and many other areas.

The technologies to support AR systems have also developed and provided a wide range of laboratory products as well as commercial ones. Historic breakthroughs in I/O devices, calibration, tracking (registration) and occlusion have paved the way to better implementation and possible future commercialisation of AR based systems. The following sections describe the current state of the art in the respective research areas.

1.3.1.1. Input Devices

Augmented reality systems, like any other IA¹, require a method to input the data. Considering the nature of AR systems, there can be two classes of input devices: classical and advanced. The classical devices would come from previously designed systems operating in 2D (like desktop computers) while the advanced devices design would apply only to MR systems or 3D oriented systems.

Table 1 presents the currently available types of inputs used by AR systems. Many devices that are included in the table are still under development and, hence, they are not yet commercially available. The classical UIs are those that have been accepted by users and are considered to be a standard UI in current systems. In the same category are the interfaces that use the basic principles of the older ones (like virtual keyboards). The advanced ones are the UIs based on a different architecture and those that use other types of input devices that are not inherited from the classical ones.

Another classification could be to split the devices into pointing devices (continuous) and typing devices (discrete). Furthermore, some input devices have a physical presence (keyboards, mice, tablets, haptic interfaces) while others are only available virtually (gesture based, sign language, nervous interface, brain wave/thought readers).

A novel approach to classify input devices is to split them into fixed and mobile. In this case, the context in which they are used is the classifier. For example, for a fixed input device, the size and context are not as important as the speed, while for a mobile device, the physical context (i.e., location) becomes very important. The user should be able to operate a mobile device while standing or walking.

¹ An Appliance is a device or an instrument specially designed to perform a specific function, like an electrical device, such as a microwave machine, for household use.

Table 1. Input device for AR systems.

Type	Description
Classical:	
Buttons	Mini-keypads (Twiddler)
	N-Fingers keypad
	Buttons from mice, trackballs, etc.
Keyboard	Smaller keyboard
	Chording keyboard
	Foldable keyboard
	Rollable keyboard
	Soft keyboard (or onscreen keyboard)
	Virtual keyboard
2D Pointing	Mice and trackballs
	Tablet and pen
Speech	Speech recognition
	Commands
Advanced:	
2D Pointing	Hand/gloves orientation/position
	Head orientation
3D Pointing	6DOF sensors
	3DOF mouse
Body sensor	Gloves
	Body sensors
	Brainwave sensors
Tangible	Video-based recognition (including sensor based) tactile detection
Interfaces	
Video	Video-based recognition based sign language
	Video-based recognition hand gestures

Despite the large amount of devices used for input, there is still much to tackle on how to use them properly while operating a mobile system and how to implement the UI so that they are less intrusive and complex. For example, using gloves is regarded by many as cumbersome, even though they are very much used for laboratory and usability testing. This is because, for a mobile user in normal circumstances, it would be hard to accept and wear them in order to interact with a mobile system (it would take time to put them on, while wearing them all the time may not be comfortable).

1.3.1.2. Output Devices

The output devices for AR could operate on various media like audio, tactile, visual or physical. While the present work is concerned only with the visual medium, it does not exclude the idea that audio could supplement it. Moreover, latest research on haptic interfaces could prove useful in applications such as, for instance, virtual prototyping

(Halttunen & Tuikka 2000). Unfortunately, the design of such devices is far from being available in a portable format and the applications to benefit from it are not yet fully feasible.

The audio medium has to deal with various outputs in order to guide the user or augment the environment. The sound could be available through earphones, spatial or a 3D mechanism and as ambient sound. There could be various applications in which the sound could supply more information than the visual and that, in certain cases, it could reduce the clutter of a scene (Barrilleaux 2001).

Visual output devices for an AR system, based on their capabilities and implementation, could be HMDs (Kalawsky 1993b) or of the Surround/CAVE type (Cruz-Neira *et al.* 1993). When referring to mobile applications, the class is restricted only to displays that mount on a user's head and which provides the means to show computer-generated images.

Another type of display that could be mobile when using a special setting is the projective one. This consists of some beam splitters, micro-mirrors and projective lenses. Operating these devices would require a special reflective material to be placed into the environment or over the objects. These settings would allow the user to see the projection reflected on the special material. Such an AR experience is personal since the projected stream comes from the user's own system and, therefore, only the user would perceive it.

In the recent years, it has been observed that commercial NTE displays have become more commonly available on the electronics and home appliances markets and at lower prices. These devices, that are now advertised as a video portable TV or a mobile DVD viewer, could easily become part of an AR system.

A NTE display based on this implementation could be either an optical see-through or video see-through one.

Video see-through has been studied since the beginning of AR research. In 1968, Ivan Sutherland (Sutherland 1968) introduced the idea of such a system based on HMD in the literature, and the video camera as is described in Fig. 6. When the system was used, the viewer was able to see the real world and, simultaneously, the virtual objects, as if they were present in the real world. Unfortunately, such systems, even if they were easier to build, are accompanied with several problems. One is that when using such a system, the user would see the world from the point of view of the video camera, which, in most cases, is located on top of the displaying device. That induces the viewer to have an illusion of being taller. To correct this, a more modern approach uses the parallax camera – or places the camera in front of the eyes. Moreover, the system has to be very fast to display the video of the real world with the virtual objects overlaid. If the system fails to align the user's head direction change with the corresponding video image, this could result in motion sickness. Such an approach also lacks the natural FOV to which a person is accustomed (in some cases users have reported a slowness in orientation). The best HMDs are not able to provide more than 40° FOV, which is less than a person can see. Even if the system is improved, the user of a video see-through system will still be immersed into an environment that is fully digitised and would not have access to the environment except through a video image. As a result, the user of such system will face the effects of cybersickness (LaViola 2000). A much better solution is available when using optical see-through displays.

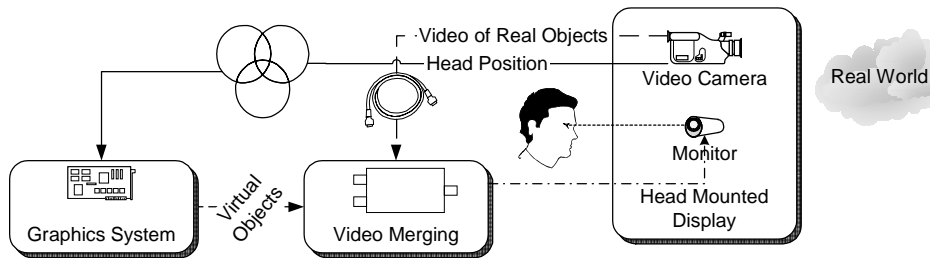


Fig. 6. Schematics of video see-through system.

Optical see-through has seen much development lately, with the introduction of commercial devices from Sony and Olympus¹. The see-through, many agree, is a better solution for AR applications, since it does not pose the problems as the positioning of the camera and the cybersickness caused by immersion associated with the video see-through system. However, a certain level of disturbance could come from the system if the virtual objects fail to properly align with the real ones. The system, as described in Fig. 7, consists of a tracking device (for the pose of the head) and a specially designed display (or projective display). The usual way to build such a system is to use a semitransparent mirror that would mix the real world image with the one generated by the computer. Unfortunately, this classical approach has some drawbacks in terms of contrast and the transparency of the glasses that makes it impossible to see through in dark places. Very recently, a possible solution was provided by the use of laser techniques and the Holographic Optical Element (HOE) that are bringing greater contrast, higher brightness, lower power consumption and a larger depth of view (DOV) (Kasai *et al.* 2000a, Kasai *et al.* 2000b).

Another implementation problem arises when using this system because of tracking (registration) problems. Compared with the video see-through approach, the optical one, in order to properly overlay the virtual objects over the real ones, requires a special calibration process between the NTE display and the eyes of the user (Genc *et al.* 2000). This calibration process, in some implementations, needs a permanent setting so the system can continuously identify the position of user's eyes in order to produce the proper overlaying of information. In some cases, the user can set the correct position of the glasses so that the objects are properly overlaid on the virtual images.

¹ <http://www.stereo3d.com/hmd.htm> for a detailed comparison of prices, resolution, year of release, for these devices.

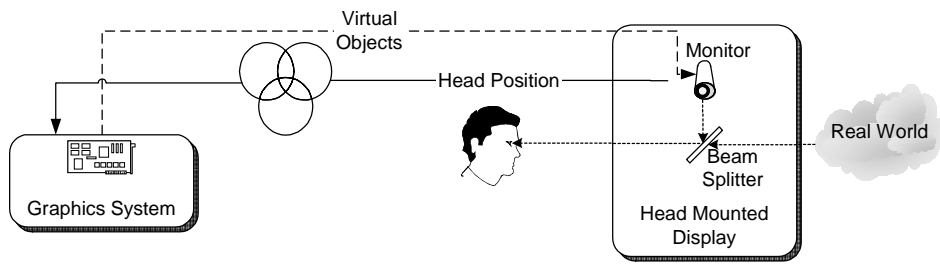


Fig. 7. Schematics for optical see-through system.

There are some arguments about AR being obtrusive to use. There is criticism that while using the see-through glasses, the user of an AR system is unable to have direct eye contact with other users (page 54 from (Rakkolainen 2002)). This argument holds good only if the users are unable to easily remove the AR glasses. Fortunately, current technologies can deliver greater transparency and flexibility to display devices (since the introduction of the micro-displays glasses (Fig. 8)) versus the large and cumbersome HMD's of the 90's. Moreover, it is a common observation that during face-to-face discussions, if eye contact is important between participants wearing glasses, some will lower their glasses or remove them (this is regarded as a display of transparency and fair play). The same pattern could be applied to AR systems, where the AR glasses could be lowered or removed, permitting the users to have a direct eye contact whenever necessary.



Fig. 8. Thad Starner's MicroOptical displays¹ are less intrusive than the LCD implementation.

Despite the advances in optics, there are still several barriers to cross until a good level of detail (LOD) and ergonomics will lead to the specifications of a commercially viable

¹ <http://www.cc.gatech.edu/~thad>

product. Recent work is concentrated on the convergence of the eye and how it could cause fatigue. Some implementations are able to cover an area from 0.25 meters to infinity in just 0.3 seconds which would cover most of the requirements of an ergonomic system for the eye (Sugihara *et al.* 1999).

1.3.1.3. Calibration, Tracking and Occlusion

Calibration and tracking are important techniques that contribute to good implementation of an AR system. Calibration is the technique used to align the virtual world with the real world from the user perspective. Tracking, also known as registration, is the method used by an AR system to recognize real objects and register them into the system so that they can be tracked and sometimes overlaid with virtual objects. Both techniques require a combination of sensors and algorithms that would enable the system with an accurate real-time response of the position of the head and the eyes of the user (in the case of optical see-through). This section treats them together since both techniques involve sensors and algorithms that belong to the same family of detecting the user-object movements in space. Similarly, occlusion takes advantage of both techniques in order to achieve correct overlapping and registration of the real and virtual object perceived by a user.

The calibration of an AR system is required only by systems that use the optical see-through display, since this needs the exact position of the eyes of the user to be referenced to the display. For the video see-through systems, this calibration is not required since the systems can interpolate the video of the real world with the virtual images without taking into consideration the user's eyes.

In order to cope with the calibration process of the optical see-through glasses, some researcher have proposed a semi-automatic solution. Their intention was to take advantage of techniques involving gathering of the data in several steps from the inputs of the user (Genc *et al.* 2000) and then provide the settings for system calibration. Unfortunately, this method has to repeat itself whenever a user removes the glasses or if the glasses shift or twist on the user's head.

Another aspect attributed to the calibration process concerns the finding of the camera or sensors parameters used by the AR system. In 1998, Kutulakos and Vallino presented their idea of a calibration-free AR based on a weak perspective projection model (Kutulakos & Vallino 1998). Since then, several others have addressed and developed the idea including traditional illumination techniques (Seo & Hong 2000a). All these techniques are contributing to better integration between the two objects, real and virtual, and hence, to a correct perception of the tracking (registration) information provided by the system to the user.

Tracking is also required for many AR systems. The co-ordinates of objects and the user are important information for the system in order to place the virtual objects. Tracking requires a variety of sensors in order to achieve the accuracy expected of a system. The techniques and the sensors can provide the applications with different levels of detail, depending on the application or the environment in which they operate. Some

systems use a combination of sensors (hybrid systems) to refine the accuracy of the output at a higher speed. Hoff and Azuma have used auto-calibration between a couple of sensors in order to correct the error induced by a compass (Hoff & Azuma 2000). Some researchers have been looking into the use of vision-based registration (Park *et al.* 1999, Stricker & Navab 1999, Seo & Hong 2000b, Jiang & Neumann 2001) instead of using magneto-accelerometer sensors. This latter method has proven to work well in outdoor environments (State *et al.* 1996). In the case of the vision-based technique, the system can register the objects more accurately, but, simultaneously, it requires a higher processing time (Sato *et al.* 2001).

The idea of using the same approach as in computer graphics, called LOD, could also apply to the treatment of errors (called Level of Error or LOE) in the tracking system (MacIntyre & Machado Coelho 2000). If the registered object is far away, the registration could be less accurate, leading to less computation and improved accuracy. In 3D graphics using LOD, the closer the objects, the better the rendering should be.

Tracking techniques that are only available for indoor use would be to place fiducial markers (Cho *et al.* 1998, Naimark & Foxlin 2002) at known locations and to use a special infrared camera that could see the markers even in variable light conditions (Sato *et al.* 2000). Latest research in the area reveals many other methods that are capable of enabling AR systems with accurate data from the trackers in real time. Several systems are now providing indoor navigation with alternative technologies and use the latest developments for tracking to operate accurately (like visual tracking, fiducial, infrared and Bluetooth techniques, and a combination of these).

Tracking concerns not only the position of a user's eyes or head but also of some other body parts (like the hands or fingers (Brown & Thomas 1999)). A very innovative technique applies the sensors on the body of a user and tracks the movements of body parts (Foxlin & Harrington 2000) providing the system with more information about the user and the interactions that occur during the operation of the system.

A relevant feature of tracking, especially when used in the MARS, is the amount of equipment needed for the user to wear or to set it up in the environment. At the ideal level, the goal should be to wear as little as possible and to have as narrow of necessary settings in the environment. Because of this, some tracking systems require a certain level of knowledge about the operating environment (i.e., building layout, street maps or characteristics of the nature around the area).

Occlusion has two meanings in AR. The first is when an object occludes the fiducial marker or other object, causing an error in the registration or recognition process (Lepetit & Berger 2000). The other meaning is when an artificially generated object occludes a real object, like a technique to block the real object's image and to overlay it with virtual ones (Takagi *et al.* 2000). Depending on the requirements of a system, occlusion can become very important in the design process.

Calibration, tracking and occlusion are the greatest problems faced in AR systems. There are several possible solutions including combining some of the techniques. An ideal design of a UI in AR should find the solution which is the fastest and with least processing required for these to work.

1.3.2 Virtual User Interfaces

Since the beginning of the studies on VR, some work has examined the UIs that could benefit from this environment. Virtual Reality is seen as the high-end of human-computer interactions (Burdea 1996b) and it has the potential to target a wide range of applications from CAD to entertainment, medicine and tourism. In 1975, Knowlton proposed the virtual push button (Kalawsky 1993a), a device that used semitransparent mirrors to mix the image of a virtual keyboard with the image of a real keyboard that the user could operate. Other early applications of Virtual Environments (Ves) were in developing flight simulations in which the operator interacted within the environment using a virtual hand (Kalawsky 1993d).

Another application was to take advantage of visual interaction and feedback. The user could introduce commands by using image recognition of the motion of body parts. For example, in Fig. 9 the user is moving the virtual hand (that is synchronized with the real hand through motion recognition and processing of an image from a video camera), and changes the state of the virtual button available on the screen (Fukushima *et al.* 2002).

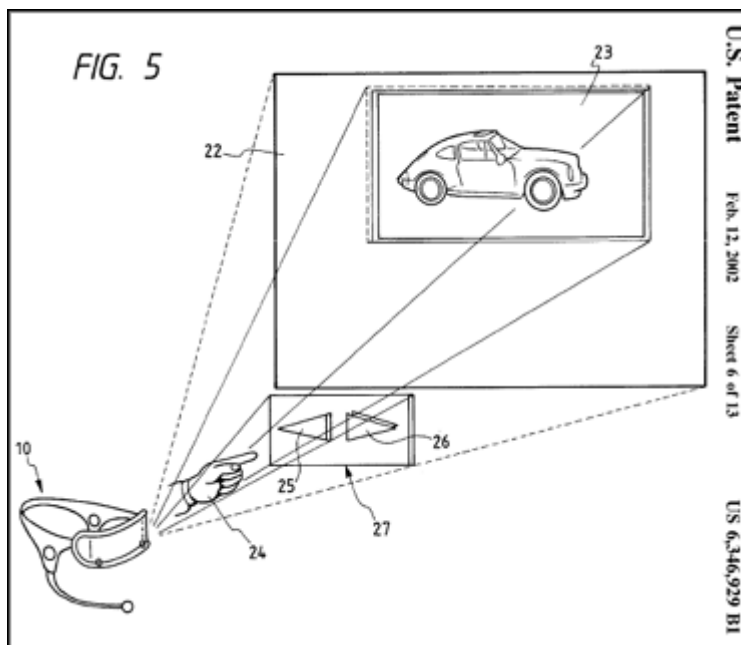


Fig. 9. Fukushima, Muramoto and Sekine display apparatus detecting the user's hand in correspondence to a virtually displayed button (on page 6 of the patent papers (Fukushima *et al.* 2002)).

Moreover, VEs provide many modalities of interaction, like 3D graphics, sound, tangible and even olfaction feedback (Burdea 1996a). This has pushed development of various applications and led to an increased interest from private companies to develop new products (Burdea 1996a).

As technology has progressed, some applications of VR have migrated into the blending of virtual and real, more specifically into MR. As in the previously described patent (Fukushima *et al.* 2002), some work has introduced real objects into the virtual world (in our example, the hand of the user). Other applications have included virtual objects into the real world of the user to enable the interaction (as in AR).

Since the introduction of the term AR (in the early 90's), until several years later, the major subjects of concern were calibration, tracking, occlusion, and devices to support them. Nevertheless, in recent years more research has dealt with the applications and interfaces for AR. Unfortunately, many of these prototype UIs are too often using the metaphors borrowed from desktop computers (Dempski 1999). However, several researchers have looked into alternative design and usage of the UIs for VEs, like conferencing enhancement (Billinghurst *et al.* 1998). Their technique and proposed interaction schemes have provided the information on the capabilities available when using AR.

More advanced interfaces are the ones that integrate physical objects with the virtual ones, like in tangible interfaces (Ishii & Ullmer 1997, Kato *et al.* 2000, Ishii 2002). When using such an approach, the users are interacting with the virtual world by means of manipulating real world objects (like a real paddle to manipulate the virtual furniture models in a prototype interior design application).

1.3.3 Mobile Augmented Reality Interfaces

Researchers have been investigating mobile AR systems and their abilities to support and enhance various other activities like work, leisure and navigation. These systems, which are closely linked to wearable computer systems, have been analysed and discussed in a paper published in 1997 (Starner *et al.* 1997). Since then, changes in area (Starner 2002) have led to a new field of application more closely related and targeted at the field of mobile devices that are available today. Their availability varies from mobile medical kits to wrist mounted wearable computers. These types of devices, some of them available commercially today, underline the importance of mobility and portability of future devices. As pointed out by some researchers (Gemperle *et al.* 1998), while the trend within society is to make tools and products more portable, the portability of the office desktop and personal computer should not consist only of shrinking the desktop into a more smaller and portable version. Instead, the potential and opportunities of the new environment would need more detailed exploration, eventually leading to a better specific human-mobile computer interaction paradigm.

Mobile devices like PDAs, mobile phones, wristwatches, health monitors and many others, all share the same problem of limited screen size. The paradigm is that while manufacturers would like to have more screen size so that a bigger UI could be available, with better functionalities, the device has to be made smaller and lighter. Unfortunately, the bigger is the screen size, the less is the device portable. Strangely enough, even with the current high pace of development and the recent increases in market shares for mobile devices, there is little research on the topic of UIs for small portable devices. This may be

due to the specialization of the interfaces (like the mobile phone) or to the lack of openness of the manufacturer in sharing the information with the public or the academia.

An advanced system was developed using AR as a platform for interaction. The system, called HIT-Wear (Sasaki *et al.* 1999), was developed at NAIST. The approach was to use a hand placed in front of the user's face while the device identified the fingertips via image processing techniques. The user wore a HMD, which later showed a menu with different options, each option corresponding to a finger. The hand had the fingers widely spread and placed perpendicular (in front) of the user's glance direction. When the user touched a particular key, the corresponding function was called. However, a drawback of the arrangement was that, even though it enabled rapid selection of a desired command or operation, there were few commands or operations that could be accessed by a user. Another limitation was that the arrangement was not suitable for entering alphanumeric or other symbolic data as the options were limited to the number of fingers available on the hand.

The future is open to many other applications of such AR based UIs. Some examples are available in this thesis. They cover areas of applications in medicine, education, entertainment, maintenance work, path planning and finding, and military. There are practically no limitations in designing and developing other novel applications (Broll *et al.* 2001).

Virtualizing the interface also means obtaining a greater degree of flexibility. While current interfaces for household appliances are physical, they have limited interaction objective, Information Appliances of the future may operate through more personalised and contextual interfaces (Eustice *et al.* 1999). This class of devices, called as a Universal Information Appliance, could enhance the communication between humans and machines. A user could have access, based on the preferences and the context, to various resources by operating just one device. This universality of devices could bring benefits as to the way the data is stored and adapted for a user's needs as well as the context of use. The future could provide an even more flexible way for operating such devices, by making their interface virtual. Such integration work is important, knowing that the IAs that will populate the future should become invisible in order to be used more naturally (Weiser 1994).

In the future, ubiquitous computers may occupy the information space surrounding a user (Weiser 1993) and could be operated by some universal interface or "intimate assistant" (Rekimoto & Nagao 1995) that may be adaptable and context aware. Such an interface or interface appliance, as this author would like to call it, could act as a universal interface that could be personalized and context aware. Combining the flexibility of a virtual interface with the universality of the interface appliance could result in a very powerful tool to interact with in the information world of the future.

1.4 Definitions

Based on the understanding of the work and surveys of the field, several definitions have to be stated in order to understand better the idioms used in the following chapters.

An *information appliance* is defined as a device specialized in processing information (Bergman 2000b). An IA refers to any device that is capable of processing information, signals, graphics, animation, video and audio; and exchanging such information with another IA device. Typical devices in this category are smartphones, smartcards, PDAs, portable PCs, and so on. Digital cameras, the older cellular phones (only voice), set-top boxes, and TVs are not IAs unless they become capable of communication and information processing functions. Information appliances may overlap in definition and are sometimes referred to as mobile devices, smart devices, wireless devices, wearable devices, internet appliances, web appliances, handhelds, handheld devices or smart handheld devices. They can be embedded in the environment, fixed or mobile.

A *mobile information appliance* is restricted to the class of IAs that are characterized by mobility, or the ability to be moved around easily.

Augmented reality is defined as the technique that enables a user to see artificially generated objects on top of real ones. In this thesis, AR is restricted to sight or visual perception and not to other human senses like hearing, touch, taste or smell, which also could be artificially augmented (Azuma *et al.* 2001). Hence, AR here is a form of VR that supplements the real world with virtual information rather than creating a completely synthetic 3D graphical experience.

A *deviceless interface* is defined as one between a user and an IA that has no mechanical interference with the user. To be more explicit, it means that such an interface is virtual and the user interacts with it via a set of devices that are not mechanical. A classical example would be voice commands. The thesis treats the mobile AR UI as such deviceless interface.

A *mobile computer* is defined as a computer that has the capability to be carried around in daily life without restrictions. There are many antithetical definitions for mobile computers. Here is an explanation why: based on some classifications (Newman & Clark 1999) there are three types of portable devices: Portable Desktop Computers (like Laptops, Notebooks etc.), Portable (Handheld) Digital Assistants – also known as PDA's (like Palm, Visor etc.), Wearable Computers (embedded in clothes or worn on wrist, etc.). A class of other smaller computing devices (chips inserted in the body, etc.) can also be added. Based on the initial definition of a mobile computer, the class of Portable Desktop Computers is not capable of being operated without restriction, as usually a user is required to sit at a table in order to operate the device. The *handheld* or *portable digital assistant* is a small, portable device that is capable of processing, storing and retrieving information. This class of device has the quality of being smaller than a laptop and being operated from the hand (hence the term handheld). *Wearable computers* are defined as computers that can be worn, meaning that they are mobile. Others had defined wearable computers as being portable while operational, having minimal manual input, being context aware and always on (Feiner 1999).

1.5 Publications and Dissemination of the Research Results

Several publications, including a United States Patent (see Appendix) are available on the subject reflecting the contribution and dissemination work from previous research on the topic. The most important publication is a granted patent (Pulli & Antoniac 2004). The work on the patent started in March 1999 with an idea shared between the two authors. The writing of the patent text (abstract, body and pictures) was carried out until December 1999, when the application was filed. The writing and the pictures were mainly the work of this author.

The second important publication is a journal paper (Antoniatic *et al.* 2002). The paper was co-authored by several people, but the contribution of this author was in introducing MARISIL, and some applications of the language.

The first public appearance of the concepts described in this dissertation was presented in Delft to 2nd International Symposium on Mobile Multimedia Systems & Applications (Pulli & Antoniac 2000) by Prof. Petri Pulli. The paper described the transition from the older research on UIs of mobile devices to the new idea of interfaces proposed in this work.

Another important publication was the introduction of the HandSmart prototype idea in a paper presented World Multi-Conference on Systemics, Cybernetics and Informatics (Antoniatic *et al.* 2001b). The paper also presented some of the new applications of the Mobile Augmented Reality Interface Sign Interpretation Language (MARISIL) class of interfaces, like user profiles, virtual office and guardian angel.

Another paper that is regarded important is a position paper presented to 6th International Conference on Concurrent Enterprising (Antoniatic & Pulli 2000). The paper is important because it received considerable feedback as it identified the importance of trust in UIs from the mobile user's business perspective.

A smaller contribution that received a lot of attention (including from highly ranked EU Commission officials presenting it as a key contribution to future mobile interfaces) was presented in Helsinki at the Wireless World Research Forum (Antoniatic *et al.* 2001a).

Other papers discussing the subject were presented by this author at International Conference on Concurrent Enterprising (Antoniatic & Pulli 2001, Antoniac 2002, Antoniac *et al.* 2004).

The public review of these papers refined the perspective over the idea and better tuned the target of the research. However, the dissemination was influenced by the restriction to freely publish anything on the subject during the first year of the research due to patent application restrictions.

1.6 Thesis Contributions

The expected contribution of this work was to:

- Design a system to allow testing of virtualized interfaces for mobile IAs;
- Evaluate the system comparing it with current systems;
- Survey the technology limitations and physical constraints for building such a system;
- Identify the benefits of using such a system;
- Specify new requirements for a future system;
- Build the framework for future development of such interfaces.

The ambition of the contribution was to raise the area of advanced UI for mobile devices specifically to identify new research fields that could contribute to the implementation of the interfaces for future mobile devices. The work would also have to provide an answer to the question of how well the new generation of mobile devices could support the deployment of applications that are more advanced and how the interaction with such applications would be supported when using a mobile device. Another expected result would be an exhaustive survey of the technologies required to enhance the interaction of mobile users.

Yet another expectation would be to build a prototype and define a framework, providing a starting point for future development and proof of feasibility of the new concepts and constructs.

1.7 Thesis Outline

The present work is structured in several parts which discuss the background, specifications, building and evaluation. Some parts are further divided in order to clearly cover the important topics such as interaction and prototyping.

The first part outlines the background work related to the subject. As the work contributes to mobility, Chapter 2 introduces and defines the mobility and the mobility requirements of UIs. In Chapter 3, the survey of earlier systems provides the findings from a theoretical angle, while Chapter 4 describes the findings from the empirical standpoint. These findings later lead to a set of specifications for the system.

Following is a more detailed description of the chapters.

Chapter 2 defines mobility and identifies the set of requirements to improve the current mobility for IAs. The chapter also includes the mobility requirements from the mobile IA UIs perspective.

In Chapter 3, the early mobile systems and their UIs are assessed. The two parts of this chapter introduce the early devices and their applications concluding with the challenges that they pose.

The last chapter (Chapter 4) of the first part of the thesis presents the early experiments and discusses the design problems that were faced in the development of a “virtualized” interface. The chapter explores two possible solutions to enhance the interaction of mobile devices. One solution was the compressed or dense displays. The

second solution, more efficient but harder to implement, was of virtualizing the interface and extending the displays.

The second part of the thesis introduces the abstract concepts and the new design constructs for a novel interaction technique that could support the “virtualizing” of the interface idea. This part includes the specifications of a new interaction mode called the MARISIL (Chapter 5) and the evaluation of the new proposed language (Chapter 6) as well as the impact it has on originating new applications and services (Chapter 7) for the future. The conclusion of the evaluation underlines the capability of the new interaction technique to extend the class of applications operable from within a mobile IA. This evaluation of the abstract constructs is followed by the introduction of the prototype implementation for a “HandSmart” device (Chapter 8) capable of interacting using MARISIL specifications, and an evaluation of the implementation (Chapter 9).

Chapter 5 provides the specification for a new interaction (MARISIL) that will enhance the interaction for mobile IAs. The specifications were based on the empirical data provided by the first part but they were also concerned with usability and intuitiveness of the language constructs.

The validity of the specifications was probed by a set of tasks commonly used on mobile information devices (in this case, the mobile phone) that have been evaluated and compared in Chapter 6. The results of the evaluation contributed to the findings on the implications of the new interaction technique in novel applications and the capabilities to handle complicated tasks (Chapter 7).

The implications or the impact that the new interaction technique would have on new applications and services is the subject in Chapter 7. The discussion analyses the impact on devices and introduces new applications that could work better with the new device and the interaction technique.

The last part includes the details of building the prototype that could handle the interaction specified in Chapter 5 (Chapter 8) and the prototype evaluation (Chapter 9). This part of the thesis also discusses some of the implications of the findings including details on designing such devices. The human aspects of the technology and how it could improve life are also discussed in this part.

Chapter 8 describes the building process and the components that were included in the system. The chapter examines both the hardware and the software implementations that handle the interaction. Examples of commercial devices and an analysis of the components used in the system are also indicated.

In Chapter 9, an evaluation of the prototype is detailed including some of the implications discovered during the development. The chapter also presents the perspective and the next implementation challenges for the system.

Finally, the thesis ends with a concluding chapter (Chapter 10) followed by a proposal for further research (Chapter 11) which discusses the future research perspectives.

2 Mobility and Mobility Requirements for User Interfaces

Mobility could be defined as the quality or the state of being mobile (Company 1994). Based on this definition, the classification for the term mobility is broad because it should cover the majority of things that move. Even so, a narrow classification is proposed, having the epicentre in the mobility of the IAs (cf., to definitions from page 36) from the point of view of the UI. Based on this classification, a set of requirements for the UI has been extracted and is presented in the last part.

The proposed categorisation is the result of combining several papers (probably the most complete work was done by Rodden and co-authors (Rodden *et al.* 1998)) that described or defined mobility from both telecommunication and Human Computer Interface (HCI) perspectives. However, the main topic was the UIs and their requirements when used by a mobile IA.

When analysed from the point of view of the data, some groups have categorised mobility as having three separate components: users, computers and information (from a database point of view (Heuer & Lubinski 1996)). From the user perspective, the mobility splits into three separate components: mobility of people, infrastructure and information. These components are independent and each can be again split into sub-components. For example: users, devices and applications.

Table 2 summarises and describes a taxonomy of these three independent components. The categories proposed do not cover all possible aspects of mobility (e.g., social implications of mobility in families, generations or groups/clans). In the proposed approach, the focus was mainly on UIs and their collateral implications (as in infrastructure, information and users).

Table 2. Taxonomy of mobility from User Interfaces perspectives.

Category	Subcategories	Correlated matters
People	Individual	Work
		Leisure
		...
	Group	Friends
		Colleagues
		...
	Organization	Companies
		Societies
		...
	Nation	...
Infrastructure	Information Appliance/Terminal	Free
		Embedded
		Pervasive
		Link (i.e., mobile IP, GSM, WiFi...)
	Network	PAN
		LAN
		WAN
		...

		Code
Information	Application	Data
		RPC
		...
		Location
	Service	People
		Infrastructure
		...
		...

		...

The following sections detail each of the cells in the table, adding some extra dimensions to some specific components.

2.1 Mobility of People

The mobility of people can be divided as mobility of individuals, groups, organizations and nations and it could be extended to larger groups. What is important is to notice that each of the categories is somehow independent. While individuals are part of a group, in terms of mobility from the UIs perspective, they are a special category. Individuals can access the information in different ways, but when they are in a group, they could act together or even use the same UI (e.g., a mobile projective display).

The mobility of individuals could embrace different aspects. From an activity point of view, it could be a leisure or work activity related mobility, and each aspect could come with more ramifications.

Group mobility could also embrace different aspects, depending on the interests of the individuals or the rules of the group. It could include family, colleagues, teams and other common interest clusters. What is important to notice is that, from UIs perspective, groups represent another dimension of interactions that the mobility embraces. The users can share documents, share infrastructure, communicate, integrate, and develop together using as much as possible of a common environment. The UIs should be adaptable and aware of the individual preferences within the group in order to support this special interaction. Moreover, the mobile aspect adds more dynamics to the concepts of a group, considering the privacy and security of a user while in a group (notice the clear separation between individual interests and the group interests).

A larger form of clustering of individuals is in organizations. This involves a more ordered type of interaction. It could also include inter and extra-organizational levels of mobility. While the grouping could be around common interests, the organization could include common rules, infrastructure and services that are available to the users. From the mobility perspective, this category concerns the maintenance and structure of mobile resources and the capabilities to adapt to changes (mainly of a geographic nature).

When looking at an individual and his/her determination to be mobile, the classification of mobility was split (Pascoe *et al.* 2000) into three components: navigation, sojourn and promenade. Another important component was work-oriented activities (mobile work, emergency work, field of combat). This concluded the splitting of the term mobility based on the scope of the motion to four components as is described in Table 3.

Table 3. Classification of mobility from user's motivation to move perspective.

Applies to:	Motivation
Individual	Navigation (the user is involved in travel activities and needs assistance) Sojourn (the user resides temporary at a different location, as a leisure visitor or on a work trip) Promenade (the user is missing a specific destination) Work (the user is having a task that by nature is mobile—commonly defined as mobile work)

The need for mobility comes from various contexts in which the user is moving. These contexts could be travelling as a tourist (Feiner *et al.* 1997, Kuutti *et al.* 1999, Dahne & Karigiannis 2002) or navigating outdoors in a city (Behringer *et al.* 2000); it could also support indoor navigation (Butz *et al.* 2000) or outside office activities (Raskar *et al.* 1999).

2.2 Mobility of Infrastructure

This mobility does not only refer to the ability of people to move, but also how they are able to use the infrastructure while being mobile. Infrastructure supporting mobility and the mobility of infrastructure are probably the most common subjects for the research in mobility. Parts of the infrastructure are the devices and the networks interconnecting them.

A good approach in classifying devices was introduced by Rodden and co-authors (Rodden *et al.* 1998). In their paper, based on their capabilities to exchange information and other resources with the environment they split the devices into three categories: *free*, *embedded* and *pervasive* (Table 4).

Table 4. Mobile devices and devices supporting mobility classification based on their information and resources exchange.

Applies to:	Type of exchange
Device	Free (the device is independent)
	Embedded (the device is enclosed in another device or environment)
	Pervasive (the device functionality is spread through the environment)

A proposed classification is based on the classes of devices and how they are used. Table 5 details this classification. Some of the fields of the classification are dependent. For example, a wearable device could also access remote resources. The classification proposed here tries to cover all the possible cases of usage found in the literature and in practice, and hence it did not seek the classification based on independent terms.

Table 5. Mobile device classified by usage of device.

Applies to:	Usage
Device	Wearable (the device is worn on user body)
	Remote (the device is a communication device that accesses remote resources)
	Portable (the device can be moved but is not small enough to fit in the hand, more like a laptop)
	Handheld or micro-mobility (the device is an information devices like a PDA that can be held in the hand)
	Ubiquitous (the devices are invisible to a user but they assist the user in mobile activities)

The proposed classification can be classified by the one suggested by Rodden and co-authors by including the classes from Table 5 as follows: *wearable*, *remote*, *mobile* and *handheld* could belong to the *free* category of Table 4; *ubiquitous* could belong to *pervasive*; while *remote* and *handheld* or *micro-mobility* could be also included in the *embedded* category. Rodden's proposed taxonomy is more generic and therefore, it has been included as such in Table 2. The taxonomy proposed here is more empirical and hence more specific.

Some have argued that the mobility of the activities should be included in the classification (Pascoe *et al.* 2000) as there might be some cases where the mobile device is used in a static activity (Table 6). However, these are just specific cases and they are

only relevant when some classes of interfaces are not operable in a dynamic activity (like writing an email from a laptop while walking on a street).

Table 6. Classification of mobile devices based on their activities.

Applies to:	Type of activity
Activities	Dynamic (the user is moving while using the mobile device)
	Static (the user is fixed while using the mobile device)

Because of the abundance of contexts in which these mobile activities occur, a device could track the attributes or context in which it is used and change accordingly. When a device is context aware (a good example is available in the Hinckley and co-authors (Hinckley *et al.* 2000)), it could help to tailor the applications for mobile devices. Moreover, it could benefit from the special nature of the context in which the user is operating it, e.g., navigation.

While the devices are an important part of the infrastructure and the closest category to the subject of this thesis, the network contributes to many of the functions used by the devices.

From the mobility point of view of the UI, the network component could be split into levels of topological access to remote resources. They could be near the user (Personal Area Network or PAN) evolving through a Local Area Network (LAN or Intranet) and a Wide Area Network (WAN or Internet). Another important component would be the link or the protocol and how well it could support mobility (there might be problems like optimal routing, handover support or roaming). Sometimes, not only the logical part of the network infrastructure would be mobile but also the physical part could become mobile (as in satellite network coverage).

2.3 Mobility of Information

This mobility is about accessing information by people, anytime (sic), anywhere (Perry *et al.* 2001). The last category defining mobility is, therefore, the mobility of information and is another category that is close to the topic of this work.

The mobility of information could be split in two categories, applications and services, based on their support for mobility and the location of the information. The mobility of applications means the ability of an application to work in a mobile environment (in other words, to support mobility) while the mobility of services is how the applications provided by service providers are accessible from a mobile environment.

Applications are defined as computer programs designed for a specific task (Company 1994). Mobile applications are, therefore, computer programs designed that they are accessible or operable from a mobile platform. This means that the code and data should be either available or capable of executing on a mobile platform or should be accessible from a mobile platform while located on a remote computer. A more exhaustive description can be found in the Fuggetta and co-authors paper (Fuggetta *et al.* 1998) that examined code mobility from the point of view of the programmer. Heuer and Lubinski,

on the other hand, looked into data mobility and how to access databases from mobile environments (Heuer & Lubinski 1996).

Services could be defined as activities performed by one party for the benefit of another. From the mobility perspective, services are applications that run on a remote location and are used from a mobile environment. Due to the mobility of data and code, and generally, the distribution of resources in a mobile environment, it is hard to distinguish between mobile services and applications. In the approach here, service is a group of applications that reside on local or remote locations and contribute to a common activity. There are infrastructure services that support the infrastructure but they are not so important from the point of view of the UI (they should be invisible to the user). There are also personal services that enhance a user's personal experiences. Another important group of services for these studies were location based services.

Personal services deal with a user's personal interests. They could handle incoming calls (like call waiting) or they could provide access to data (calendar, office applications). Parts of the personal services interact with location based services, particularly the context-aware data gathered through these services. Probably the most challenging research is the one to study privacy protection and anonymity of the user when using personal and location services due to the amount of data gathered on user's preferences and customs.

Location based services are a group of services that are aware of the geographical position of the user and provide more specific output to the user. Sometimes, knowledge of the location of the user is not sufficient and combinations of other sources of information (like user preferences, type of user's activity or the time) contribute to better access to service resources (Giaglis *et al.* 2003).

Giaglis, Kourouthanassis and Tsamakos (2003) classify location based services as: *emergency, navigation, information, advertising, tracking and billing services*.

Emergency services deal with emergency calls and how to handle the situation (sometimes having to reveal automatically the location of the user).

Navigation services provide a user with fast routes, traffic information, indoor and outdoor directions.

Another location based service is the *information services* that could provide a user with important information on various sources of data like yellow pages, travel services or even infotainment.

Advertising services are another part of location based services that contribute to better information access on products or other services from specific locations. They could include alerts, advertisements, banners and guides.

Another important part of location based services is the *location sensitive billing services* that could facilitate mobile commerce combining the location with a purchase (it adds a new dimension to advertising).

While this classification was quite broad, Gialis and co-authors failed to take into consideration other categories than individual users. Even services for individual users should support other activities than leisure time activities. For example, services to support work-related activities, like sharing of resources, or services supporting mobile engineering. Some work-related location based services include support for a person working on maintenance (Neumann & Majoros 1998, Klinker *et al.* 2001, Doil *et al.* 2003) or inspection (Webster *et al.* 1996, Tang *et al.* 2003). These services require

infrastructure support and special settings. They could be classified as belonging to both emergency services (if the work is related to emergency situations) as well as information services (as work related information sharing and access) without necessarily being either of them. Because of the importance and the number of services related to work, the classification should include *mobile work services* as another class. These services should support the tasks of a mobile worker in outdoor and indoor locations, and they should be concerned with security and reliability along with services supporting collaboration, sharing and accessing of resources.

2.4 Mobility and User Interfaces

While previous sections defined the mobility from a broader perspective, this section analyses the implications of mobility on UIs. The UIs for IAs (cf., definitions from page 36) have a long history of design, starting with the old 1950 batch mode cards with punched-holes, to the more advanced graphical UI (GUI) or even the “post WIMP¹ GUI” of the future (Dam 1997). Throughout history, these devices have all been static. They were on large shelves in computer rooms or on the desk and difficult to operate on the move. The challenges of today are to make these IAs mobile, and similar to any IA, mobile devices need a UI. Unlike the UI of desktop computers, those for mobile ones have the special requirement as to be operable while the user is mobile.

A mobile IA should be able to provide similar resources as a fixed or desktop computer, with the additional quality that it could operate while a user is mobile. Even so, the abilities to unbind the links that keep IAs fixed are hard to break. The current approach in designing UIs for mobile IAs are concerned mostly with the emulation of the functions available for a desktop computer. So far, very few researchers have been emphasising the mobile aspect of the interaction when using mobile devices. Some even argue that mobility is not always necessary and that a system, even though designed to be mobile, could eventually be used as a fixed one (Perry *et al.* 2001). However, the wave of support for various types of mobility should lead to important changes on how input and output for these devices are being designed and built (Lyytinen & Yoo 2002).

While mobility and interaction when mobile are important, the context in which mobile devices are used is also relevant. In general, it is safe to say that the user’s privacy is critical, but in certain contexts, the infrastructure should also allow a certain level of proximity or location information to the system. In certain aspects of mobility (like mobile workers) the information on possible disconnections from the network takes precedence over privacy. Failing to do so could induce dangerous consequences on how data is interpreted and reacted to it in special situations like safety critical fieldwork (Rodden *et al.* 1998). Mobile interfaces should provide, in case of an error or disconnection, information to others in a group, hence enabling them to access the resources properly.

¹ WIMP means Windows, Icons, Menus and a Pointing device, typically a mouse. WIMP GUI is the class of graphic UIs that current operating desktop computers are using.

Even though recent technological advances have reduced the size of system components, and even after incorporating latest developments in the field of mobile IAs, the interfaces and the design patterns being used are still paradigms from desktop computers that are not suited for the mobile environment (Dempski 1999, Newman & Clark 1999). For example, while in desktop computers a user can handle the interactions quite naturally in front of a big monitor and tabletop keyboard, mobile devices are bound to small screens (Brewster 2002) and little keypads. Many approaches have been made to resolve these issues: audio, voice and facial recognition enhance the experience of a mobile user; various sensors also contribute to a better context and location awareness; augmented and MR increase the information availability and the way it is represented. Combining all these technologies could generate a wider range of implementations and development for these types of devices in the future.

2.5 Mobility Requirements for User Interfaces

Information appliances when becoming mobile require certain tailoring in the design, especially knowing that mobility usually means that the size of the devices will decrease. Unfortunately, becoming smaller also requires that the screen is smaller. With smaller screens, even if using enhancements like sound (Brewster 2002), interaction could become difficult. From this perspective, the first and most important requirement is to extend the screen size.

Desktop computers, due to their nature, were merely concerned with desk activities. These activities are involved mostly with work related tasks, but entertainment and games applications have also been available. Mobile devices, on the other hand, are free of such a strict limitation. They have the potential of reaching a larger segment of the population that is not bound to a fixed environment, like the desk. As a result, an analysis of usability requirements of the UIs for mobility is important and needs consideration.

The definition of usability given by Nielsen (Nielsen 1993) described it as being about: learnability, efficiency, memorability, errors and satisfaction. Taking the context in which they are used (Johnson 1998), the UI design of a mobile device should consider the following aspects: *networking*, *mobile vs portable*, *lightweight*, *ergonomic* and *non-intrusive*.

The *networking* property of a mobile UI refers to the quality of accessing remote information and resources. Remote access of information also necessitates a certain security level for the communications. Even though this requirement concerns the infrastructure, remote access of information and the ability to interface with other devices is important for a mobile device (Eustice *et al.* 1999).

The *mobile vs portable* aspect refers to the capability of a UI to operate while a user is in motion. It is difficult to operate a desktop computer, even if it could be moved, (like a portable/laptop computer) while walking or driving.

The *lightweight* attribute concerns the device and how the interface operates from the point of view of weight. A heavy device would cause weariness to a user and could lead to abandonment when long tasks need to be carried out.

Due to the situation in which mobile IAs operate, their UI design should also concern the *non-intrusive* aspect of the interaction. A UI that will block, distract or scare a user while driving could cause accidents. Moreover, knowing that mobile devices should assist a user during various activities at work and during leisure time, they should include some personalization features. The requirement for more personalization in mobile devices could handle, better and more efficiently, the *non-intrusive* aspect of their operation (Julier *et al.* 2000).

As mobile devices become smaller and more powerful, the interaction and applications would require more space for presentation. While an alternative is to provide solutions for better use of the limited screen, increasing the size of the screen appeals to developers as well as users. Even with more effective combinations of sound and touch screens (Brewster 2002), the screen of a mobile device would require more space to allow more complex interaction and better presentation.

Another requirement comes from the point of view of marketing. Mobile devices, while being personal, should also be affordable and attractive. Attractiveness implies a more flexible approach in their UIs (adaptable) and an inventive approach that would generate better adoption of these devices. Even though many people have expressed the need for a mobile IA, they have most often failed to find one that satisfies their needs or attracts them sufficiently, primarily because of the weight as well as the difficulties to learn how to operate them.

2.6 Summary

While the importance of mobility increases due to technological advances and increase in social demand, further studies should provide answers to the generic requirements for mobility. Moreover, a holistic view of what mobility means and how to develop future IAs in order to accommodate mobility are also important.

Setting the requirements for a system to support mobility could be a laborious task. This is due to difficulties in forecasting and mapping cultural differences, various individual needs, social patterns and behaviour. However, some stronger threads are present and when extracted they provide the basis of the requirements for mobility, some of which have been described in this chapter. For example, while devices are required to be smaller in order to be handy and portable, display size or density should increase in order to allow more interaction and presentation. In addition, while technology advances to allow more advanced applications, social diversity requires simplicity, eventually leading to adaptation and personalisation of the mobile UI.

Other requirements are remote information access, ergonomics of interaction (non-intrusive, light) and the ability to learn to operate a device in a shorter time (intuitiveness of the UI and flexibility).

These requirements are mostly from the UI perspective. Other important requirements can be deduced from the information and network perspective. While more services become available and personalisation could enhance the current ones, the importance of

security increases. Therefore, important requirements are the security and privacy of mobile IAs of the future.

To conclude, the list of requirements from the point of view of mobility is concerned with: display size, lightweight, ergonomics, flexibility, adaptability, network, security, personal, privacy and simplicity to learn. These have been identified as the basic requirements for a mobile IA. In subsequent chapters, the evaluation of these basic requirements helps to derive a new class of requirements, expanding the list and their implications.

3 Prior Mobile Systems and Their User Interfaces

The challenge of virtualizing the interface is complemented with the challenge of having a system mobile. The mobility of the interface is, therefore, an important feature of a system. A popular definition of mobility, away from the classical one, is to access information at anytime (sic) and anywhere (Perry *et al.* 2001). The choice to have both a mobile system and a virtual interface could be realised by implementing it in a MR environment (Milgram & Kishino 1994), or more specifically, within AR.

Augmented reality enables the user to see the surroundings while overlaying virtual objects into it. It has the base for better interaction by adding more scalability and flexibility to the UI design (Rodden *et al.* 1998). Moreover, AR, by augmenting only a user's view, results in better privacy when the user performs tasks in public. Using concepts like eye tracking (Flickner *et al.* 2001), or the scrambling of the interface keys (Hoover 2001), would provide a user with enough security guarantees to allow access of sensitive resources (like documents, banking including entering passwords or other sensitive data) in public spaces without fear of possible intrusion.

Even though traces of AR systems date back to year 1896 (Stratton 1896), the research field, as it is today started its existence around year 1950 (Sutherland 1968, Kalawsky 1993c). However, AR systems are rarely available in UI implementations. In fact, the first approach of using AR for building a UI appeared in the late 90's as a possible solution for interaction in a wearable computer (Pouwelse *et al.* 1999) and for a controlling a waste-water plant (Bertelsen & Nielsen 2000). Even so, the interfaces were more like an extension of the current ones. They took advantage only of the ability to overlay information on objects (as in augmenting the interaction (Rekimoto & Nagao 1995)) and not interacting within it.

It can be concluded that until the late 90's the application of AR to UIs was rare, if any. The research was mostly concerned with the feasibility of a system rather than of the UI. Several implementations of AR systems that were close to the present work and some of their applications are seen in this chapter.

The chapter is a survey of past implementations of mobile systems, with a special emphasis on their capabilities to support virtualizing the interface or augmenting the view. The survey, hence, primarily presents the research results from the AR field as they appear to be most relevant for the purpose of this dissertation. In a following chapter

(Chapter 7), a more detailed discussion will focus on some of the more advanced features used with AR.

3.1 Devices Made for Mobility

This section discusses the availability of various mobile platforms for IAs. More specifically, the section looks into the designs of wearable computers, mobile devices and ubiquitous computers, trying to identify the challenges and the drawbacks of their implementations. The examination of their design looks mainly from the UI perspective, and how virtualizing their interface is feasible.

3.1.1 Wearable Computers

Wearable computers are IAs that are worn on the body. They could continuously operate or assist (i.e., they can be continuously operational and accessible; see also the definition available on page 36) a user in activities (Billinghurst *et al.* 1998). The difference between this type of computer and a mobile or portable computer is that wearable computers are clothing on the user's body. Many wearable computers laboratory implementations use a laptop or portable computer as the computational device. A good extension for wearable computers is the augmented view – combining the mobility of the platform with the mobility of accessing the information in a better visual way (Klinker *et al.* 2000).

As wearable computers, by definition, operate in a ceaseless manner, they should also provide the user with more permanent access to their output. A solution is to use see-through eyeglasses. This would allow the user to see the surrounding while still having some information available on the display.

One of the first groups to present the benefits of using this computing platform in combination with AR glasses was the MIT Media Lab (sic) Vision and Modelling Group. In 1995, they presented a technical report (Starner *et al.* 1995) in which AR was used to enhance the use of their system (Fig. 10). This system did not interact with the user via an AR based UI (it was only for displaying purposes), but the team were the first to identify the benefits of using AR in combination with a mobile system, such as a wearable computer. Their applications were various, from supporting students (data storage) to augmenting memory (face recognition and tagging the person situated in the visual range).



Fig. 10. MIT-Media Lab Vision, Wearable Computer made from PC104 boards and the application of it in a Private eye™ display mounted on safety glasses (Starner *et al.* 1995).

Unfortunately, their system did not interact with the user via an AR based UI, but instead used what they called “the Twiddler” (Fig. 11). This was a keyboard-like interface that was used to input text and interact with the device.



Fig. 11. The MIT Twiddler. One hand keyboard for wearable computers to interact while the system was using AR to overlay information to the user.

A new wave of commercially available technologies, identified by Behringer and co-authors (Behringer *et al.* 2000), provided researchers with new tools and devices for wearable AR systems. These new technologies also encouraged researchers to look deeper into this field and to come up with new applications. For example, a novel idea was to operate a wearable computer by using AR views and tracking of the head (Billinghurst *et al.* 1998). In their project, Billinghurst and co-authors talked about a conferencing application in which a “wearable user” carried a HMD with a tracker. The participants were virtual and they appeared to the user as static pictures. The user could choose which participant to listen to by turning the face (some audio culling techniques were also involved). The display was monoscopic (which was popular at that time) and,

therefore, the system was not “fully augmented”. Even so, the application was one of the first in which a user interacted with the system using augmented objects.

Another proposed approach to a fast and light wearable computer based AR was to split the system into the mobile and fixed parts. The fixed part processed the heavy computational work, while the mobile one handled only the interactions. The link between the two parts was done like in the mPARD system (Regenbrecht & Specht 2000), via a radio frequency with 2.4 GHz analogue radio or some other wireless technology. This distribution or balance of the computational load via wireless network reduced the power consumption (Pouwelse *et al.* 1999) and increased the processing power available (Pasman & Jansen 2001). The distribution of the resources varied. The remote provider could support the full computational process or it could support only part of it (like in Pasman and Jansen where only the rendering was supported). Such systems had a certain latency due to radio wave transmissions, so the choice when selecting the method was between having the processing power versus the latency of the answer provided by the remote provider (sometimes the same result could be provided in a shorter time by a slower processor, compared with the shorter processing time but with the added latency).

A better and more recent example of a wearable computer with AR was the one provided by Reitmayr and Schmalstieg (Reitmayr & Schmalstieg 2001). In their approach (called Studierstube), the mobile person interacted with the system via a tablet and a pen. The wearable system was a combination of a laptop and an AR system (camera and see-through glasses). The inputs from the various devices came from the camera, detecting the position of the tablet and pen. Even so, if the pen touched the table, the tracking shifted from video sensors to the tablet’s own touch sensors. In this way, a more accurate input was available for the system. An application for the system included playing chess, but other virtual or AR applications were possible. A picture of the system in which a video camera is present on the helmet and the processing unit is a commercial laptop is presented in Fig. 12. On the left, the components used by the system for the I/O operations are visible.



Fig. 12. The Studierstube (Reitmayr & Schmalstieg 2001). System includes Wacom graphics tablet and pen. The devices are on the left and the user wearing them on the right.

The implementation of an outdoor combat environment for the army also benefited from the combination of AR and VR running on a wearable computer (Piekarski *et al.* 1999). The system took use of both AR and VR in order to have a complete simulation

environment. The Tinmith-II, called such in the Piekarski and co-authors paper, was not the only military application which took advantage of combining AR with Wearable Computers. In their paper, Julier and co-authors also described such a system used for information filtering in a civil defence exercise (Julier *et al.* 2000). Their system was able to filter the information and show only relevant data to the user. The application was in the area of urban warfare, for a sniper prevention system. The filtering included several criteria like location, objective and user input.

Azuma's survey (Azuma *et al.* 2001) also gave an example of a wearable computer used by the Naval Research Laboratory, called the "Battlefield Augmented Reality System". The system was able to provide field operation personnel with 3D information (such as goals or hazards) that were otherwise only available on 2D maps.

While there are many benefits of using wearable computers, some people (Rhodes *et al.* 1999) have identified some problems when using the systems in certain applications. Even though an important requirement for a mobile system is privacy, accessing localized information (i.e., the information about a location), resource management (i.e., access by multiple users to the same resource) or even storing unprotected personalization information, could pose a threat to privacy. This is because by using localized information, the system would require tracking of the position of the user, information that is, by definition, private (unless otherwise indicated by the user). Moreover, storing the personalization information could pose the risk of leaking it, hence providing insightful data to intruders on the habits of a user. An often seen solution is to cloak the user's identity and to encrypt the data. However, the risk of breaking the privacy of users is high leaving the problem unresolved.

The social aspects are also a matter of concern when mass use of such platforms is considered (Feiner 1999). Imagine the damage that could result if the recording capabilities of such system were to track or monitor surrounding activities. When entering a secure military establishment or other places where secrecy of information is the enforced policy, such devices could supply a significant amount of secret data to third parties and, therefore, unknowingly break security restrictions.

More important for consideration is the contribution of such a platform to enrich the social life of the future. Entertainment and interaction have been given a new dimension when used with this new design. Consider, as a simple idea, the use of wearable computers in a theatre (Cheok *et al.* 2002), and how traditional theatre could change its forms of expression in the future.

Research on wearable computers has yielded good results on how to interact and benefit from a virtualized interface. The results have permitted scientists to look further into applications of AR when used with a smaller system like a mobile device.

3.1.2 Mobile Devices

Like wearable computers, mobile devices are also inclined to support applications of AR. To clarify the distinction between wearable computers and mobile devices, the latter do not require to be worn on the body. Unlike wearable computers, which are at present

mostly in the research laboratories, mobile devices have reached the level of becoming commodities. They have penetrated markets and are present in every day life. Handheld computers, laptops, PDAs and even mobile phones, are classes of mobile devices.

The most popular class of mobile devices is the subclass “mobile phones”. This class had advanced rapidly towards being a mobile information communication device, since the latest models embed large processing capacity and include advanced applications. Moreover, commercial products are more available and able to support the combination of mobile device functionalities with several features of AR systems. One example is the portable 640x480 resolution display designed for a handheld Pocket PC (IIS 2003). Fig. 13 shows the level of miniaturisation reached for these devices.



Fig. 13. Interactive Imaging System’s Second Sight M1100 Display for handheld pocket PC (IIS 2003).

The pace of development has encouraged researchers to look more into the future of the field. Latest research deals with the size, weight, power, processing capabilities and display quality. Besides the engineering and technological problems, social and human aspects are also important.

This work mostly concerns the problems of interacting with this class of devices and subsequent sections give a more detailed description of this area.

3.1.3 Ubiquitous computers

Ubiquitous means being available everywhere simultaneously. The vision is that the development and maturity of the technology will allow deployment of information devices on such a scale that the devices would become invisible to users. Researchers define such devices as being everywhere and enhancing humans with a better understanding of their surroundings. The ubiquitous computer was defined for first time

by Mark Weiser (Weiser 1993) as a computer that is present everywhere. The driving force for the field, as stated by Mark Weiser, would be to provide hundreds of wireless computing devices per person per office, of all scales (from 1 inch displays to wall sized ones).

Such infrastructure, when deployed, would require new interaction techniques suggested by the new level of access and availability of information provided by the pervasive nature of ubiquitous computers. As the infrastructure develops, these new interaction techniques should cope with the problems of ubiquitous computers (to mention two, privacy and personalization (Rhodes *et al.* 1999)). The ubiquitous computers infrastructure necessitates a different interaction paradigm that goes beyond the current one of desktop computers. The new interaction should be more natural, context-aware (Abowd & Mynatt 2000), and allow more physical-real exchange of information (Ishii & Ullmer 1997).

An interesting approach would be to utilize AR to support better interaction between the human and the ubiquitous computers (Newman & Clark 1999). Even if the environment should be “invisible” to the user in respect to the definition, some interactions would be required. Augmented reality can contribute to this as it allows the exchange of information unobtrusively between the surrounding information infrastructure and the user by overlaying the view of the user with certain data provided by the ubiquitous environment.

Pairing AR with ubiquitous computing integrates the information provided by ubiquitous computers into the world of the user. Even though some argue that these two approaches are sometimes complementary (Rekimoto 2001), AR could become an extension of the ubiquitous computer in the user’s world. To give an example, the user could enter a room in which the ubiquitous computer is operating. The user can then see what resources are available by using the AR glasses as the output device. The glasses would provide this information to the user only if required, for example, by using the magnifying glass metaphor (Rekimoto 2001), hence ensuring the non-intrusive nature of ubiquitous computers.

Although introduced over a decade ago, the goal of ubiquitous computing to provide seamless interaction between human and the surrounding informational environment has not yet been achieved. However, the benefits of having such an infrastructure are important, especially when leading to seamless interaction, or everyday pervasive computer access. Even if ubiquitous computers do not imply mobility by default, the omnipresent nature of the deployment supports the idea of a user moving freely within the informational infrastructure, hence encouraging, indirectly, the mobile nature of the user.

Unfortunately, some of the problems with ubiquitous computers are the cost of deployment and reliability (are they going to work together and how will it be known if they are functional or if they are not). Coupling ubiquitous computers with AR could solve some of these problems, harmonising the interaction between the human and computers available in the surroundings by sharing the information between the real and digital world.

3.2 Applications for mobile augmented reality systems

The previous section described mostly devices or hardware platforms. In this section the focus is on the potential use of such devices, specifically the applications running on them.

3.2.1 Support and Maintenance

The first project in which the term AR was first coined was where the two Boeing researchers, Thomas Caudell and David Mizell (Caudell & Mizell 1992) were dealing with the problem of wire bundling for airplane assembly lines. The continuation project that dealt with the ideas promoted by Caudell and Mizell was later started in January 24, 1999 (Mizell 2001) and the main topic was to handle “formboards” used by Boeing employees when grouping wires into bundles that could be used later in the wire framing in airplanes. The task was to automate the manual reading of the diagrams showing the routes, route and tie off the bundle. Their first prototype dealt with the technological and financial problems. The demonstrator was not a see-through HMD but, by the year 1999, the system emerged into a fully functional system capable of substituting the paperwork included in the bundle kits.

From the tasks of supporting a user in the current job to maintenance use, it was a small step. The AR found many applications in maintenance activities like equipment (KARMA Project (Feiner *et al.* 1993), aircraft (Neumann & Majoros 1998)), building (Augmented Reality in Architectural Construction, Inspection, and Renovation (Webster *et al.* 1996)) and remote maintenance (for power plants, chemical refineries etc. (Navab *et al.* 1999, Stricker & Navab 1999, Klinker *et al.* 2001)). An example of the use of AR for assembly and construction support is available in Fig. 14. The real picture of the bars, the virtual bar and the augmented view can be seen from left to right in the figure.



Fig. 14. Augmenting the job of construction worker (Available from Computer Graphics and UIs Laboratory at Columbia University, project Augmented Reality Spaceframe Construction (Feiner *et al.* 2004)).

3.2.2 Manufacturing

Another area of applications in which AR can play an important role is in manufacturing systems. Many times in Flexible Manufacturing System (FMS) there is a need for human intervention in the process (errors, material handling, supply, maintenance). In such cases, the AR system could supply the operator with information about parts, the tasks to be performed, with directions or navigation information (Barfield *et al.* 2001). A more detailed representation of the activities that could benefit from AR applications is available in Table 7.

Table 7. Applications of AR to Manufacturing Activities (compiled from Table 23.1 (Barfield et al. 2001)).

Manufacturing activity	Task	Application
Product Design	Using CAD tools for prototyping parts and assemblies	View CAD models in 3D overlapped on real workspace environment
Fabrication	Machining parts manually or supervising machining operations	Provide instructions or diagrams in real time to workers for machining parts, or supervising the machining process
Assembly	Joining parts together in the correct order and orientation	Provide instructions or diagrams in real time to correctly matching and assembling the different parts by overlaying them on the real objects
Inspection and Testing	Visually inspecting items for defects. Taking measurements to establish properties (dimensions, hardness, weight, roughness)	Help inspectors locate where visual inspection performs and/or where measurements are taken. Provide evaluation and disposition instructions
Material Handling	Locating equipment in the facility to obtain suitable material flow pattern. Moving parts, tools, equipment from one location to another	Allow designers to place virtual machinery on the shop floor and run simulations in real time, aiding in facility layout and redesign. Provide location and handling information to the workers

A powerful example of how AR could improve the work in manufacturing is the layout of virtual information over the real environment of a factory floor. The integration between the real objects available on a factory floor and the planned image or possible arrangement of equipment available digitally, could provide useful information when planning future assembly lines. Fig. 15 shows how the virtual machinery integrates into the environment when using AR. Such an approach in planning shortens the time to complete the task, leading to cost-reduction and other improvements in the manufacturing-planning field (Doil *et al.* 2003).

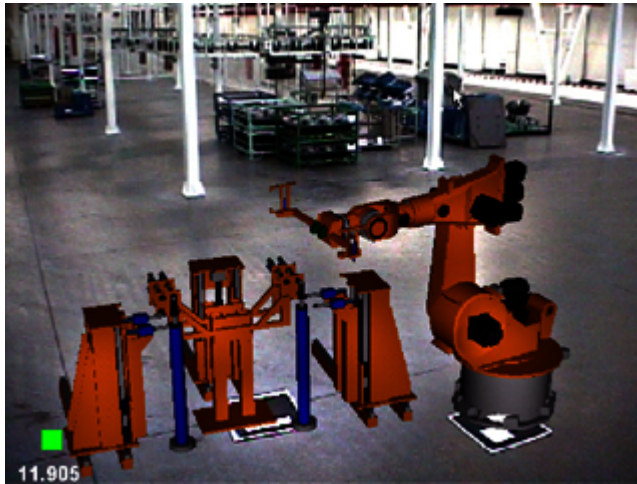


Fig. 15. Augmented reality assisting manufacturing-planning activities (Doil *et al.* 2003).

Using AR in various applications for manufacturing could also increase productivity and relieve mental workload. However, some researchers (Tang *et al.* 2003) found that while offering benefits, overlaying the information could also distract the attention of the worker from the task. Therefore, special emphasis should be on how and when the information is overlaid. This requires that the UI would be non-intrusive and context-aware. Even so, AR offers high flexibility and adaptability features, and these have been well recognized and needed for the development of the modern manufacturing industry.

3.2.3 Games and Entertainment

The primary purpose for playing games is to train and stimulate activities. This can be observed not only in humans but also in the behaviour of the animals. The history of games can be traced back 5000 years when the Chinese invented the war game, the so-called Wei-Hai ("encirclement") (Smith 1998) — now called Go, for entertainment purposes (Smith 1999). Since then, many changes and challenges have been cultivated in order to preserve and encourage fun, entertainment, educational, collaboration, as well as the exercise and training aspects of games. Recent data about games and entertainment (ESA 2004) showed that there is great potential for the growth of the games sector. As some have stated (Bendas & Myllyaho 2002), almost every information device is a viable environment for games, from desktop computers to mobile phones. Moreover, games could be stand-alone or collaborative. Mobile and portable devices would become more popular if they were to incorporate more games. Unfortunately, due to the physical restriction of these devices (like the size of the screen and the processing power) games are only slowly starting to migrate towards mobility.

Games are not the only thing attractive to people. Other entertainment activities could be equally attractive, like watching a movie or a play. These activities are also hard to

access from current mobile platforms (due to viewing capabilities and the screen size). Augmented reality could provide more interaction and visual space allowing them to migrate to mobile platforms. An interesting combination is the use of AR to interact with actors. In their approach (Cheok *et al.* 2002), the user could communicate and interact with virtual actors (Fig. 41 on page 108).

Commercialisation of some devices used in AR systems (like the Sony Glasstron) already arrive with applications supporting movies, entertaining or private TV – a more traditional application that takes advantage of the new technologies. However, the combination of AR with games and entertainment would be beneficial (Starner *et al.* 2000b) and it should increase the acceptance and impact of both fields.

3.2.4 Sport and recreation

Sport and recreation, even if challenging, are sometimes not so entertaining. While they contribute to the health, many find it hard to exercise because of the monotony of the repetitive actions done in a non-interactive environment. Watching movies or listening to music helps. A better approach should include interaction. Systems like the Ping-pong plus (Ishii *et al.* 1999) demonstrated how the new technology could enrich entertainment of old games (i.e., the ping-pong). New research should uncover more applications and eventually invent new games combining the new technologies and the practicality of physical exercise.

Sport and recreation activities can benefit even more from the use of MARS, particularly those that take place outdoors. Displaying health information and the functions of the body in real time could help better adjust the quality of the training. Moreover, coaches could get an overview of the condition of athletes and adjust training programs accordingly. The augmentation of the view of the sportive could provide unobtrusive access to such information without having to interrupt the activities (as opposed to placing the screen of the monitor in front of the eye).

For indoor use, such systems could provide the user of a sports appliance with some interactive actions. For example, motivation is possible when there is a group of people, but collaboration and socializing is hard to achieve in an individual training programme. Physical activities that are performed over a distance could become more appealing (Mueller *et al.* 2003) since the missing aspect of inter-human communication could be provided. Augmented reality systems could make an important contribution to this area by overlaying and enriching the interaction as they are light, easy to set and more individualistic (Szalavári *et al.* 1998).

3.2.5 Medicine

The field of medicine could also benefit from the use of AR based UIs. An application with immediate possibility is in assisting surgeons during an operation (Fig. 16) by

overlaying live information about the state of the patient, or other views like information from an x-ray, ultrasound (Stetten *et al.* 2001) or a computer aided microscope (Birkfellner *et al.* 2000, Figl *et al.* 2001). Such an application would require very accurate registration so that the information overlaps correctly on the body of the patient. On the other hand, such a system may not be required to be mobile (State *et al.* 2001).



Fig. 16. Surgeon assisted by AR system (State *et al.* 2001). The system requires accurate registration.

Mobile applications of AR UIs in medicine are more likely to be used for activities like nursing, home visiting, telemedicine. For example, a nurse could receive information about a patient via a communication link and also obtain remotely provided instructions from a doctor on where and how to assist a patient (Umeda *et al.* 2000). The doctor could have access to what the user sees through a camera attached to the glasses of the user.

In emergency cases or in military applications, future application of the MARS could enhance the abilities of the medical personnel by providing information about location and status of the wounded.

3.2.6 Tourism

Another leisure activity that could benefit from the use of MARS is tourism. Instead of using the interface just for navigation, the system could provide a user with other information, like the history of a place. An application illustrating this point is when a user, instead of looking on a schema describing some ancient ruins of a site, could instead be provided, by the system, with a view of the site augmented with the virtual walls (Stricker & Kettenbach 2001) and other architectural information (like changes that occurred over time, etc.). Other applications could inform the traveller about the restaurants (Fig. 17) or nearby hotels, relieving the trouble of exploring an unknown crowded urban environment (Feiner *et al.* 1997).



Fig. 17. Navigation in town assisted by Cyphone mediaphone see-through glasses (Pyssysalo *et al.* 2000). This is a typical example of how AR can contribute to better inform the tourist or traveller in a foreign environment (courtesy Cyphone project).

3.2.7 Architecture

Architecture is another field that could benefit from AR (Webster *et al.* 1996, Tripathi 2000). The benefits could be not only in displaying on-site information about a new design but also by displaying information about the building, its maintenance or repairing of sites. Using Mobile Augmented Reality (MAR), an architect can receive real time data about the progress of the construction of a building or other information required (Fig. 45 on page 114).

Helping the interior design architect is another possible application for the MARS. The system could help the designer to visualize the setting of furniture and share the view with colleagues at a remote office (Tripathi 2000).

3.2.8 Collaboration

People interact in many ways in working places. These Human-Human interactions – called collaboration – are moving towards computer supported mediation, that has been called Computer Supported Collaborative Work (CSCW). Augmented reality methods are valuable for such tasks since they address two major issues in CSCW – enhance the reality and support the continuity (Billinghurst & Kato 1999). In a classical collaboration scenario, the participants would be able to seamlessly shift the view from the working table (or shared space) to personal communication (face-to-face conversation). Using AR

methods, the users may be able to experience the same workplace feeling and continuously shift between the shared space and the person-to-person conversation, in the same manner that they experience in classical collaboration, with the observation that in this case the other persons could be virtual and remote.

Such tools for collaboration are already available today, but they are mostly tied to fixed computer infrastructures (desktops). In 1998, in their paper (Luff & Heath 1998), Luff and Heath noted that the features of mobile systems are neither present nor exploited by the new tools in computer support for collaborative work. In many working activities, the collaboration tools have to be flexible and portable.

While AR enriches the interaction of a CSCW system, some people (Kato *et al.* 2000) have constrained the collaboration environment by tightening the interaction to some fixed object, in this case the table. Removing these constraints could provide the collaboration tools with greater flexibility and a wider application area (Reitmayr & Schmalstieg 2001).

Examples of using MAR in collaboration include indoor and outdoor meetings, mobile shared-information spaces (Fig. 43 on page 110), telepresence work assistance and outdoor support for collaborative work (e.g., of two architects sharing their design on a site).

3.2.9 Military

The military has had AR as a research target for a long time, from HMDs for pilots (Kalawsky 1993c) to MARS for training and simulation support (Piekarski *et al.* 1999). Some of the applications used by the military and based on AR systems are concerned with real time information updates (Julier *et al.* 2000), navigation support, training, simulations, reconnaissance, building or terrain information support, and so on.

Simulations and modelling are important for training and military applications. Augmented reality can enhance military simulations through the integration of virtual objects on the training ground in real time (Fig. 18). It is important to note that even though governments are spending a lot of money on researching applications of AR, the best results are expected to come in the future from the academic and commercial areas (Smith 1998, Smith 1999).

MARS could provide more information and enhance the cognitive and coordination capacities of combatants, not only on the training ground but also in real situations. As the director of the Virtual Reality Laboratory at the Naval Research Laboratory¹, Lawrence J. Rosenblum said: “The war fighter of the future will have to work in an environment where there may be no signage, and enemy forces are all around. Using AR to empower dismounted war fighters and to coordinate information between them and their command centers could be crucial for survival”.

¹ Naval Research Lab (sic) is part of Information Technology Division and it can be found at the following web address: <http://www.itd.nrl.navy.mil>



Fig. 18. Real world and virtual world. On the left, the user can see through the AR glasses the virtual helicopter. On the right, the view rendered for the user and the virtual helicopter (pictures from (Piekarski *et al.* 1999)).

3.2.10 Business and brokerage

As the market of mobile phones initially targeted the selling of devices to the business community, doing the same for MAR devices could also make it acceptable quicker. Modern business people require more information access from various places and personal information management combined with a MARS could fulfil their needs of accessing information any time, anywhere (Antoniac & Pulli 2000, Antoniac 2002). There are already some companies that offer specialized solutions or kits for AR (like Shared Reality¹ or TriSen²). The services and applications offered for a business person could vary from browsing the stock exchange to accessing a virtual office and sharing documents on the fly. Such an infrastructure could also provide for collaborative work (Reitmayr & Schmalstieg 2001) and the access to virtual spaces in which multidisciplinary teams would be able to work over VR models. Using an AR interface that extends the view, is mobile and can handle more interaction, would help the user to achieve faster results and at any time that the work requires.

3.3 Challenges and Test-beds

A popular question researchers are asked is whether AR offers practical benefits. One answer is to try it out, but current platforms and technologies are not able to fill all the

¹ Shared Reality is available at: <http://www.shared-reality.com>

² TriSen is available at: <http://www.trisen.com>

criteria defining a complete AR system. The best is to provide the users with a testing platform for AR.

Several researchers have addressed the issues of designing a test-bed for AR (Behringer *et al.* 2000, Sauer *et al.* 2000). The problems when building such a system arise from both the hardware and software available. In order to provide a test-bed, it is necessary to implement all the specifications that come with an AR system. The system should properly handle the calibration, tracking (registration), occlusion and the real time augmentation.

When adding the requirements set by a wearable or MARS, the challenges for the test-bed are even greater. Such a system should also handle user registration (for the navigation), it should be smaller, it should be wearable or handheld and it should operate continuously (as in seamless access to resources and fast power up time).

The challenges are still ahead for AR before it can be made available for commercial implementations. Even with recent advances, some questions remain open, like the size of the system (smaller is required) and processing power. For a real AR system, tracking (registration) at high rates and with good accuracy are essential.

An AR based UI test-bed faces even greater challenges. It should handle the interaction in a real-VE without confusing or disturbing the user movement. Such a system was under consideration when designing the prototypes (Section 8.1.6 on page 124). The prototypes had to have good registration, be mobile and ergonomic.

Even with limited implementation, the present chapter demonstrates the vast area of applications in which AR could impact. From civil to military and from engineering to leisure activities, the need for computer generated images to augment the view of the user are arising. Such environments, in order to become more useful, should include a UI that will emphasise its features. New designs should look away from the desktop paradigm and orientate towards a more flexible interaction available any time and anywhere. The mobility feature of the system should provide the user with more alternatives for where and how to use information technology. The next chapters will highlight some results of this research.

4 Experiments and Design Problems

Based on the discussion of the previous chapter, a basic requirement of a mobile system is that while being small in order to be more portable it needs to allow better interaction. This chapter will explore the experiments for implementing a better UI for mobile devices. The requirement for better interaction was focussed on finding the ways to extend the display. It began by exploring the method the information was arranged on the screen. Later, through experiments and after new iterations, from the simple idea of a mobile interface, a new and complex UI for mobile IAs was proposed. The described incremental steps also contributed to a better understanding of the usability and requirements for UIs of mobile systems. At the end of the chapter is a list of requirements and their implications are proposed, which will lead to the next chapter that will introduce the new concept of sign language (MARISIL) and its evaluation.

4.1 Dense Displays

In order to find a way to overcome the lack of screen space, one hypothesis was that by using a special layout for the information displayed, the interface “compresses”. The approach here was to “iconize” some common words and use a tabular layout. Others have placed more emphasis on sound-enhancement (Brewster 2002), semitransparent widgets like buttons (Kamba *et al.* 1996), transparent layered displays (Harrison *et al.* 1995) or toolglass and magic lenses (Bier *et al.* 1993). The following sub-sections describe the experiments carried out to test the hypothesis and the results that were obtained. Even though the experiments are little related to the subject of this dissertation, their results formed part of the basis of the present work.

4.1.1 Tabular Interfaces for World Wide Web

The first experiment dealt with the ability to layout information in a tabular manner that would allow faster and better access to data. The meaning of the word “tabular” concerns the way the information is laid on a display so that the user is able to read it.

The experiment started in 1998, as a means to serve a large community of researchers in the field of Concurrent Engineering. The research task concerned the access and dissemination of a database of people. The best solution at that time was to develop a tool that could be accessed from the web and be available to everyone to browse the database. The problem was on how to design such a tool and how to give access to it independently of the platform.

At that time, Sun Microsystem was making steps towards Java™ technologies leading to the decision to select the Java platform (at that time already version 1.2) as the solution for solving the portability problem. The main concern remained as to how to lay the information in a compressed format. The database contained many links and a categorisation of these was required. Fortunately, a special task in the project dealt with the categorisation and taxonomy, so the research focused on how to develop the tool (Kerttula *et al.* 1998) for accessing the information as fast and ergonomically as possible.

The database contained several fields for browsing and searching. Many argued that the best way to access the data was to present it as a list of items. However, listing many records that contain separated information is hard to read and it is even harder to find data that a person is interested in without reading the whole list. A more optimal way of presenting the data would be to fit it into a table that could be then reordered based on the columns – each column containing a category.

Tables are a method of presenting or arranging the information in a tabular form. This arrangement provides the information in a more condensed format, especially when using databases (Rao & Card 1994). While lists can only present successive items, tables have the property to split the information into rows and columns, hence improving the readability and compaction of the content.

In a digital form, tables can be very useful since they can rearrange the information contained in them dynamically. For example, by clicking on a row heading, the user could rearrange/resort the information from that row in an ascendant/descendant manner.

Moreover, by only showing the most relevant items (sorted by the rows) the user would only see a segment of the whole, something which is more relevant to the user's interest. Doing this would speed up the search for information. The conclusion of the experiment was that, by using tables to contain representation of the information with dynamic resorting of rows and browsing only a small segment of the database, it increased the browsing speed of a user. This conclusion was deduced from user feedback over a couple of years (between 1998 and 2000) but no serious quantitative measurement was possible as the tool was only available at a remote and inaccessible site. However, the conclusion that the tabular interface could perform well on small mobile devices was also confirmed by other researchers (Terveen *et al.* 2002) and was used as the hypothesis for the next experiment that dealt with media-phones (following section).

After some tests and several presentations, the tabular representation of the data with the rows filled with QFD symbols (Akao 1990, Day 1993) was able to comply with the

requirements for compact representation and usability (at that time generic browsers were running on 800x600 pixels screen resolution). Fig. 19 represents the layout and how the database displayed on a web page.

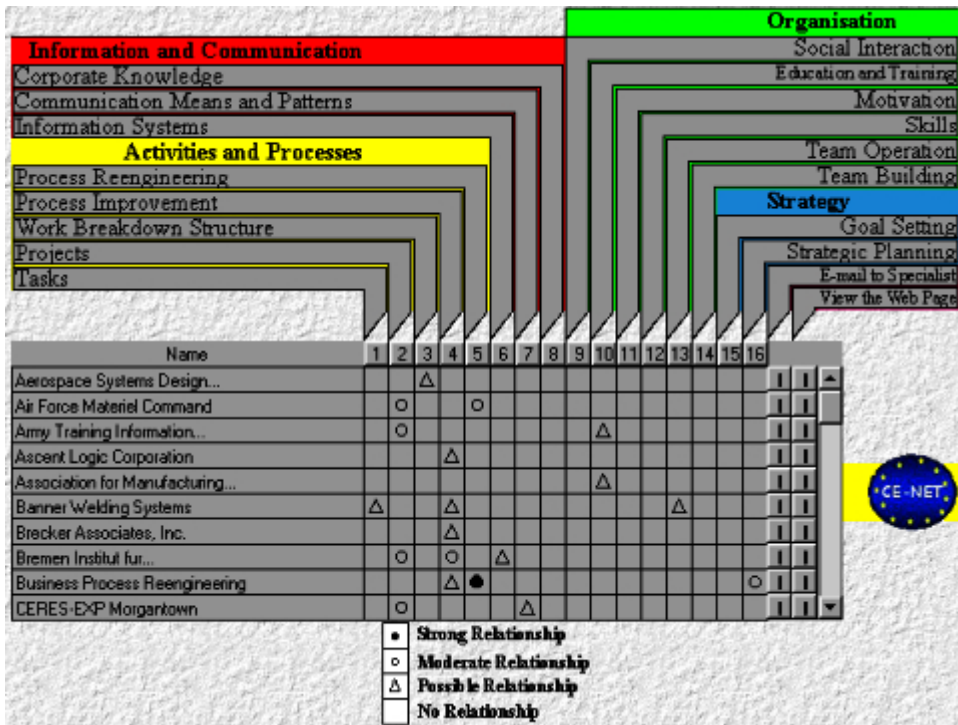


Fig. 19. CE-NET Database Viewer ¹.

The research, done for the Concurrent Engineering – Network of Excellence (CE-NET) project, was very well received. A continuation of the work was done for a “Who is Who in CE” task in a new project and it is still being used successfully when a small database with categorised information is needed.

What was highly valuable was the idea that when using a table the information layout is available in a more compact form. From this fact another project benefited later. This time, the focus of the project was on future media-phones, a subject that is closer to the present work.

¹ Available on the web on a backup server at: <http://www.tol.oulu.fi/~peter/work/ce-net/Information/taxonomy>

4.1.2 Tabular Interfaces for Mobile Phones

The second experiment dealt with the same problem of fitting more information in a small screen space. The difference was that, this time the platform had more restrictions in terms of size. The screen size of mobile phone is smaller compared with the one available on a desktop computer.

The experiment was done for the Cyphone project (Personal Virtual Services Based on Picocellular Networks joint project funded by Technology Development Center of Finland - TEKES) at the University of Oulu. The project had the requirement to fit location information into a small screen area. The device used was a WAP¹-compatible phone. The location information was personalised for the user and displayed, as in the previous project, in a table using QFD symbols (Akao 1990, Day 1993). In order to fit the information to a small screen, such as a mobile phone screen, the compromise was to pick the correct number of columns that the user could browse simultaneously. The table could also allow easy reordering of information and scrolling. Another interesting result was the impact of QFD symbols versus other icons. The problem was to choose between three simple symbols and a more flexible representation of the quantity (like bars, disks, etc). Finally, the choice was to keep the first schema, since it was easy for the user to remember the symbols and the accuracy of the representation was not so important (the bars were able to show the quantity more linearly, i.e., from 0 to 100%, while with the QFD symbols the information was split into 4 classes 0-25, 26-50 represented by a triangle, 51-75 represented by a circle and 75-100 represented by a disk). Fig. 20 is a snapshot of the working prototype taken from Nokia's WAP toolkit.



Fig. 20. Table interface running on WAP phone in a location-based application (Cyphone project).

While the results of the experiment were encouraging (Antoniatic *et al.* 2000), the limitations of the screen size were still encountered when dealing with more complicated

¹ WAP – Wireless Application Protocol is now an open international standard for applications running on wireless communication devices, e.g., Internet access from a mobile phone.

information. However, the benefit of using tabular displaying of information was acknowledged. Compacting the information and the increase in the speed of browsing through volumes of data that could be categorised, were observed. Moreover, the tabular display of items could easily become a switchboard. Using tabular interfaces was found to be important for speed and usability when it came to presenting categorised data.

The research from the CE-NET (ESPRIT Project No 25946) and Cyphone Project cleared the path towards a smaller canvas for displaying information. Unfortunately, the physical dimension of the screen always restricted the size and resolution of the display. This constraint was “food for thought” and it led to the idea of using some kind of virtual panel that could be extended as much as possible. The idea materialized further and envisaged the experiments using a virtualized and extensible screen.

4.2 Extended Displays

While compacting the information provided more speed in browsing data, the size of display still reduced the amount of interaction available within mobile devices. The problem of extending the display while maintaining the requirement, in terms of size of the device, was slightly contradictory. It was logically impossible when thinking of physical restrictions, but it was realisable if implemented in a VE. The problem with the VE was that the user would be unable to see the surroundings. The compromise was, therefore, to allow the user to see the environment and only for the interface to be virtual. This was called as “virtualizing” the interface.

Virtualizing the interface had an impact on other requirements. While physical interfaces could be seen by others when operated, for a virtualized interface, only the user could view and know what was being operated. Others could deduce the results of a user’s actions by tracking the user’s movements (like stealing the password by watching the keys pressed on the keyboard), but they could not see the interface. This implied a higher privacy of use, an important feature knowing that mobile IAs are used in the presence of other people.

Another important implication is that using a “virtual” interface would mean that the mechanical problems of the physical interface disappeared. Mechanical faults occur more often than electronic ones, and even though the present implementations of the virtual interfaces are still error prone, they could be better by being non-mechanical.

The following subsections describe the experiments to “virtualize” the interface and hence extend the display.

4.2.1 Augmented View

Virtualizing the interface was an important step in this research. However, to find the methods to achieve it were still unclear. Some ideas came from the Cyphone project. One of the tasks was to design a navigation system that would take advantage of a location

based service. The Cyphone device included a pair of video see-through binoculars through which the user viewed the world. From the data provided by a GPS device and/or bluetooth based trackers, the system was able to calculate and provide the user with some sense of navigation. When the view of the user was overlaid with synthetic arrows, the device augmented the reality surrounding the user. This technique very well satisfied the requirements to extend the screen size while keeping the physical size of the screen small.

Using AR, the user was able to see as much information as the system was able to provide. While using the video see-through system (Fig. 21) the limitation was only in the FOV of the device, but the user was able to look around, move and turn. However, there was a limitation: the user had to keep the binoculars in front of the eyes. In order to provide the user with even more freedom, another idea suggested was to use optical see-through glasses.



Fig. 21. Augmented view of a virtual meeting. Notice the participants' pictures from the binoculars' view.

Another technology considered in the project was the GPS. The GPS was supposed to contribute to outdoor navigation and registration of the system in a travel example (scenario of a business person travelling to a remote and unknown location). A GPS device uses the position of the satellites (via radio link) and is based on a triangulation process to provide the position of the system. Unfortunately, the accuracy of the GPS devices was 30 meters compared to 10 meters that is available today. This was caused by the Selective Availability or SA¹ feature that rendered the GPS data unusable. The outdoor registration was changed to a more robust implementation by using the Differential GPS (DGPS). The accuracy of the DGPS in the outdoor environment is

¹ At that time the US military had a feature called Selective Availability (SA) that was causing a random timing error in the GPS signals. This intentional miscalculation prevented the commercial GPS devices from obtaining accuracy any better than 30-100 meters. After 2nd of May 2000 the feature was turned off and now average accuracy is 10 meters outdoors.

around 1-2 meters (Pyssysalo *et al.* 2000). Unfortunately, for indoor use, the DGPS loses accuracy dramatically. Hence, for indoor registration the device had to use a more robust implementation like Bluetooth, local sensors and video scene recognition.

Sadly, the optical see-through devices available at that time (late 1999) were very expensive to build. Those that were available on the market were of poor quality. Later, with the introduction of the Sony Glasstron (more information on glasses are available in Table 10 on page 118), the research was re-started, now having the target to find new applications for optical see-through displays that would lead to the specifications of the MARISIL interaction and the building of the HandSmart prototype. However, an important achievement was the discovery of the method to include virtual objects in the real world (as the windows with the participants pictures in Fig. 21).

4.2.2 MARISIL and HandSmart Introduction

Mobile Augmented Reality Interface Sign Interpretation Language is the base of the specifications for a mobile UI based on AR interaction. The language specifications need a hardware platform to support them and hence the introduction of HandSmart interface. The connotation of the words in the name HandSmart emerged from the idea of using the hands in a more intuitive way when interacting with the computer. Every computer user utilises the hands to type or to move the mouse. The question was: what if the hands themselves could act as the interface? This kind of interface in which the hands are part of the interface was named as the HandSmart interface. The language description or how the hands are used in order to achieve from the HandSmart interface the desired results in the form of a computer interaction was called MARISIL.

The idea to build such an interface came when the research was struggling to find a way to increase the size of the display when using a mobile device while keeping the device size as minimal as possible. At that time, VR was becoming popular since video cards and commercially available systems were available at lower prices. The obvious approach was to use more the outcome of VEs in the hope of unblocking the physical restrictions of real displays. As the previous experiment showed, AR had the potential to combine virtual objects in the real world, allowing the user to operate in a normal environment. From the use of AR and having the table interface in mind, the next step concerned the design of the UI. The interface idea, as it is today, came during the research done in the Paula Project¹. At that time, the focus was on how to have tele-reality and virtual meetings. During some brainstorming meetings, the idea to use video recognition in combination with AR and to overlay the hands of the user with some kind of virtual interface developed. The main process dealt with recognizing the fingers of the user's hands, overlaying them with a table (either a mobile phone keypad with numbers and letters or a table with information) and tracking the pointing finger of the other hand when laid over the panel hand. The panel hand is the hand being overridden, while the pointing hand is the other hand that acts as the input. In Chapter 5 which describes the

¹ See the <http://paula.oulu.fi> for more information.

sign language, a full description of the method is given. The language was called MARISIL while the system using such sign language was called HandSmart as the interaction was done via the hands of the user, and Smart since the hands are used as a panel and not other physical objects, so they became “smarter” and part of the interface.

The HandSmart interface consists of the recognition system (a video camera with high colour resolution) and see-through glasses (Fig. 22). The see-through glasses received the information from the processing unit (via cable Fig. 22, item 4) that was able to recognize and register based on the camera (Fig. 22, item 3) attached to the glasses (Fig. 22, item 2).

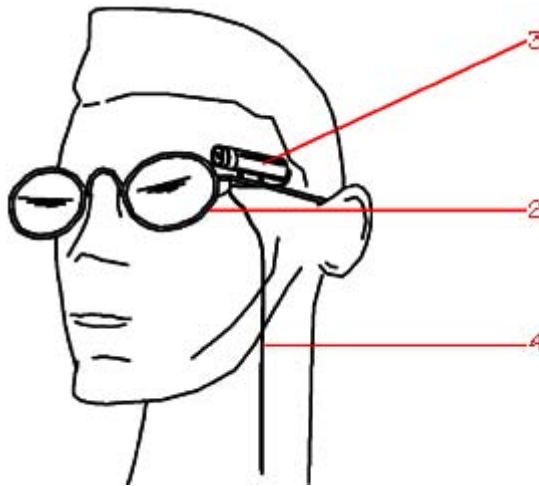


Fig. 22. HandSmart system containing the camera (3), the see-through glasses (2) and the link or cable to the processing unit (4) (picture from the US patent (Pulli & Antoniac 2004), see also Appendix).

After generic specification of the device and a brief description of the interaction, the next challenge and experiment dealt with understanding the interface and improving the interaction. Moreover, while in the process of writing the patent specification, the need for a more intuitive presentation of the idea to the patent attorney led to the writing of a scenario and creation of a demo movie.

4.2.3 MARISIL Scenario and Movie Creation

While the idea of creating a movie can scare some engineers, acting is able to highlight many usability requirements (Nielsen 1995). Hence, the inclusion of the description of the movie creation in this document.

The scenario of the movie dealt with placing a call, making a mistake in typing the number, talking and hanging up the phone. The purpose was to show how the MARISIL interaction with HandSmart is similar with the interaction of placing of a phone call from

a plain old telephone system (POTS). The movie pictured the user's hand and how the number was dialed. In the post processing, the movie was supposed to include the augmented view of the user's hand and the interface (Fig. 23).



Fig. 23. Dialing the number. A shot from the movie.

The recording of the movie started in October 1999 in a studio specially arranged at the University of Oulu. The camera used to capture the movie was a normal Sony Digital 8 HandyCam (DCR TR-7000E PAL) that allowed uploading and digitising the raw movie faster. The raw movie consisted of 6 hours of video. From this, the final movie was edited to just 1 minute. The same camera was later used in building the first prototype. The work in the studio showed the importance of lighting for the recognition process.

Several problems were identified from the process of creating the movie. Some problems that were addressed were the technical settings, while others mostly concerned the usability of the interface. The first problem identified was the discomfort of keeping the hand still in front of the camera for long periods. Another problem was of keeping it high up and looking down at it. Better implementations should consider using a fish-eye lens (lenses that operate similarly as omni-directional cameras) that enables an image-grabbing camera with a greater FOV. This should also encourage the user to use the hands without looking at them (and hence without seeing the overlay) and to keep them in a lower but in a more comfortable position.

While filming the hands, the settings of the studio allowed the use of multiple light sources that permitted the removal of shadows. Such settings are not available in natural light or in a normal environment. Proper implementation should consider the calibration of the system to handle the recognition of the hand in various lighting situations.

Another observation came from the post-processing work. While the movie was made with the camera fixed on a tripod, hence, maintaining a still position for the camera, the hands from the movie were shaking. This created an annoying vibration or movement of the interface picture. This movement could be accentuated, particularly when the camera was mounted on a user's head (versus the tripod when filming the movie). Using some

filters (a Kalman filter could be used to estimate, interpolate and smooth the motion) would be necessary in order to remove this shaking of the displayed interface.

A good observation was that the hands were capable of more actions than just pointing or pressing some buttons or icons. The hands can change shape and hence change the input type of the interface or even input data. Even though the current technology and the prototype work were not capable of implementing or providing such features, they are worth mentioning as part of the observations on the movie.

Another observation was that while using the interface, the user took advantage of the tactile feedback from the pressing and touching of the hands, therefore, increasing the repeatability of the actions (or what is called spatial memory).

After the realization of the movie, the potential of implementing such interaction for mobile interfaces became clearer. The problem of extending the screen space of mobile devices became a more understandable solution.

4.3 Summary

This chapter presented observations on the experiments carried out. Some basic requirements can be extracted for an interface that could provide mobile IAs with improved interaction. The most significant requirements listed were:

- *Extended screen size*: As the technologies progress, it will be possible to deliver more processing power to mobile devices, and there will also be an increase in the level of applications running on these devices. Applications would require more display size in order to provide a better interface for the users. The lack of physical size of the display could be corrected by various techniques (like icons, sound-enhancement (Brewster 2002), semitransparent widgets like buttons (Kamba *et al.* 1996), transparent layered displays (Harrison *et al.* 1995) or toolglass and magic lenses (Bier *et al.* 1993)), but the ideal solution would, however, be to extend the screen size.
- *Tabular layout*: Using this layout provided the information in a more condensed form, especially when using databases (Rao & Card 1994). While lists can only present successive items, tables have the property to splitting the information into rows and columns, hence improving the readability and compacting the content. Mobile devices already use such access to information (when opening the phone agenda, the calendar or the to-do list).
- *Use of icons*: Icons are also a way of compacting the information. Icons have been used in the past in various interfaces (starting with the WIMP paradigm of desktop computers) and, therefore, its importance to be included in the list.
- *Virtualizing*: By virtualizing the interface, the user gains access to a “virtually infinite” display. The use of VEs in which artificial objects are generated could provide the user with access to any size of display. The problem would be on how to access the virtual objects in the real world, a problem solved by the use of a MR technique like AR.
- *Optical see-through*: The use of AR implies the use of certain devices to enhance the user’s environment. One technique was to use video see-through. This has the

disadvantage of blocking some of the user's FOV. A better implementation would be to use optical see-through glasses.

- *Tactile feedback*: Often, when using VEs, the user misses the tactile feedback that is available from interacting with real objects. As the tactile feedback could add a dimension to the person's understanding of the surroundings and even more, it could enhance the memory of a person (what is called spatial memory). A good system should consider including tactile feedback as a requirement.

The scenario of a user operating in such an environment identified the challenges of implementing such a system. The lighting problem, the movement of the user and the fatigue of using the hands upheld in front for long periods have been identified as possible problems.

Despite the challenges described, the experiments also led to some important discoveries. It was found that while using the proposed interface and by using the hands to operate the interface, the user had a spatial memory that enhanced his capabilities to memorise and repeat an action. Moreover, the user could use the hands shape to change the state of the interface (as in a sign language). Virtualizing the interface provided the user with better privacy, a more flexible interface, a non-physical and hence non-mechanical interface, and a more easy to learn interface which has the potential to be used almost everywhere.

These aspects are discussed in the next chapter that specifies the basic constructs for the sign language that would allow a user to operate and interact with such interfaces.

5 Interaction Specifications for MARISIL

Humans interact in many ways in order to communicate and exchange ideas. A simple, but quite efficient, way to express ideas or feelings is by using the body or hands. The body or sign language expressed during dialogue could contribute the missing emphasis or add meaning to a conversation. How to use such communication, which is present in many cultures, with computers is the subject of this chapter. While the subject of many research projects (Wu & Huang 1999b), body and sign languages were not explored enough from the natural point of view. Some people have addressed the issue based on standard sign language used for disabled (Kuroda *et al.* 1998); others were looking more into gesture recognition (Kjeldsen & Kender 1996, Wu & Huang 1999b).

The research on future interaction methods for the 3D computer environment identified hand-gestures as a possible solution to current problems (Nam & Wohn 1996). A hand-gesture is *a motion or movement of the hands made to express or help express thought or to emphasize speech*. There are a large range of hand gestures which include gesticulation, language-like gestures and sign languages. The majority of the research on hand gestures is on gesticulation and sign languages. In the approach here, the gesture was meaningful only if augmented with synthetic information provided by the AR UI. This was done by overlaying the symbols on the hand and pointing at the symbols, something similar to pressing the buttons on a keyboard. In the case of using a more advanced system (such as when using the fish-eye lenses cameras that provided a larger FOV), the user could still interact with the system by learning the layout of the symbols and pointing at them without looking at them (in this case, the laying of the symbols on the hand was not done, as the user was not able to see the hand).

In the philology of gestures, Kendon (Kendon 1986) described them as: gesticulation, language-like gestures, pantomimes, emblems, and sign language. Sign languages include vocabulary and grammar and they might require special skills and learning. Some have also classified hand gestures (Wu & Huang 1999a), depending on their context of use, as: conversational, controlling, manipulative and communicative gestures. The conversational and communicative gestures are related mostly to the act of speech and speech emphasis, while the manipulative and controlling gestures support distant speechless communication. In the proposed classification, the complexity of the hand movements was used to categorise the gestures. Fig. 24 shows the proposed classification

of gestures used with vision based algorithms for hand movement recognition. From left to right on the *Hand Movements* scale, based on the complexity and learning process, the hand movements were less or more complex (an example of complex sign language is the American Sign Language for disabled, that takes time and certain training to learn).

A command-like gesture video-based recognition system would be easier to implement than the sign language video-based recognition system, due to the structural complexity of the sign language. By using only image-processing hardware that could video recognize hand gestures, the system was able to interact with the user. An example of use would be to command house appliances, like the fireplace or doors, to open or close by doing special gestures (Starner *et al.* 2000a).

The use of pointing gestures as an interaction technique is the focus of this work. Using pointing gestures, the user could control an information system (Fukushima *et al.* 2002). In Fukushima *et al.*, the system was immersed in a VE. A better approach would be to use AR and to overlay digital information in the real world so that the user could point and select the virtual objects (a technique called *Tangible Interface* is discussed in detail in Chapter 7.7 on page 112). The pointing gesture tracking combined with the AR synthetic generated images, was a more sophisticated approach to pointing gestures as the system worked only when combining the gestures with the synthetic image generated by the AR system that enhanced user's view. In the system used here, the table of Kato and co-authors (Kato *et al.* 2000) was substituted by the user's hand. Without the combination of the image overlaid on the user's hands and the gestures, the gesture alone had no meaning to the user and neither to the machine. While in other hand movement video-based recognition systems, the user was required to have a certain qualification or to complete a learning process (i.e., when learning the sign language), the AR combined with the pointing gesture was more simple and intuitive as the AR assisted the user with synthetic objects, and hence, it was easy to adopt. This alone made this novel approach more appealing for mass acceptance and a better user-friendly interface than common gesturing or sign language video-based recognition methods. The AR-command gesture (the pointing gesture in combination with the AR technique was called AR-command gesture) was richer in commands than the command-like gesture. This was because by combining AR with the gesture, a small variety of simple gestures could be turned into a large number of commands that resulted from adding the images on the commands (as done in overlaying virtual images on the real objects when using AR).

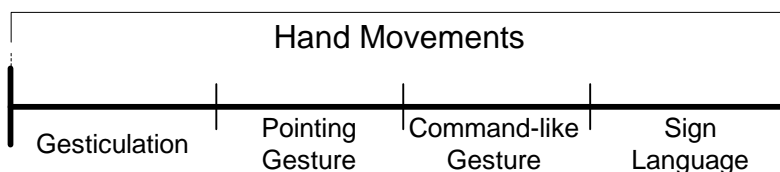


Fig. 24. Hand Movement representation, from primitive Gesticulation to complex Sign Language.

While gesticulation is a personal way of expression when talking, it is the least capable of transmitting information than any other hand movements (as the information is more

vaguely interpreted by the other party with this form of communication, without even taking into account the personal and cultural variations that are involved).

Using AR-command hand movements could improve the speed of learning the interaction gestures, while maintaining a low number of gestures needed to be learned or used by the users. The following sections propose a set of input gestures that could be used to introduce data by using a MARS. The gestures are defined so they keep highly intuitive gestures for interaction for the benefit of a faster input pace.

5.1 Foreword to Input Gestures

The input gestures were intended so that the interaction with the interface would provide enough input commands as required in a mobile phone device. This includes the ability to place a call, to write short messages and to browse the phone book. For this purpose, a *Core* set of gestures was identified: the way to introduce numbers, to browse the agenda and to place a phone call. Probably the most interesting observation at this stage was the one-to-one correspondence between the dial pad of a mobile phone and the number of natural partitions of the fingers (without the thumb, four fingers with three partitions per each finger).

The evolution of mobile phones into media phones led the research into looking for other possible uses of the interface and the expansion of the initial *Core* set into a more complete set (which is called the *Basic* set) that would allow applications like viewing a table, *undo*, *multiple selection* and *deselection*, *speed/fast key* and *volume/linear scale setting*.

The last extension of the proposed input gestures (the *Advanced* set) dealt with the language part that was based on gestures that were too complicated to be recognized by the video-based recognition system and that would also require more of a learning process by the user. Their purpose was to support the same level of interaction that was available with a desktop system. They included operations like mouse pointing (Quek 1996) and character recognition or handwriting, which were harder to implement with current video-based recognition systems.

The MARISIL proposed in the following section laid the framework for the interaction with a mobile device, allowing the user to have access to the same level of interaction with a video-based recognition system as with a current desktop computer. In this way, the input device became the user's hands and was, therefore, ubiquitous and easier to access than other input devices available currently for mobile systems (Section 1.3.1.1, page 27). Additionally, the user of a MARS could benefit from other enhancements, like memory and/or visual enhancements, allowing a growing number of applications to be deployed on the mobile system (Chapter 7, starting at page 101).

When specifying the MARISIL, the process had to also look at the anatomical constraints of human hands. In some cases, even though some gestures might look appropriate and logical, they may be impossible to achieve by some persons.

Much like the current mobile phones, in which the user can place a call and also write text, the proposed language enabled the switch between phone mode and IA mode. The

switch could be done either by the user (*selecting* from a list of menus what mode to operate, or by *closing* the current mode) or by the system, based on the context of use (when receiving a phone call, the mode would be switched automatically to the phone).

5.2 Core Input Gestures

The core gestures to interact with the interface were *opening* and *closing* the interface, *selecting* or *pointing* to a symbol, numbers, and characters overlaid on the hand and *placing* or *closing the call*.

In order to operate the interface, the user had to set the interface to the “enable mode”. To do that, the user had to place the hand used for interaction (called the palm panel) in front of the face. The palm panel had to be at a comfortable distance at which the user could read the text overlaid on it. The interface was *enabled* when the user could see the hand overlaid approximately perpendicular to the axis of view (glance of the user) as described in Fig. 25. By turning the palm away, the user could *disable* the interface. The video recognition worked based on a classifier identifying the shape of the palm (discussed in more detail in Section 8.2.3 on page 130). If the palm was not situated correctly, i.e., more or less perpendicular to the view axis of the camera (or the user’s eyesight per se) and vertical, then the interface would be disabled.

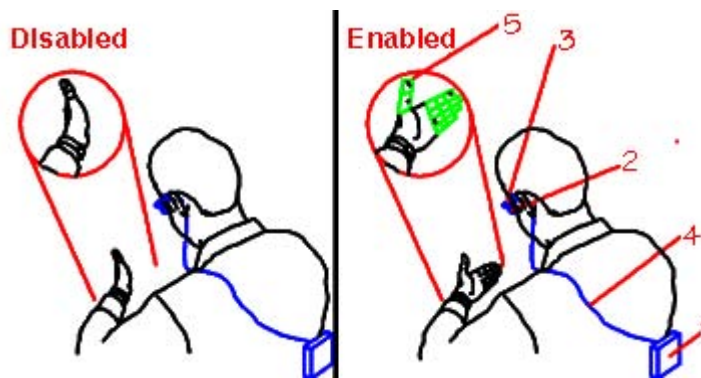


Fig. 25. User looking at his hand would enable the interface. User could also disable it with a gesture (left). The picture on the right shows the user, and the HandSmart appliance (1), the see-through glasses (2), the camera used by the recognition (3), some of the connection cable (4) and the interface seen by the user (5).

The video recognition process could identify the segment/region of the hand containing the item and the location of the pointer (the pointer finger from the other hand). Later, based on the pointer location and the “item” displayed on the hand, the gesture became the input. The “item” existed only in the AR world and its location depended on the hand location. Usually, the locations and the boundaries of the “items” should be defined by the natural partition of the fingers on the hand so that the user could identify the locations faster (a better example can be seen in Fig. 28 on page 84). Since they are visible only

when using the AR system, only the user of the system who wears the see-through glasses would be able to see them. The pointer (usually the pointer finger of the opposite hand) in the current implementation had a special marker attached to it. This marker — called the fiducial marker — had a colour that the image recognition process could distinguish quickly in the normal life environment (a colour that is not present too often in the surroundings of the user). This approach was accepted because with the current implementation of the system, the video recognition process failed to extract the pointing fingertip based only on the natural colour of the fingernail.

The user could *select* an “item” by placing the pointer over it for an extended time. To *unselect* the selected “item”, the user had to again place the pointer over it for an extended time. For example, if the user wanted to introduce text characters, by selecting the corresponding letter of the alphabet from the table (the interface had to be in the keyboard mode) the letter became available in the editor or word processor. The *unselect* function worked in certain modes, e.g., in the *information mode*.

To view the information from multiple pages, (like in *page up* or *page down*) the user had to place the pointer over the thumb of the canvas hand that had to be tagged with the *page up* or *page down* symbols (Fig. 26). If the user placed the pointer finger over the upper segment of the thumb that was held perpendicular to the palm (Fig. 26, item 13), the movement triggered *page up*. The *page down* command was activated when the pointer finger was over the base of the thumb (item 14, Fig. 26).

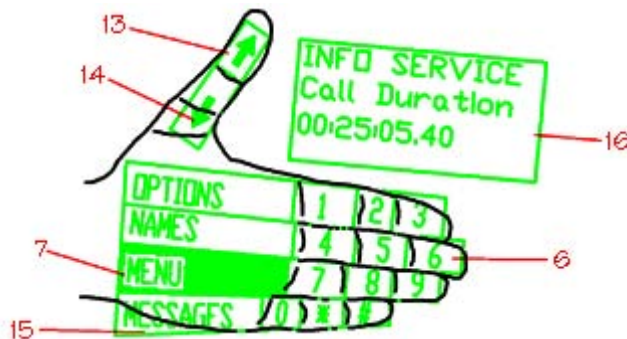


Fig. 26. Hand augmented with page up and down option. Figure shows normal items (6) or selected items (7), *page up* (13) and *page down* (14) icons. Placing the pointer on *page up* (13) the user browses up; if placed on *page down* (14) the user browses down. Figure also shows information panel (16) that is neither interactive nor tangible (is only virtual).

For *shifting* or *viewing left* and *right*, the user held the thumb parallel to other fingers. The user could use this feature only when the thumb was in the left-right shifting mode (Fig. 27). The user identified the right mode by viewing the icons/symbols displayed over the “items” at the base and upper parts of the thumb (Fig. 27, item 10).

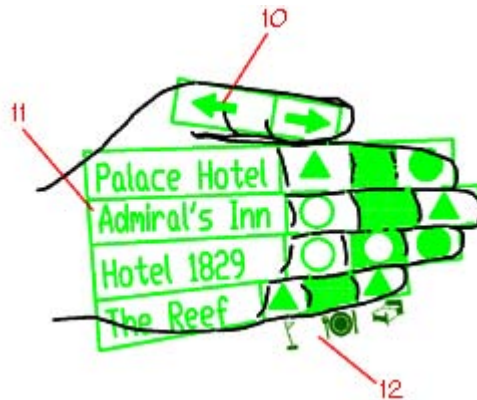


Fig. 27. Hand augmented with left and right option. The figure shows the left-right items (10), the normal items (11) and the selected column (12). Placing the pointer on left or right arrows (10) the user browses left or right. The figure also shows the icons (callout 12) that represent the table rows type.

Another *core* command was to *place* or *close a telephone call*. The hand and the thumb had to be in the *telephone mode* (with the thumb augmented with specific icons for this mode). This mode was selected by the user by picking it from a menu (like in Fig. 26, by pointing at the menu item activating the *telephone mode*) or it could be automatically enabled by the system, e.g., when the user received a phone call. In this mode, the user could use the natural partition segments of the fingers as the dial pad (Fig. 28). Each finger had overlaid on it several numbers and, when placing the pointer over one of them, the user could dial that number. Placing the pointer over the thumb's lower or upper part would cause the interface to *close* or *open a call*.

In some circumstances, where the privacy of the entered characters was required, a more secure mode of entering the information could be activated. As the user of the system was the only person that could see the interface (through the see-through glasses), by scrambling (changing) the order of the numbers overlaid on the palm panel, the user could confuse an onlooker from tracing the movement and hence deducing the input from the user's chain of selections. This gave a good sense of privacy when using the interface in public places.

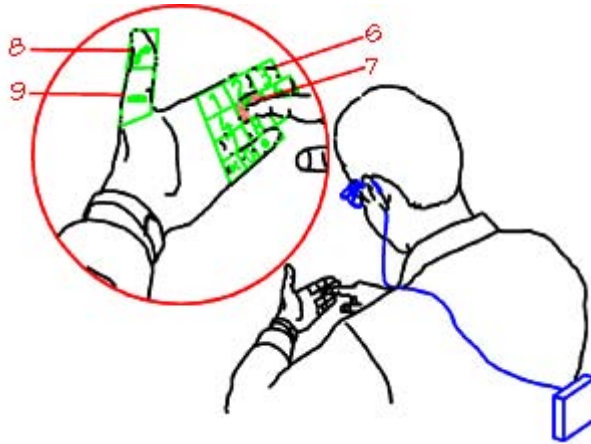


Fig. 28. Hand augmented as telephone. The figure shows the dial pad (6) selected/dialled number (7) and the place call (8) hang-up call (9) items. Placing pointer on call item (8) will start a call. The user can hang-up by placing the pointer over the hang-up item (9).

As observed during the production of the movie (Chapter 4.2.3), the hands could be used in a more advanced manner. Combining the shape of the hand with the augmented interface could provide faster access or shortcuts to menus or operations for the interface. The following section discusses the next level of gestures used for interaction by the HandSmart device.

5.3 Basic Input Gestures

The *basic* gestures to interact with the interface were a little less obvious than the *core* input gestures. They involved actions that were more complicated and required a more advanced recognition process. Their purpose was to improve the interaction and to speed up the input process. The main commands were: *multiple selection/deselection*, *undo*, *raw removal*, *fast key*, and *level/volume*.

The *multiple selection* operation was defined by the gesture of closing together the forefinger and the thumb of the pointing hand (not the panel hand) and placing them over an item. The “item” would be selected and included into the list of selected items.

The *multiple deselection* operation was the reverse of the *multiple selection* and consisted of placing the same fingers, the forefinger and the thumb, over the selected items and hence removing them from the selected items list.

The *undo* command was executed when the user was doing a fast *disable* and *enable* gesture as defined in the *core* input gestures. This consisted of removal and replacing of the panel hand in the viewing area. The interval time was calculated based on camera speed and user preferences (slower than camera processing frame rate and faster than the user preference).

An *alternative undo* command was defined by the gesture of placing the thumb on the palm, enabling the selection of the undo command. This was a more complicated action and some users could have difficulties in using it. An image grabbed from the movie describes more accurately how to select the *undo* command (Fig. 29).



Fig. 29. *Alternative undo* command operated only with the thumb of the panel hand. Image from the movie.

Removing a row operation was defined by folding over the corresponding finger of the row that was required to be removed. The user had to help the folding over finger with the pointer finger from the other hand, as shown in Fig. 30 (since some users had difficulties in folding over the middle finger without changing the shape of the hand).



Fig. 30. Hand augmented as browser. The user is removing a row from the list. Notice the help of the pointing finger from the other hand to correctly fold over the finger.

Fast key or *speed dial* was defined when the interface was in the phone mode and the user had defined an index of phone numbers that were assigned to correspond to a single digit number (as in current mobile phones). When placing the thumb of the palm panel over the natural segments of the fingers of the same hand, the interface would interpret it as selecting one of the numbers corresponding of the speed dial number. If the interface was in the browse mode, the operation could provide the user with access to the associated information corresponding to the item selected (one possible use could be to implement a help/info display of that item).

The command to *open* an item (for further actions, or advanced operations) was defined by the gesture of opening and closing the palm panel at a predefined rate. The rate had to be shorter than the rate for *closing* (as described next).

The command to *close* was defined by the gesture of opening and closing the palm panel at a longer predefined rate than the *opening* command. Both *open* and *close* commands were used for opening and closing, not only selected items but also interface modes, tables, menus, etc. One example is when the user was in *telephone mode*, placed the call, and wanted to switch to *information mode*. By doing a *close* operation, the user could switch to *information mode* (i.e., the user could close the phone mode causing the next mode to be available). If more than one mode was defined for the interface, the closing of one mode would automatically switch the interface to the next polling mode.

Finally, the important *basic* command was the *level setting* operation. To set the level, the user of the interface had to make the palm panel fold over into a fist shape so that the outside part plane was perpendicular to user's axis of view (Fig. 31). When doing so, the interface entered the *level mode*. The level mode was not in the polling list of interface modes like the *telephone mode* or *information mode*, and it could only be started by using this gesture. Once in *level mode*, the user could set the level for various variables, by pointing at one of the four reverse sides of the fingers plus the thumb from the palm

panel. One example was to set the volume of the ring tone or the volume of the speaker. Other uses were for setting the contrast of the screen or other values that required scaling.



Fig. 31. Hand augmented in level mode. The user is setting level 3 when pointing the finger at the corresponding finger for the level setting desired.

5.4 Advanced Input Gestures

As the core and basic input gestures are important in order to operate the interface in the same manner as current desktop interfaces, the advanced input gestures should be for the user to define.

An example of an advanced gesture is the *link* command. The user could close both palms as fists and link them together, generating a *link* command. Alternatively, if the user rotated the panel palm around the hand axis, this could be interpreted by the system as a *next menu* command.

Another example of an advanced gesture would be the *recognition of the characters* based on the tracking of the moving of the pointing finger over the palm. The next figure shows how the user enters the letter “A” by drawing a “V” upside down on the hand (this process is described in the evaluation chapter, Chapter 6).

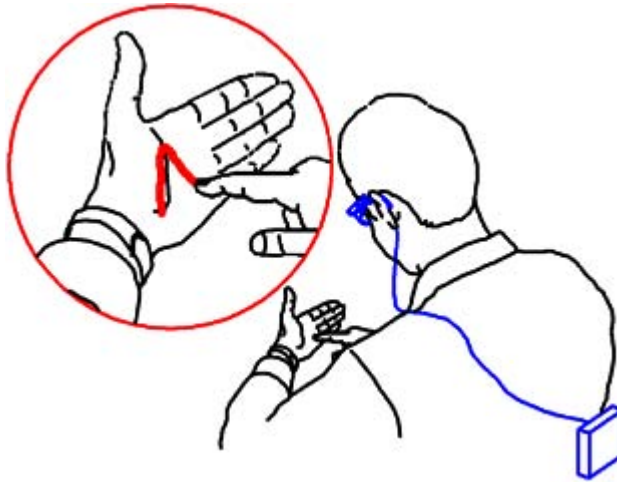


Fig. 32. Hand Augmented in writing mode. User can introduce letter by writing on his palm with the pointing finger.

Another advanced operation could be the implementation of such interactions as described by the FingerMouse system (Quek 1996). In such a case, the user could operate the interface in the similar way as when operating a mouse device.

This mode could include other ways of interaction based on gestures. It was not excluded that, in the future, by adding to the system a generic learning process for the video-based system, the user would be able to define a sign language. In this way, the user could access a custom set of commands.

Advanced input gestures require more accurate and advanced image recognition. The implementation of some advanced input interactions, as the *recognition of characters*, is difficult to achieve without implementing an accurate stereoscopic video-based system, since the detection of the touch of the hand is important for determining the character being input. Such a system would require more resources from the system (at least two cameras) and a higher resolution. Moreover, the operation should be carried out in real time, making the task even more difficult as the amount of data would be double that with the current single-mode camera system. Additionally, the generic process would be hard to implement with the current algorithm, as the shape of the hand is used by the system in the extraction process of the hand from the surrounding environment. To conclude, because of the complexity of the video-based recognition task, with all the current knowledge in artificial intelligence, heuristics and neural networks, it would be still hard to deploy the *advanced* input gestures on a portable device. Their inclusion in this work is purely informative and speculative.

5.5 Summary

This section specified the initial set of gestures that could be used to operate a MAR based system. The proposed interaction had the requirements of being intuitive and simple, hence being easy to learn and memorise by an inexperienced user. While some *advanced* gestures would be hard to implement with current technologies, the *basic* and *core* ones were experimented on a HandSmart system implemented with present commercial devices. The next chapter evaluates the proposed interaction and sees if it could provide a sufficient set of actions for the user of an advanced mobile phone.

6 Evaluating the MARISIL

This chapter presents the evaluation of the constructs specified in the previous chapter. The evaluation consisted of sample scenarios of using an advance media phone and comparison with the two types of interaction: the classical media phone versus the MARISIL operating on the proposed HandSmart device. As the proposed input method replaced the mechanical buttons of a classical mobile phone with virtual overlay images of buttons on the user's hand, the pursuit was a comparative analysis between the two input methods. The analysis confronted the proposed new interaction technique against the one currently available in media phones. The aim of the tests was not the usage, but rather the comparison of features in the respective UIs. The usability tests were not possible at this stage of implementation of the HandSmart prototype. However, in future implementations such tests would be required in order to improve the usability of the interface.

6.1 Features Analysis

The test aimed at comparing the features available in present media phones to those available with the proposed MAR interface based on the common actions of users of media phones. Based on personal experience and some reports on the use of mobile phones (Klockar *et al.* 2003), common actions (or functions) required from a media phone were identified. From these actions the more generic operations were selected, like: making a phone call (by entering the number or searching the contacts), writing text (present with functions like sending short messages, emails or entering a name) and the use of media features (like taking pictures or video). These three activities included all the interactions needed to operate a mobile device, like: input text, browsing, using media capabilities and receiving and placing phone calls. The next question focused on the interactions available within the suggested interaction technique in supporting the functions of the current media phones.

The reference for this analysis was a Nokia's 6600 (Fig. 33), since it is a media phone with an extended screen and included extended multimedia capabilities. The physical

characteristics are mostly the ones to support the multi-media capabilities as presented in Table 8.

Table 8. Comparison of physical characteristics between Nokia 6600 media phone and the latest prototype of HandSmart device (introduced later in Chapter 8 on page 116).

Characteristics	Mobile Device	
	Nokia 6600	HandSmart
Weight	125 g	90g camera+120g display+440g battery+900g Portable Computing Device≤1600g
Display Size	176x208x64k	Physical 832x624x3 (can be virtually ∞)
Video Capture	Capture, download, view, preview	Capture, download, view, preview, edit, real time recognition
Applications	Some, ported, Java MIDP	All possible as in PC today
Input methods	Phone Keypad	Phone like keypad; in the future hand recognition methods could be available

As can be seen from the table, the HandSmart prototype available is currently heavier than the media phone. This is obvious since the device includes see-through glasses with a powerful processing unit that requires more battery life. Moreover, the HandSmart physical characteristics belonged to the current mock-up or prototype that had not undergone a design rationalisation process involving optimisation for size or weight.












Another difference was in the display size. While the media phone had a small display size, it provided a better colour range. The current see-through glasses used for the HandSmart, due to price was bound to have lower colour range. However, due to the software implementation of the virtual display, the screen resolution was as large as the user could see (considering the user's FOV and the device capabilities). Moreover, if the HandSmart system had included a head tracker for the position of the head (compared with the body or maybe against the Earth magnetic field) the display would have virtually extended to cover all the surroundings of the user.



Fig. 33. Nokia 6600 The mediaphone's keyboard and joystick.

The next step was to analyse how the tasks carried out with the media phone (Fig. 33) were available on the proposed HandSmart device. Following is a comparative table (Table 9) between the phone's buttons and the MARISIL corresponding gesture.






Table 9 Comparison between Nokia's 6600 and MARISIL gestures.




Nokia 6600 button	MARISIL mode	MARISIL set	MARISIL gesture
 -Enter	Phone/Information	Core	Open
 -Back	Phone/Information	Core	Close
 - Select	Phone/Information	Core	Select
Numbers/Keys 0-9, a-z	Phone	Core	Select symbol from table
 - # sign	Phone	Core	Select symbol from table
 - * sign	Phone	Core	Select symbol from table
 - Left	Information	Core	Left
 - Right	Information	Core	Right
 - Centre	Phone/Information	Core	Open
 - Up	Information	Core	Page up
 - Down	Information	Core	Page down
 - Cancel	Phone/Information	Basic	Undo, Alternative Undo
Volume	Level setting	Basic	Level/Volume setting
T9 mode	Phone/Information	Core	Select symbol from table
		Advanced	Hand writing
N/A	Phone/Information	Basic	Multiple select/deselect

As a better representation of the table and based on the actions selected as generic operations (making a phone call, writing text and the use of media features–video and pictures) a set of four scenarios (examples) is narrated:

6.1.1 Example 1: Making a Phone Call

Here is a list of interactions as they are presented in the user's manual to use the Nokia 6600 (Fig. 33) for making the call. The manual could be found from www.nokia.com website. Only the basic interactions are presented, conference calls and more complicated menu interactions being irrelevant since the menu construction was not the purpose of this comparative analysis but rather the input mode. The present evaluation compared the keys required when placing a phone call with the Nokia 6600 with the interaction (or the access to the virtual keys) of the proposed language MARISIL. Here are the two possible interactions when placing a call:

1. NOKIA 6600 – Making a call when the number is known:
 - a) While the phone is in standby mode, the user keys in the phone number, including the area code. The interaction is to press  or  in order to move the cursor. If the user requires modifications in the number, pressing  will remove a number. For international calls, pressing twice  will generate the + sign for the international prefix (the + character replaces the international access code) and then the country code can be keyed in, the area code without 0, and the phone number (ITU-T 1995).
 - b) Pressing  will call the number.

- c) Pressing  will end the call (or will cancel the call attempt).
2. NOKIA 6600 – Making a call using the Contacts directory:
- a) The user has to open the Contacts directory by entering the Menu → Contacts.
 - b) To find a contact, the user has to scroll to the desired name, or key in the first letters of the name. The Search field opens automatically and matching contacts are listed.
 - c) Pressing  will start the call. If the contact has more than one phone number, scrolling to the number desired and pressing  will start the call.

When using MARISIL with a device like HandSmart the interactions described in the previous list are as follows:

3. MARISIL HandSmart – Making a call when the number is known:
- a) User will have to place the hand in front in order to operate. In phone mode (by default the hand is in phone mode), the user can key in the phone numbers by pressing the natural partition of the fingers (Fig. 28). Each finger has a number and when placing the pointer over the partitions (Fig. 34), the user can dial numbers.
 - b) Placing the pointer over the thumb's upper part (Fig. 34) will cause the interface to open the call.
 - c) Placing the pointer over the thumb's lower part (Fig. 34) will cause the interface to close the call.

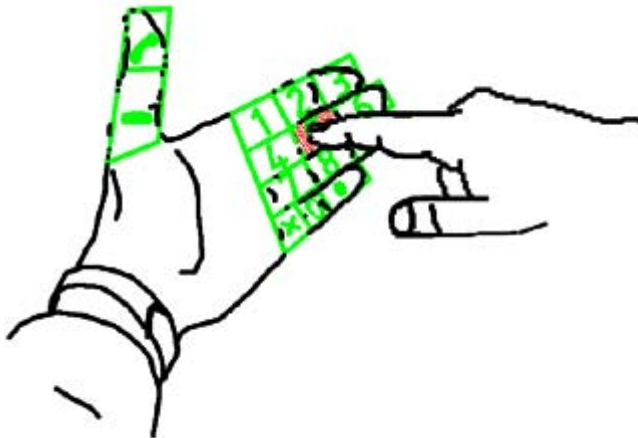


Fig. 34. Making the call by entering the numbers.

4. MARISIL HandSmart – Making a call using the Contacts directory:
- a) To open the Contacts table the user will have to put the device in *information mode*. To do that, the user's thumb finger should be parallel with the other fingers (or the panel palm, as in Fig. 27). A menu will appear showing various

options. One option is the Name section that contains the stored address book with names and other information. The user can select an option by placing and keeping for an extended time the pointing finger over the location of the hand that is overlaid with the desired option.

- b) To find a contact the user can scroll by changing the setting of the interface in “page up page down” mode or it can key in letters using either the phone keypad as in Fig. 35 or, if implemented, by using the handwriting recognition (as proposed in Fig. 32).
- c) Pressing the call menu will place the call







Fig. 35. Contacts browsing and searching.

6.1.2 Example 2: Taking pictures

Described here are the interactions before taking a picture with a device. Beside the action of taking a picture, various interactions can be performed to ensure the desired quality of the picture. These interactions are irrelevant for the purpose of this analysis since they only describe a set of menus and configurations for the camera or the storage of the pictures. Nevertheless, since the MARISIL operating the HandSmart prototype is able to function only by means of a video-based system, it is an obvious application to take advantage of the hardware and use the camera as a photographic device. This example will show the variants in which the user interacts using the MARISIL to take pictures compared with the Nokia's 6600 media phone.

1. NOKIA 6600 – Taking a picture:

- a) The user will select the **Menu → Camera**. The Camera application opens and the user will see the view that can be captured. The user can also see the viewfinder and the cropping lines, which show the image area that will be captured. The user can also see the image counter, which shows how many images, depending on the selected picture quality, fit in the memory.

- b) Pressing  will zoom in on the subject before taking the picture. Press  will zoom out again. The zoom indicator on the display shows the zoom level.
- c) To take a picture, the user will press . It is advised not to move the phone before the Camera application starts to save the image. The image is saved automatically in the Gallery.
- d) If the user does not want to save the image, selecting **Options → Delete** will remove the image. To return to the viewfinder to take a new picture, the user should press .

2. MARISIL HandSmart – Taking a picture:

One possible scenario walk-through for implementing the menu for such application is as follows:

- a) To open the Video menu the user would have to put the device in the *information mode* (by either closing the current mode, or by selecting the mode from the menu). After that, the user's thumb finger should be parallel with the other fingers (or the panel palm, Fig. 27). A menu would display showing various options including the Video option. Once opened (by placing and keeping, for an extended time, the pointing finger over the location of the hand that is overlaid with the *Video option* menu) the user could point the see-through glasses camera towards the area that should be in the picture while keeping the hand panel in front.
- b) The user could zoom in and out using the interface's *page up – page down* mode. Depending on the camera capabilities, a digital or optical zoom could be available. The display seen by the user with the panel hand in front of the target area looks like in Fig. 36.
- c) Placing the pointing finger over the Snapshot option will trigger the picture tacking method. When the user removed the hand from the front look, the device would then take the picture (this way, the hand of the user would not show in the picture).
- d) If the user desires not to keep the image, using the *close* mode (by opening and closing the palm panel in a longer predefined time than the *opening* command; see also pages 84-87) would discard it. Placing the hand in front of the eyes would start the Snapshot option for the next picture (saving the previous one with a predefined name).








Fig. 36. Taking picture using HandSmart device.

6.1.3 Example 3: Recording a video

Described here are the interactions of recording a video clip with the devices. Beside the action of recording, various interactions can be performed to set the desired quality, edit or remove the video from memory. These interactions are irrelevant for the purpose of this analysis since they only describe another set of menus without describing new forms of interactions.

1. NOKIA 6600 – Recording a video:

- a) The user would select **Menu** → **Video recorder**. With the video recorder on, the user could record video clips to the memory. To start recording, after opening the Video recorder, the user should press .
- b) To pause recording at any time the user should press . Pressing  again would resume recording.
- c) Pressing  would zoom into (digital zoom) the subject area. Pressing  would zoom out.

2. MARISIL HandSmart – Recording a video (camera on glasses):

Presently, there are two options to video record using the HandSmart device with MARISIL. The first one is to use the camera mounted on the glasses while wearing the glasses as described here:

- a) To open the Video menu the user had to put the device in the *information mode* (by either closing the current mode, or by selecting the mode from the menu). After that, the user's thumb should be parallel with the other fingers (or the

panel palm, as in Fig. 27). A menu would display various options including the *Video option*. Once selected the *Video option* the user could point the see-through glasses camera towards the area that should be in the video while keeping the hand panel in front (as in Fig. 36).

- b) Pressing record would start the recording when the hand is removed from the view (as in taking the picture example). The glasses will display in 2D the video feed from the camera (the user can shut this feature off if necessary). Placing the hand back in front would stop the recording.
- c) Zooming is not available in this mode.

3. MARISIL HandSmart – Recording a video (camera in hand):


The second option for Video recording was to use the camera after removing it from the mounting point from the glasses (holding the camera in the hand) as follows:

- a) First, the user should open the Video menu as described in the previous method, MARISIL HandSmart – Recording a video (camera on glasses):. After that, pressing the settings and setting the camera to be un-mounted (removed) would set the un-mounted mode.
- b) The user could now remove the camera from the glasses after pressing record and hence the access to a very handy tool for making video records. The camera must have small buttons for stop, pause and record as well as zoom in and out (as the video based system of HandSmart device is not able to handle the MARISIL interaction without the camera mounted on the glasses). At the same time, the user could see the live feed from the camera via the see-through glasses.
- c) Zooming in and out is now possible using the built-in camera zoom-in/out buttons. Remounting the camera and pressing stop will resume the MARISIL operation of the UI.








6.1.4 Example 4: Writing a text

This example emphasises mostly the text writing capabilities of both interfaces. Here is a list of actions for writing a text message:




1. NOKIA 6600 – Writing a text:

A person could key in text in two different ways, using the method traditionally available in mobile phones or using another method called predictive text input (also known as T9 (Grover *et al.* 1998)). To switch between the two methods, the user should press  twice quickly when writing text. Following is a description of how to key in the text without predictive text input:

- The user should press a number key repeatedly until the desired character that is printed on the key appears. Note that there are more special characters available for a number key than there are printed on the key.

- To insert a number, the user should press and hold the number key until the number appears. To key in more than one number, the user could switch the mode to number mode by pressing and holding the  key.
- If a user made a mistake, pressing  would remove the last character. Pressing and holding  would clear more than one character.
- The most common punctuation marks are available under . Pressing repeatedly would change the punctuation mark to a desired one. Pressing the  button would open a list of special characters. Using the joystick the user could move through the list and select a character. To insert multiple special characters, pressing  to select it, and the user could scroll to the required character and press  again until the desired number of special characters are inserted.

When using the predictive input (Grover *et al.* 1998) user interaction is as follow:

- a) The user could write the desired word by pressing the number keys only once for one letter. The word probably changes after every key press. For example, to write 'Hand' when the English dictionary was selected, the users should press the key: 4 (ghi) 2 (abc) 6 (mno) 3 (def). For the same stream of keys the word 'Game' would be available. To change the word, the user could press  repeatedly or by pressing  and selecting **Dictionary → Matches** would show a list of matching words. It is possible to find the desired word by scrolling and selecting it by pressing  of the joystick.

2. MARISIL HandSmart – Writing a text:

There are some options for the user of a MARISIL HandSmart system to write a text. One option is to use the MARISIL HandSmart system as a virtual projected keyboard (this idea is similar to the one implemented by Virtual Devices Inc.¹). In this scenario, the user needs a table on which the HandSmart device using AR view will project a keyboard-like template. When typing over the table on the location of projected keys, the camera will recognize the location and interpret the input as the key over which the user's finger was placed.

Another way of writing the text that can be used with the MARISIL HandSmart is to recognize the user's finger movements in the palm as the palmOne™ "Graffiti 2©" or hand writing characters (Fig. 37). A more detailed description of this method was presented in Section 5.4 including Fig. 32 on page 88 with a schema of the interaction.

¹ <http://www.virtualdevices.net/>



Fig. 37. The Graffiti© characters for hand recognition by Palm Co. devices.

For the comparative study, a third possible use of the MARISIL HandSmart device is available when the user has his/her hand augmented as a phone keyboard. Following is a description of how to key in the text without predictive text input (predictive text has the same interactions as they were described in the previous chapter):

- a) To open the *Text writing* menu, the user had to put the device in the *information mode* (by either closing the current mode, or by selecting the mode from the menu). After that, the user's thumb should be parallel with the other fingers (or the panel palm, see also Fig. 27 on page 83). A menu would display various options including the *Text writing* option. The user should select the *Text writing* option by placing the pointing finger over the selection.
- b) To key in text, the user would press the natural finger partition that is augmented (overlaid) with a number and the letters as in Fig. 38.

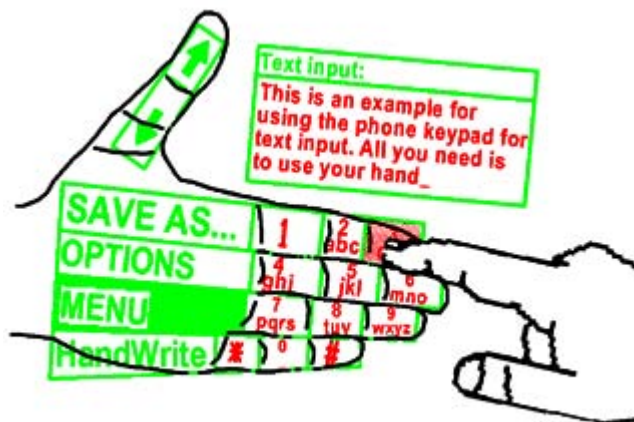


Fig. 38. Writing text with HandSmart using the phone keypad mode.

6.2 Summary

The examples above showed that most of the functionalities of an advanced media phone are possible to be also accessed by using the proposed HandSmart device operated using the specified sign language (i.e., MARISIL). The weaknesses were more in the technical domain and implementations. The recognition process took more time to recognize the user actions than a normal keyboard. This could be improved if, in future implementations, HandWrite as Graffiti is used for text input. Simple interactions, like buttons or placing a phone call, were faster and easier to handle.

The strengths of the design are in the application domain for HandSmart. The proposed device could act, not only as an interface for a media phone but also for a PDA, Tablet PCs, Laptop, and in industrial environments.

This chapter presented the evaluation of the constructs of the proposed MARISIL that could provide the means for interacting with an information device in the same way as when a user interacts using a physical keyboard of the mobile phone. The main benefit of this approach was to extend the screen size and to remove the physical restrictions of the present UI designs for mobile devices. The evaluation of the new interaction revealed that, between the mobile phone input and the MARISIL interaction, there exists at least a one-to-one mapping of the interaction. Compared with the media phone, the new interaction technique has a wider scale of applicability (that will be demonstrated in the next chapter) and it is easily customisable. The next chapter will provide an extensive look on the applications available for this new type of device, most of the applications being inoperable with the current media phones because of their limited screen size.

Having an alternative for the current UIs for media phones is another benefit. Moreover, this kind of interface is capable of handling the interaction in a 3D environment. The user can see artificially generated 3D objects in the space surrounding him while using the interface, extending the personal area of the user with another degree of freedom.

By removing the mechanical aspect of the interface, as well as combining it with the new interaction technique, another area of applications (like VR and MR applications) became possible. This interface also fits very well when used in special environments like clean rooms or sterile medical environments. Another possible application is, and it has been at a particular moment of the development been proposed, for astronauts working in space to use it. Using the MARISIL HandSmart makes the keyboard virtual and hence there are less physical objects floating around when operating, for example, in gravitation free spaces.

The tangible feature of the interface is another benefit of the new interaction technique. The user is co-ordinating both hands and using the eyes in order to input the data. This adds a tactile memory beside the visual one. It is believed (it has not been yet demonstrated) that by doing so the user would be capable of more easily remembering and repeating the gestures than when using a keyboard, as the menu from a media-phone or computer system.

7 Evaluating the New Applications

In the evolution of computers, since the invention of the microprocessor, human-computer UIs have changed considerably. In fact, the history of computer technology goes hand in hand with the history of its UIs. First, the technology pushed for better design of UIs. Then, it was the later ones that made the technology progress. The future of mobile IAs also means the future of its UIs and how they will evolve. It is just a matter of time until the adoption of new ways of interaction will change the human-computer interfaces of these types of devices.

Section 3.2 introduced AR and its applications in various fields. It was concluded that AR systems were applicable in a very broad area from support and maintenance to entertainment and medicine. The early systems presented in Chapter 3 could all be part of the applications of the proposed interface based on AR. Moreover, in Chapter 5 the specifications for the new interaction techniques called MARISIL were listed. The proposed sign language promised that by combining video-based recognition of hand gestures with the use of AR, mobile devices would be accessible in a more pervasive and nonchalant way. This chapter discusses how the proposed interface device called HandSmart (first introduced in Section 4.2.2) and MARISIL (presented in Chapter 5 and evaluated in Chapter 6) could contribute to the implementation of future applications for the new generation of mobile devices. It consists of present applications, or “application papers” describing future utilizations for AR systems and how to apply the novel design to them.

The area of possible applications stretches from low-resource ones, like simple set-ups as in a telephone interface, until it increases in complexity to supporting heavy 3D simulation and virtual enterprise infrastructure.

7.1 Mobile Devices

The main application for the MARISIL based devices is to operate future mobile devices. This particular case is of an appliance that has the characteristic of operating while in motion. One example of such a mobile device is the mobile phone. Other examples of

mobile devices are PDAs and wearable computers. The research community has embraced the latest ones (i.e., the wearable computers) and hence the multitude of innovative ideas including the area of UIs for them. This was possible because, in this case, the research was responsible for the technology push and not the other way. The reverse scenario was the case in mobile phones. The technology was available to build them, but the UIs were not ready. The present UIs paradigms are concerned more with the porting of the common desktop interfaces to work for the mobile devices.

With regard to the number of innovative ideas involved in their UIs implementation, the PDA's are in the middle, between mobile phones and wearable computers. They have a better implementation of the UIs and include interaction that is more intuitive (Bergman 2000a). On the other hand, these devices lack both the processing power and the display space.

As discussed in previous chapters and by others (Brewster 2002), these devices require a larger screen size. The only ones who have addressed the problem so far are the scientists and research projects in the area of wearable computers (starting with (Starner *et al.* 1995)). Even so, they have not provided the users with a good mechanism for interaction. This is where this research contributed: extending the display size and the interaction (use of intuitive sign language).

The display size, when using AR systems, had to deal with a different task compared to traditional two dimension (2D) interfaces. Instead of one area of focus (e.g., the LCD display), the application had to handle a potentially unlimited display space surrounding the user, of which only a very small portion was visible at a time. Even if it sounds intrusive (as the display can pop in front of the user), if the application handles the head movements well, the transition from a small 2D area to an unlimited display use could be done seamlessly. In the approach highlighted here, the head movement would be trivial since the information display point of reference is the user's hands and not where the head is oriented.

In the subsequent sections, evaluations of both mobile phones and PDA's with the new interaction technique are presented. What are the benefits and drawbacks of using AR for these systems?

7.1.1 Mobile Phones

Mobile phones were the first to be considered as the target application for the MARISIL implementation. Functions like browsing, searching, opening, closing and dialling were the first to be described for the sign language. The language definitions are available in Chapter 5. Based on them, the movie about how to make a phone call was made publicly available. Screenshots from the movie highlighted the dialling and placing the call process (Fig. 39).

Beside the use of the phone to place a call, a person could use the interface to operate various services that otherwise were hard to visualize using on the move. A good example of an application would be navigation. Consider a person who is on holiday or a business trip visiting an unknown location. The user might want to have access to various

services, such as restaurants and hotels or to some town tours or museums. While using the interface for placing a phone call, when the interface is in stand-by mode, the user could receive information about the surroundings and navigation information. This is because the user wears see-through glasses all the time when using the interface (in order to see the augmented information). It would be very convenient to use the same way of displaying the interface on the hand, to place this information on other objects.

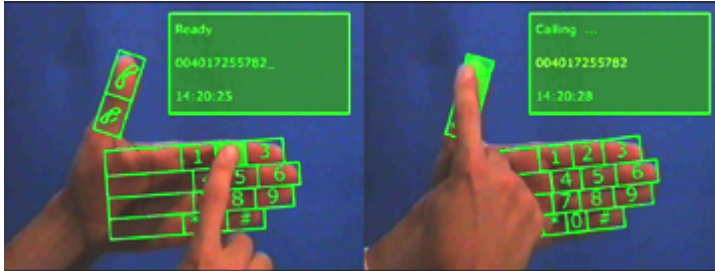


Fig. 39. HandSmart-Dialling a number. Screenshot from the movie describing the language and how to use the interface as telephone.

Using such a system also provides a better sense of privacy. Only the user can see the interface. Hence, it would be impossible for others to guess what the user inputs especially if the interface shuffles the buttons.

One drawback is the inconvenience of wearing glasses. Even if the technology evolved and micro-optical powered glasses that are less intrusive were available, some people might find it uncomfortable to wear any glasses at all. Another issue is the speed of the input due to the detection of the touch of the hand – the current implementation considers a button pressed if the pointing finger is in one region for a certain period of time (2 seconds).

7.1.2 Personal Data Assistants

Personal Data Assistants are a growing segment of mobile devices. Starting from their introduction by Apple (Newton) and followed by more mature devices from Compaq (Palm), the growing number of applications and services have proven that innovative UIs combined with latest technological advances are key factors in consumer adoption.

The high number of applications available for PDA's encouraged the extension of the MARISIL specifications to include a hand recognition concept. The user could use the panel palm as a tablet and write letters. In this way, the correspondence between the PDA user and this interface approach was reciprocal. This should imply a more complicated video-based recognition algorithm than the one used in the recognition software from PDA devices, as it has now to deal with the recognition of the gestures and translate them into writing. In a video-based writing recognition system, as described in Fig. 32 on page 88, the detection mechanism started when the user's hand touched the palm. To detect this, the algorithm should have access to depth information. The depth information could

be included in the system by using a stereo video stream (and a technique called triangularization). Another approach would be to start the recognition based on a period that the finger is held still over a spot and to stop it when the finger again enters a same spot, or by changing the position of the panel hand's thumb (from perpendicular to parallel to the other fingers) to signal the start of a new character. This would mean that the process could suffer from the time taken for these movements, since the user has to keep the finger still for a period at the beginning and at the end of the movement.

Sometimes, the PDA comes with hardware extension, like a keyboard. For the HandSmart device user, the keyboard input provided by the interface is the nine keys of the pad overridden with the 26 letters of the alphabet (as in mobile phones). Another way would be by defining on a flat surface, like a wall or table, the virtual keyboard space. The application should identify the margins and "project" on the surface but keep the virtual keys within the margins. Doing so, while providing the user with a full virtual keyboard, would restrict the user's ability to operate the system while in motion. The same effect is observed when using an extensible keyboard with a PDA.

Even if the extension of the screen provided more space for advanced applications, the interaction speed, as a result of the image recognition process that encounters difficulties in detecting the touch, is slow. A compromise would be to use the interface in combination with a classical PDA tablet that would allow both interactions while having the extended screen benefits.

7.2 User Interface Appliance

In Chapter 4, a discussion on the new interaction techniques introduced the sign language that could help interact with an AR based UI. If the device was a portable mobile communication device, the sign language and interaction technique used was MARISIL. With the specified sign language and with a powerful image processing unit, the interface could morph into anything. It could be a phone, a browser or a video player – basically any IA. When using such interfaces, once the user assimilated them, it could operate any other devices, not only phones or computers (see the principles behind Universal Information Appliance (Eustice *et al.* 1999)).

The UI of an IA comprises of hardware and software components that facilitate the communication between the human and the computer (also known as human-computer interaction). In the case of the UI of a desktop computer, the system requires some hardware input devices such as keyboard, mouse, touch screen, microphone, etc. Moreover, the system requires output devices such as a display, printer, speaker, etc. When using a MARISIL HandSmart device, due to the nature of implementation of the devices, there was a less obvious separation between what was hardware and what was the software component of the interface. Now, because of the video-based recognition system, the component that was previously called keyboard – and therefore, part of the hardware component – had become "virtualized" keyboard and, hence, it also had a software component to handle image processing. Because of this new software component introduced in the system, some of the information processing done by the

microprocessor went to these new tasks, rather than just the interaction with the UI (i.e., image processing, tracking and calibration). These limitations have suggested the separation of the MARISIL HandSmart device interface part from the device that it served (the IA). This concept is called as the “user interface appliance” since the interface per se became an independent device. A HandSmart appliance is a device that understands the sign language defined in the previous chapter (Chapter 5) and with which the user could interact with the computer system using the MARISIL.

When using such a device, the user utilised the hands to perform tasks like input text or numbers. This kind of device had a high level of customisation – i.e., by using the overlaying technique, the user could change the way of interaction with the device. The user could use the interface with various other IAs. Because of that, the interface itself became an appliance, like a button or pen, only that it would incorporate the functions of all interfaces, i.e., it could be a keyboard (Taylor 1999), mouse (Brown & Thomas 1999), tabletop object manipulation (Kato *et al.* 2000) and almost any other physical device that could be used as an input or immersed virtually. However, most importantly, it would extend the display space available for use (Fig. 40).

Fig. 40 demonstrates how the user’s right hand is resizing the navigator window in use to fit the preferences. The dimensions of the window are limited only by the user’s own FOV as the window resides in a virtual space and, if the system is capable of head tracking/orientation, it can be seen as fixed in the “user’s space” – more like a monitor sitting on a table. This is an important requirement when thinking of the future of media phones as the user could be surrounded with such windows, extending the amount of information that was viewable at a time.

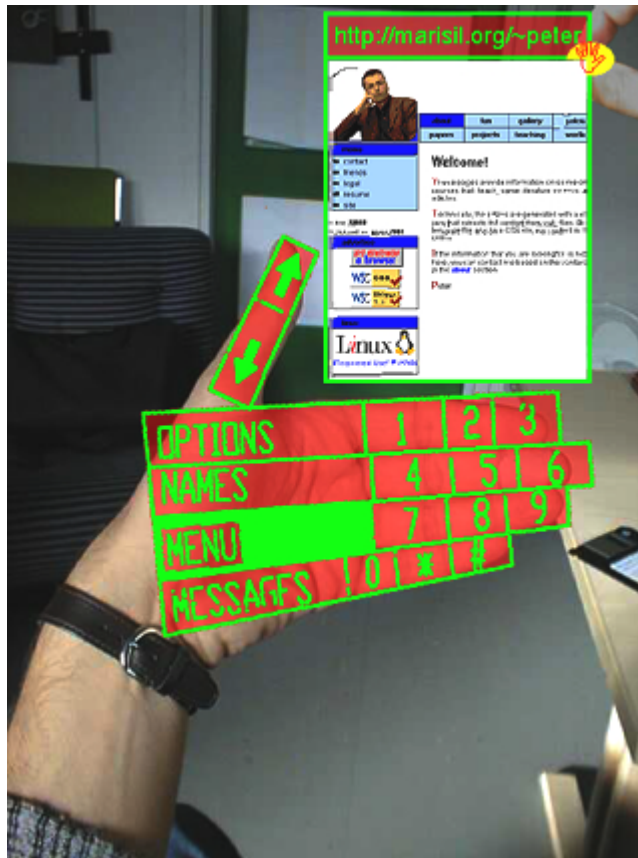


Fig. 40. Hand-browser display resizing from upper right corner. User can resize view to fit preferences with the help of the right hand, extending display to an almost unlimited size while browsing.

By separating the interface from the device, the level of adaptability of the user with the interface was higher. For example, when using a pen, a normal person needs 2 to 4 years to learn how to write (assuming that it is learnt in elementary school at the early age of 6 or 7). Once having achieved this level, the same person could change many pens but would not need to spend any more years to learn again how to write. The same rules apply when operating the MARISIL based interface and hence, it demonstrated the benefits of the separation.

On the software side, current operating systems (OSs) come with a default interface. The interface layer is available either built into the hardware (embedded, as in PDA's) or as a software application (X-window system, etc.). An example of separation between the application software and the interface software is currently provided within Java. Some people also consider web browsers as an example of interface separation. However, a higher and better separation is possible when using a hardware device like HandSmart. Already the new version of the Linux kernel (as of 2.6.0) is modular so that the developers can decide which class of I/O the system requires.

Once the software of the interface is separated from the applications, the link between the user and the machine will always go through the interface. By doing so, the interface would be able to better track the user actions and could adapt, hence becoming more customisable. Even if the user were to change the operating device, the interfaces would always be the same (a good analogy is when keeping the keyboard when switching computers). Because of this migration of the interface, the level of personalisation improves more than with current designs of UIs that change along with the appliances or OS, making it harder to consolidate the user actions in one central repository.

Another consideration is that when the coupling between the UI and the application logic is high (as in most of the current applications available) it could cause serious problems to the development team. This is because having a high coupling means knowing both the application and the UI, while usually the developer of the application is different from the one who develops the UI. A popular example of an implementation of this separation is the XUL Project¹ or the Mozilla Browser. XUL stands for XML-based User Interface Language and it proves that this kind of separation enables the development of easily customised applications for various markets with different languages and looks.

An interesting result from the use of a separated interface is the ability to filter the information based on the user's profiles and intelligent agents (Bannon 1997). Counting on predefined or custom-made filters and on an intelligent interface agent, the device could provide the user with a more intuitive access to information, hence decreasing the weight or information load on the user.

The use of such interfaces with wearable computers would mean better usability and could extend the class of applications. This kind of interface appliance could connect not only with wearable computers, but also with any IA that has a socket for it.

As a conclusion, AR combined with sign language and separated from the device that operates on both the hardware and software level into an interface for IAs, could provide the extended screen for mobility of these devices. Not only that, but it would be a device where it is possible to learn once and use everywhere, which would be another benefit. Moreover, the interface would be more personal, private and being non-mechanical, it could also be more adaptable and not need hardware replacements. Another benefit is that, even if it is virtual, it could have tangible feedback that aids the memory of the user (as in spatial memory). The drawbacks are in the awkwardness when using the glasses and the pace of the input (as discussed in the previous examples).

7.3 Mixed Reality

If mobile devices provide the hardware platform for flexible and movable computers, the MR would mean flexibility between the “fixed” real world and “dynamic” virtual one. Combining the power of such concepts could provide future devices with plentiful applications covering all the current areas of desktop computers applications and beyond.

¹ XUL Project could be found at: <http://www.mozilla.org/projects/xul>

Theoretically (Fig. 3 on page 22), MR covers everything from the real world to the virtual one. A system that is capable of handling AR should be capable of handling virtual reality. To do that, the system should just turn the real world surrounding the user blank and display only the virtual one. In the case of see-through glasses, the user would be able to switch the background to blank (usually it would have a transparency/black level). Because of these capabilities, a person with an AR system could operate in the VR environment and hence would be able to use VR applications.

It is hard to separate the example of applications provided earlier as those that emanate from the VR environment. To give additional examples, other applications could be: virtual conferencing, virtual meeting and collaborative work (Kiyokawa *et al.* 2001). A more artistic touch in applications is the “Interactive Theatre” (Cheok *et al.* 2002). The authors chose MR as a new media to express art. In Fig. 41, a spectator is viewing and interacting with the virtual actors as they are seen in the 3D view.



Fig. 41. Interactive Theatre (Cheok *et al.* 2002), live virtual actors waving at spectator, Hiroshi Sasaki.

Mixed reality also has good applications in the automotive and aerospace industry. The user could interact, browse, examine and generally achieve an alternative view in 3D in such an environment. A combination of physical mock-ups with virtual objects added to them could enhance the view of the designer or architect of the final product. Some people have addressed the subject (Fiorentino *et al.* 2002) but their approach had problems with the UI design and gesture recognition. A combination of such a system (Fig. 42) and the MARISIL specified gestures could bring more freedom to the future designer, in terms of both design and mobility. Using the HandSmart device for input, the user would not need any other object to interact with, but could use only the hands to change and operate the system. Tracking the finger instead of the pen (unlike in Fig. 42) is also possible if high-resolution stereoscopic video cameras are used.



Fig. 42. Automotive industry taking advantage of MR in design process. From Spacedesign application (Fiorentino *et al.* 2002).

The benefits of using AR as an extension of VR are obvious: the system could extend the interaction modes and it could flexibly adapt to the needs of the application. The user could use the same system for 3D work as well as for writing text. It would be up to designers of future implementations to demonstrate the power of this concept.

7.4 Virtual conference

One of the uses of telephony is to provide teleconference services. The latest developments in the area of Voice over Internet Protocol (VoIP) have ported teleconferences to Internet applications (like Virtual Meetings). Such services provide not only voice but also images, shared space, text chat, video telephony and much more.

A system using an omni-directional camera could provide a user with the surround view desired. The camera had a special convex mirror that provided 360° video-image grabbing. Applying a specially calibrated matrix to re-map the distorted mirror image into a normal view, the system could provide the full surrounding view, and by giving an angle or direction of view, the user would be able to see a specific sector around the camera. By combining these methods with the HandSmart device, that included an orientation tracker attached to the see-through glasses, the sensors could provide the direction information for the omni-directional camera view. The user would be able to participate in the meeting and it would be possible to view the surroundings from the remote meeting place by simply moving the head.

An extension to the virtual meeting teleconferencing application is telepresence (Hickey *et al.* 2000). In such applications, if the participants in the meeting were also wearing see-through AR glasses, they could view the remote participant as a virtual

person. The remote participant would be able to see the people at the meeting by using an omni-directional camera (Fig. 43) data provided via a wireless network.

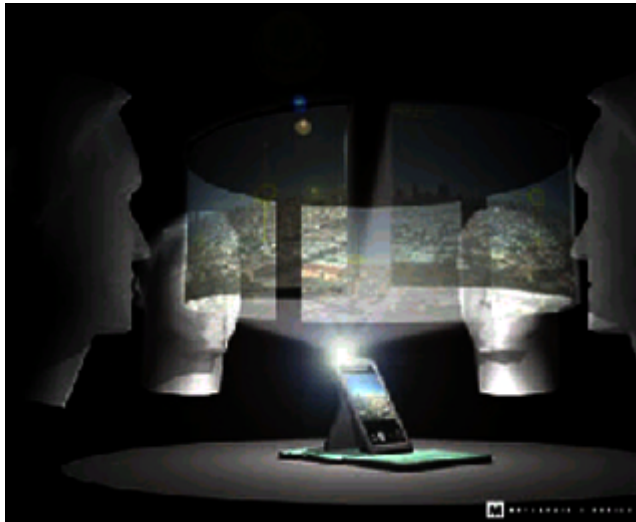


Fig. 43. Telepresence system. Virtual participants can see meeting people through omni-directional camera situated in the centre. Meeting participants are able to see virtual one through AR system. (courtesy Metsävainio OY).

A more advanced system could provide information about the participant in 3D. In order to provide such view in real time, a solution (provided in (Uchiyama *et al.* 2002)) would be to separate the image of the participant and the background and apply, what the authors called, the “dirty” method for occlusion and 3D spatial integration. Another new and sophisticated hardware and software could even provide a smoother immersive projection environment and better 3D video acquisition, enhancing telepresence systems (Gross *et al.* 2003) and collaboration environments. Future advances in image compression should also provide the systems with smoother video feeds.

As seen in the example, the mobile aspect of the interface could provide access, not only to basic applications already available in mobile phones (like phone calls and conference) but also to an extensive set of advanced applications. Virtual conferencing is one example where the user could use the interface to have access to resources that are impossible to reach on a small screen. The virtual participants could be integrated into the real world, providing the user with the ability to see the surroundings while using the virtual conference services. The interaction would also be friendlier and natural, and by detecting the user’s head orientation, the tool could provide the participants with a sense of direction, knowing which person is being addressed. These features, even if hard to implement on a small scale, could become reality in the future, aiding the interaction between people when communicating.

7.5 Virtual Prototyping

In the past, when a product was designed (usually in the automotive or aeronautics industry), in order to test its capabilities, a prototype was created. Mathematics and computer systems have evolved so far that now it is possible to simulate and model, with a certain degree of accuracy, the dynamics of the final product. This method, called virtual prototyping, could provide the designer with the visual enhancements as well as the simulation data as much as a real prototype would do. Major advantages of the method are the reduced time to develop and costs (Pallot & Sandoval 1998).

Since the HandSmart AR system provides the user with the capability to have a VE, virtual prototyping is, obviously, a good application for the system. The only problem was that a mobile system, as the one provided by HandSmart devices, lacked the processing power for such calculations as required for virtual prototyping and modelling. A possible solution would be to use remote resources or distributed processing and then use the mobile system only for the visualisation of the results. The applications of such systems can vary from modelling (Yin *et al.* 1999), design (Fiorentino *et al.* 2002) to virtual enterprise collaboration (Doil *et al.* 2003).

The drawback of using distributed processing is in the latency of the response. Sometimes, the simulation could take place on a platform, but the results would need to be uploaded so that the visualisation would be easier and more natural. Current graphics cards provide enough power and are small, so are probably satisfactory to fulfil the requirements for decent mobile 3D visualization.

The benefit of using AR and the MARISIL also comes when the prototype work is being done virtually (like in the virtual enterprise paradigm, and collaborative work). In such a case, virtual prototyping would be combined with other applications, like virtual meetings, providing an integrated tool for the user. These benefits have a larger impact when they are deployed on a mobile platform, allowing the user to work and operate in various locations.

7.6 Trusted Interface

One important aspect when using tools, like collaborative work or virtual meetings, is about how to build trust between participants. Traditionally, humans build trust based on past information and feelings. A future partner would go through a scrutiny analysis of, for example, the way of handshaking, the way of dressing or the look in the eyes. Presently, this process has reached the stage of being a social science. Books are available on the subject (Fukuyama 1996) and universities have presented different solutions on how to build trust between contacts. Unfortunately, when it comes to VEs, the inability to use natural senses has obstructed the construction of trust. This could be one of the explanations for the slow adoption of virtual meetings systems.

One solution could be the mapping of all the senses to the VE. Unfortunately, this is impossible with current technologies, and especially for a mobile system, since the body is very sensitive to humidity, smell and noises, which are impossible to reproduce

currently in a VE. Not to mention that the system would need to provide access to this type of information in real time, hence requiring undefined bandwidth and no delay in the transmission.

Another solution would be to select what kind of information determined an increase in trust and what did not. Unluckily, this comes with a high variety of preferences for each individual. Moreover, there is the problem with cultural differences. For example, in some Asian countries, it is considered that the user is nervous if he or she is smiling, while in Europe it means that the person is comfortable. Future systems could provide the data about the user, but this data could result in a breach of the privacy of a person. Intelligent agents and brokerage of data through a smart “judge” could bridge this information for the users in the future. An example of an application would be the use of monitoring data of the user (heart beat, body temperature, context-awareness, historical data), compiling the data and then using the previously accumulated knowledge of preferences and then providing the trust level rating for the person. The systems would negotiate what level of trust to exchange, and an intelligent system could decide what kind of information to display for each participant. Although this may sound futuristic, the video-telephone of the future could include such an application.

By using a HandSmart system on a daily basis, as the system provides a vast coverage for information access, the user would become more comfortable and hence possibly trust the interface and virtual partners. Including applications like “the trust factor” would support a friendlier and less sterile environment for communication. Even if many perceive the VR as a cold place, such a combination of AR and intermediate feedback on trust could add a more agreeable and warmer ambiance for work. Intelligent agents and brokerage networks could solve the privacy issue by handling the security and privacy by making only the necessary data available to others (as in real situations). The future could provide more answers to this topic, as technology and society evolve to more integration of humans into the Cybernetic space.

7.7 Tangible User Interfaces

Using AR as a method for users to interact with a computer could enable a novel and natural input style for information systems. By overlaying real objects with suggestive virtual images, the system could turn such real objects into a source of input for the interface of the system. A system that enables the interaction based on real objects that interact with the user on a tangible scale has been called: “Tangible augmented reality” (Kato *et al.* 2000).

Tangible UIs have the quality of being real (or having physical presence) while being augmented with virtual information. The user would be able to use the touch sense when interacting. An example of an application is the tabletop keyboard (Fig. 44). The user uses the paper on the table in order to make the selection (the square in the picture).

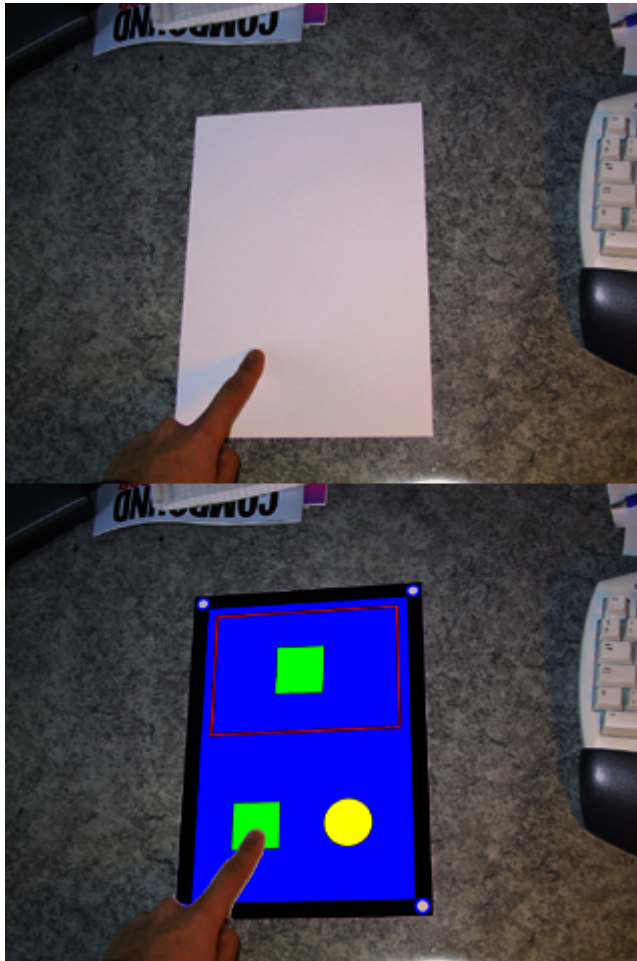


Fig. 44. Tangible interface, table and paper are overlaid with virtual information so that user could input data.

The HandSmart interface can find its use in the tangible interface example presented above. In this particular case, the hands of the user are the real objects augmented with the interface. However, in some applications there is a need for a larger surface to support an ergonomic input. For example, if the user decides to sketch a plan or to select a colour, the interface could use another surface tracked by the video camera and then be used as a table/panel interaction surface – usually a wall, floor or a large table.

Other examples of applications taking advantage of tangible properties of various objects are in architecture and interior design, entertainment and aerospace. The tangible aspect of the HandSmart interface also has an impact on the usability of the interface. A user that has tactile and spatial feedback (as when using the hands to input as specified in MARISIL) could remember the interface's menus faster and better than when having

only visual feedback (as with current GUI and their WIMP approach). This alone is an important aide for learning and memorising how to operate an interface.

7.8 Mobile Engineering

The term mobile engineering applies to people that have to perform an engineering task while being unrestrictedly mobile. The mobile engineering workers mainly perform engineering tasks in an outdoor environment (power cable installers, GIS planners and construction workers). Two characteristics define the mobile engineering field worker: the capacity to perform an engineering task, and the capability of being autonomously mobile. For an engineer that has to work with computers (mobile computer engineer), in order to achieve both characteristics, some special designed devices would have to replace the normal desktop computer.

The MAR interface could enhance the user's view with information about location, maintenance problems, cabling, wire frames (Fig. 45) and other data about the environment (Zhang *et al.* 2001).



Fig. 45. Displaying wire frame from application of AR system in architecture design with icons for different elements (Tripathi 2000).

An example of use of the HandSmart device in interior design could have the designer or the architect visiting the designated location. The person could enter the room, set the interface for *information mode* so that the device could collect pictures and data about the visited location. The interface would use the video camera attached to the glasses and record the surroundings (rotate the view, look around and take snapshots). After the visit, based on the video and/or the snapshots, the user could model the room and set the “tracking markers”. Usually, architects need to access the plan for the site, so a 3D model of the room could be easily provided. The next visit to the location would have the

system set to overlay some information or to even have the office plan for the virtual furniture in place. The calibration and registration could be made using either hybrid tracking (meaning another sensor) or a method for tracking with registered images as described by (Stricker & Kettenbach 2001) for outdoor navigation.

HandSmart could provide various interactions to mobile workers. From navigation and information to recording and playback. The augmentation of the view of the user combined with the portability and flexibility of the input covers a wide area of applications for this domain. Future applications are likely to also address the field of maintenance, administration, inspection and even emergency work.

7.9 Summary

The possible applications of the proposed interaction techniques in mobile information systems were discussed in this chapter. An evaluation of the applications from the points of view of mobility and usability demonstrated the impact and benefits that the techniques add to the interaction when mobile. It is possible to conclude that the argumentation resulting from the use of AR as a platform for implementing new and extended interactions for mobile devices is beneficial for the future of mobile IAs. A review of past applications, the new interaction techniques and the novel ideas of applications for this kind of UIs were evaluated. The conclusion is that by using AR, the developers of new mobile devices would be able to provide access to a wider area of applications (up to advanced areas like 3D simulations, games or Virtual Design). As pointed by Leppälä et al (Leppälä *et al.* 2003), this new class of new products (called “Smart Products”) would provide valuable services and utility functions for the future and they are, therefore, important for the development of society.

To summarise the work so far, it can be said that by using AR with various applications, the display and the interaction space extends. The application coverage increases from current desktop applications that are made mobile up to high-level simulations. The interface could be used in medical/sterile environments; it could also enhance the work of mobile workers as AR could be used in power plants, remote teleworking, and for other outdoor activities. Additionally, the interface could be easily integrated into a multimodal environment. Moreover, everything would be available any time and anywhere, since the system used is mobile, making it more appealing to a larger mass of people (including not only the desk-bound people).

Despite the present drawbacks in technology and implementation, the proposed technique could provide future mobile platforms with better interfaces allowing them to embrace newer and sophisticated applications required by the mobile people of the future. The next chapter addresses these drawbacks and presents possible solutions, evaluating the opportunity to deploy such systems in the future.

8 Building of the HandSmart Prototypes

In order to explore the feasibility of the proposed system, a prototype MAR UI was built. Based on commercial components, the hardware was particularly selected to give the best performance, optimum size, power consumption, as well as wearability and flexibility. Even so, latest technological advances should already be able to provide a version that would be smaller and with more capabilities than the one constructed and described in this manuscript.

The implementation of an AR based UI for mobile applications and services requires a meticulous study of the hardware available to build a mobile platform capable of video recognition and high rendering as well as good software to be able to easily deploy the various applications and services.

The prototyping was done in three phases. The first phase was concerned with getting the system and the software running and of testing the combination of hardware and software. The second phase was for testing the portability, while the last was in trying to achieve the best compromise between flexibility and portability.

8.1 Hardware Architecture

The requirements of a MAR device that support the MARISIL are hard to specify, since the technology in the field is developing rapidly and many applications are becoming available. Nevertheless, some high-level requirements are already known, such as being aware that the system should be mobile and that it should have real-time registration for the UI while being flexible enough to support at least some of the applications mentioned in Chapter 7 (Evaluating the New Applications).

The performance of the system in terms of applications available is limited since, at the beginning, the target applications were simple, like placing a call or browsing a table. Moreover, in the future, the performance will improve as the processing power of the processors increases while their size and power consumption decrease, making them more suitable for mobile platforms.

The accuracy requirement of the system is not as high as for a common AR system. In a normal AR system, the user should experience a high level of realism when overlaying real objects with virtual ones. These kinds of requirements do not necessarily apply when dealing with the proposed variant of the UI. In the current implementation, more important were the speed and the accuracy of the tracking.

Another factor that is very important in AR is the resolution and the size of the pixels of the NTE display (or the see-through glasses). For the UI, this requirement is average and falls under the limits of the current consumer products available (this system had sufficient resolution at the 800x600 provided by the Sony Glasstron).

The important factor for this system was the video camera and the grabbing speed, as well as the processing speed of the video feed. The system needs high speed input data of high resolution images and enough processing power so that the user could easily interact with the interface in real-time. This was a big challenge for this research since in a mobile system, the processing power is also an important requirement.

In the early studies, another approach was used to interact with the system via VR gloves. These devices could track the hand and fingers position and, hence, they had the role of video recognition. Later, after some studies, this alternative re-evaluated version became obsolete because the gloves reduced the portability of the system, and implied using another piece of hardware for the user to wear.

Additional factors of consideration were the brightness, contrast, as well as the degree of see-through. The degree of see-through is important since the user has to be able to handle the interface while moving, and therefore, it is important to be able to still see at low luminosity levels like indoors, dusk or in cloudy weather. These challenges were left for further research since new hardware appears to be able to deliver better see-through glasses than the ones currently available.

The following subsections survey the hardware to support a MARS.

8.1.1 Displays






The display is the most important part of the system since the user should wear it during the work. The choice was hard to make since the see-through HMDs were available with various capabilities and different shapes and sizes.








One problem in choosing the display is that the producer or manufacturer fails to provide easily accessible data on the capabilities of the display. For example the resolution could mean the true resolution (for example 640x480x24). But sometimes, data on the colour pixels was missing (so it is hard to find the real resolution).

The displays considered are listed in Table 10. The prices varied from € 500 up to € 50,000. It is obvious that at the higher price, the better was the resolution, colour and the FOV. The technologies used were also important. The LCD had less see-through transparency, while with the cathode-ray tube (CRT), that usually uses a semitransparent mirror, it was able to permit a higher level of light to pass through (depending also on the quality of the semitransparent mirror used). Moreover, the LCD type of display had less FOV than the CRT. Another technology used for the NTE display was the micro-optical

one. They were lighter, quite fast but had less colour (usually they are black and white) and had less FOV.

Table 10. NTE and HMD devices available for AR use.

	Brand	Model	Year	Panel	Display type	Resolution	FOV
	Canon (Japan)	GT270	1999	2	LCD	270k	
	DAEYAN G E&C/ Personal Display	Cy-Visor DH-4400VP/D H-4400VP-3D aka 3D-i-visor	2001	2	LCOS	1440k	31.2°
	Hitachi	Wearable Internet Appliance	2001	1?			
	Kaiser Electro-Optics Inc.	ProViewX L 40/50 STm		2	LCD monochrome	1024x768	50°
	Kaiser Electro-Optics Inc.	SimEye XL100A	2001	2	CRT	1024x768	100°

	Brand	Model	Year	Panel	Display type	Resolution	FOV
	Micro Optical	Clip-on, Model CO-1 QVGA		1		320x240	
	Micro Optical	Clip-on, Model CO-3 VGA		1	LCD	640x480	10°
	Micro Optical	Invisible Monitor, Model EG-7 QVGA		1		320x240	
	Personal Monitor	Personal Monitor wired/ wireless	1998	1	LCD	263x230	19°
	Sony	Glasstron PLM-S 700 (E)	1999	2	LCD	1.55Mpix	28°
	Virtual Vision	V-Cap 1000		1	LCD	640x480	
	Xybernaut	MA IV		1	B/W		

Some laboratory prototypes also address the occlusion of the objects. In most AR systems, occlusion is partial. The real objects are still available, even if overshadowed by the virtual ones. This is not an impediment, but it could become one if the system is required to be used for certain applications (like virtual prototyping). However, a solution to total occlusion is available. The remedy is to use another layer of display that would block the area needed for overlay (Kiyokawa *et al.* 2000).

The first prototype handled AR on a normal monitor (similar to what a video see-through system provides). Access was available through special see-through glasses but the implementation of the mobile processing unit was not available. Moreover, the video camera used was a Sony Digital 8 HandyCam (DCR TR-7000E PAL) with DV output. Later, for the second and third prototype, the see-through glasses were the Sony Glasstron PLM-S700 (E). This device had good resolution at a decent price and it was light enough to be worn for longer periods without causing problems related to pressure on the nose or head. In the future, as micro-optics based devices become more reliable, the market should provide cheaper and lighter glasses than any currently available.

Future technology should enlarge the options to pick the best display for an AR system. The research in organic light emitting diodes (OLEDs) forecasts production of very thin layer displays that would replace the current semi-transparent mirror or LED based displays. Additionally, the development in reflective technologies, like micromirrors, have been able to demonstrate some good results on some prototypes and even commercial devices (that are presently available only in mono-colour format). One commercial device available is the Nomad from MicroVision. The display is monocular with 800x600 pixels (meaning one colour - red) and it weights 500g. Unfortunately, like other LED-based or LCD displays, it cannot operate at low temperatures (below 4° C). This could be problematic in countries like Finland, if used outdoors.

8.1.2 Video Cameras



The camera was the second important device to be considered. The user should wear it in order to operate the system (either on the glasses or on the shoulders, as a tracking device). Because of this, the requirements for selecting the cameras were concerned with the size, resolution and colour range.

Another aspect involved in the decision was the performance of the frame grabber. When using a frame grabber, it is necessary to know the lag that the device might have between the camera and the processing unit. Usually, these devices come with their own device controller and this restricts the direct access of their resources. The software should use the image data as fast as possible and, if realistic, without using the processor at all. One technique is to take advantage of the Direct Memory Access (DMA). Because of this, the choice of the camera changed in the latest stages of the development. The first prototype used a Matrox video grabber, but later, in the last prototype, a FireWire or IEEE 1394 camera was available. The IEEE 1394 cables and hardware made possible the full 640x480x8 to be transmitted directly to the memory (at 400MBs rate) and hence, operation on the data was faster. The IEEE 1394 (also known as FireWire or iLink) as a

protocol provides low-cost devices that operate high-speed data transmission. It is scalable and it supports both asynchronous and isochronous data transmissions. These make it a very suitable protocol for data transfer in time critical applications like video broadcasting.

In Table 11 the features of the two models used are detailed. The WAT-230 camera had an inconvenience that it came with a frame grabber. Even though size was a better fit to the requirements, the resolution was insufficient to allow a more accurate and robust implementation.

Table 11. Video cameras selected for MARS.

	Brand	Model	Resolution	Power supply	Angle of View
	Watec	WAT-230	320x240	5.4-7.5V	80°
	Orange Micro	iBot Pro	640x480	N/A	62°

A serious drawback of the WAT-230 was the power supply. The WAT-230 required a small power supply, while iBot used the voltage from the cable provided by the system. This was another important reason to adopt the iLink/Fireware/IEEE 1394 based camera instead of the WAT-230. The only problem with iBot was the need to re-engineer it to fit on top of the glasses.

The need to use the camera in dark places was another requirement. This situation could become problematic if the requirement of availability of the operation of the system extended to dark places. A possible solution would be to use of an infrared camera in parallel with the normal camera and a beam splitter to get the beams to match. Sasaki and co-authors described a system in their paper (Sasaki *et al.* 2001) but the implementation was too large to be mounted on the see-through glasses. Possible implementations and future research could provide a solution with the system that would allow the operation even in dark spaces.

8.1.3 Video Cards

In order to output the data, the system requires a video card (or graphics card). The choice of the video card should be based on properties as low power consumption, small size and with not too high a resolution or 3D acceleration. Of course, the more the

features available, the better, but they are not as important as keeping the power consumption low and occupying a small footprint in the mobile computer.

The resolution of the video card was implicitly low since the see-through glasses used (Sony Glasstron PLM-S700) had a resolution display of 800x640. Another important annotation was that the lower the resolution, the lower the power consumption. At high resolutions, the video card spends more energy to process the extra data, buffer and store the images, increasing the battery usage. It is more important to have a small video card with less video power consumption than a fast one with more power demands.

For the prototype system (Fig. 46), the graphic card engine was a CT69030 card. Today, the designer can pick from a larger number of low power and small footprint video cards. A popular one was the ATI Radeon Mobility that included 3D acceleration. Another more powerful were the ATI Radeon or nVidia – based on the chipset NV30 – which were powerful mobile video cards.

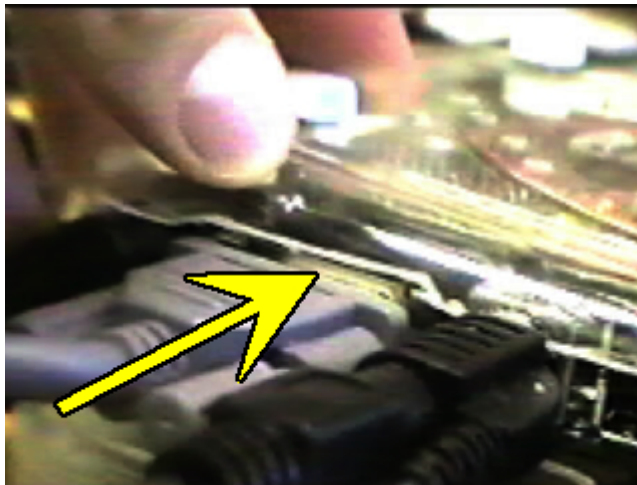


Fig. 46. Mobile system's video card. Side with the connectors.

8.1.4 Network

In order to operate mobile applications and services, the system necessitates access to various communication infrastructures. The prototypes were able to use both WaveLAN (IEEE 802.3x) PCMCIA based cards as well as GSM cards (either via RS232 serial port or via the PCMCIA). Other possible solutions could include satellite link, Bluetooth, or any other form of networking supported by the PCMCIA slot. It is up to the designer of the system or the user to pick the best device that would satisfy the networking needs.

8.1.5 Calibration Sensors

The calibration sensors for the system dealt mostly with the user's eye-glasses position. The devices used were a type of eye tracking ones. These can be found, for example, in small video cameras to assist focus setting. Their role was mainly to find the pupils of the user and calibrate the distance between the glasses' display and the eyes of the user so that proper image overlay could be achieved.

The technology to build these devices has applications in other domains, like in the aerospace industry, for the simulations (what is called gazed control by the pilot). Another application is for car manufacturers, to detect driver fatigue based on the eye movements. The devices are available from various companies (like iViewX HED – which is a head eye-tracking device) but of different sizes, and the one used would depend on the cost and accuracy.



Fig. 47. Head-mounted Eyetracking Device (HED) records eye movement. Courtesy SensoMotoric Instruments GmbH¹.

Another use of the eye-tracking device could be in detecting and setting the preferences of the user based on gaze control (or on which part of the interface the user is looking at mostly). This could add another input to, lets say, a personalisation module that takes care of user preferences.

¹ <http://www.smi.de>

The current implementation was not able to include the calibration sensor due to its size and weight. The calibration could be set manually and this should be sufficient for the system to be able to operate only with this calibration.

8.1.6 HandSmart Prototypes

As discussed at the beginning of this chapter, there were three prototypes. The first prototype was a normal PC, Pentium 166, with 128 Mb RAM. The system used a Matrox frame grabber. The software ran under MS Windows and it was problematic to have access to high video rates. The best solution was to work with Linux for better access to resources and debugging (the Windows based OS crashed more often and did not provide good handling of the errors in terms of messages).

The second prototype was a Pentium 266MHz with 256 Mb RAM (built by Mr. Hiroshi Sasaki). The system used an Orange Micro OrangeLink FireWire PCI. The card was able to operate at 400 Mb/s speed and there were drivers available for Linux. Hence, the obvious solution was to provide the hardware with the Linux OS. Moreover, the FireWire port provided the 12 V power required so that the camera did not require another source (more flexible and less weight). The next requirement was to find a DV/IEEE 1394 camera. At that time (1999-2000), there were no digital cameras based on IEEE 1394 specifications. Therefore, the natural choice for that moment was to use a Sony HandyCam until a better camera appeared on the market. Another new part of the system was the display. The second prototype used the Sony Glastron PLM-S 700 as the see-through glasses.

The last prototype (Fig. 48) looked more at portability. It used a Mobile Pentium 3, 500MHz. The RAM was decreased to 128 Mb. It operated an OrangeMicro iBot digital camera over a FireWire/iLink/IEEE1394 with an OrangeMicro PCMCIA FireLink card capable of 400Mb/s. The only problem was that the PCMCIA card required supplemental power (the 12 V was not available via the PCMCIA bus). Even so, the system was more portable than earlier implementations and able to run the application.



Fig. 48. Prototype parts, unconnected (left) and prototype at work (right).

With the last prototype, most of the applications described in Chapter 1 were realizable. The hardware, even though it was not the fastest possible, provided enough processing power to carry out the recognition task and some small applications. Future implementations should access faster processors with better power consumption and be off smaller size. However, the prototypes provided enough information to defend the idea and demonstrate the possibilities of implementing the interfaces for future mobile information systems. Even though the requirements for mobility were only partially met, future research and technological progress should provide, hereafter, the components to build such a system.

8.2 Software Available

One of the key requirements for the software platform was to be flexible. At the beginning of the implementation, the hypothesis was that in the future the hardware and the software would change, so the start-up plan was to have a modular approach in development. Another issue for debate was the choice of the OS to run the applications and the augmented UI engine. After careful analysis, the final decision led to Linux as a development platform and Java as the programming environment. The quality of being open source gave the development a more robust and insightful approach. Moreover, by using Java technology, the software was portable to other platforms.

The main concern after choosing Linux and Java was speed versus portability. It was known that Java was slower in some tasks demanding fast processing (I/O, memory etc.) but the promise from Sun Microsystem to improve Virtual Machine (VM) encouraged continuation with this approach.

Fig. 49 shows how the Java platform integrated the OS with the Java VM. In the approach, a novel library to deal with the IEEE 1394 Digital Cameras was developed. The library was called jLibDC1394 and the source code was made available online on the popular and large Open Source software development website (sourceforge.net).

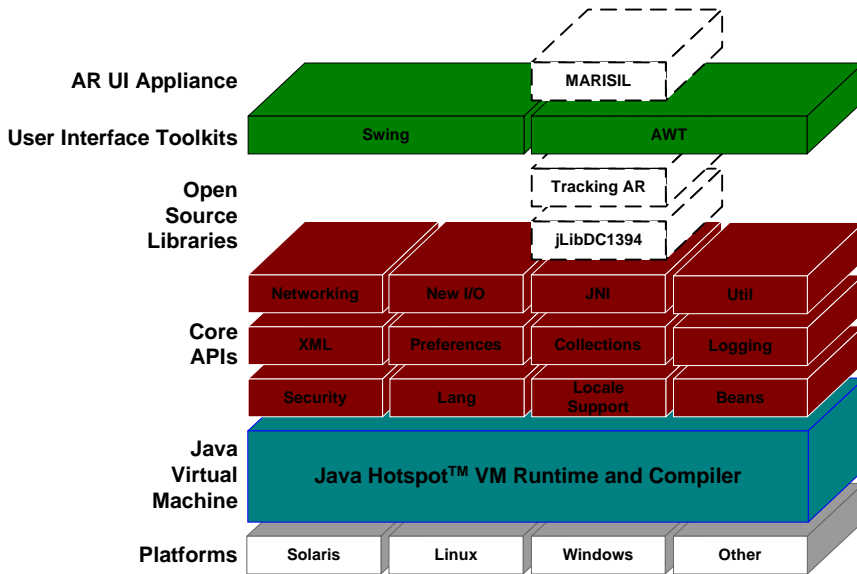


Fig. 49. Sun's Java libraries including custom made and open sources ones developed for prototypes.

The next section details each module from the software package and what were their special functions involved in the process.

8.2.1 jLibDC1394 Library

The jLibDC1394, from the administrative point of view, included one project adminstartor (the author) and four developers (including the author). The author also did the core code and the maintenance, but others contributed their ideas and helped in debugging or porting the code to other platforms (mostly Linux, but there is now a merging with another project to add support for Windows).

From the beginning of the project until the writing of this text, the project received more than 5000 hits (page views) and the library source code has been downloaded more than 400 times (meaning, at least 200 people have been using it). Feedback was received from various users and several versions and upgrades were released (they are not all anymore online).

In order to operate a digital camera within the Java VM, an interface between the VM and the IEEE-1394 digital camera was required. Unfortunately, there was no publicly available library for Java to use the IEEE 1394 devices. There were a couple of libraries for Linux and several native applications (in C or Python). This was the base for starting an open project and designing the Java library.

The IEEE 1394 system for Java decomposes itself into several sub-modules (Fig. 50). The first is at the kernel level, the second at library level (user level) and the third is on the VM and the Java Native Interface (JNI).

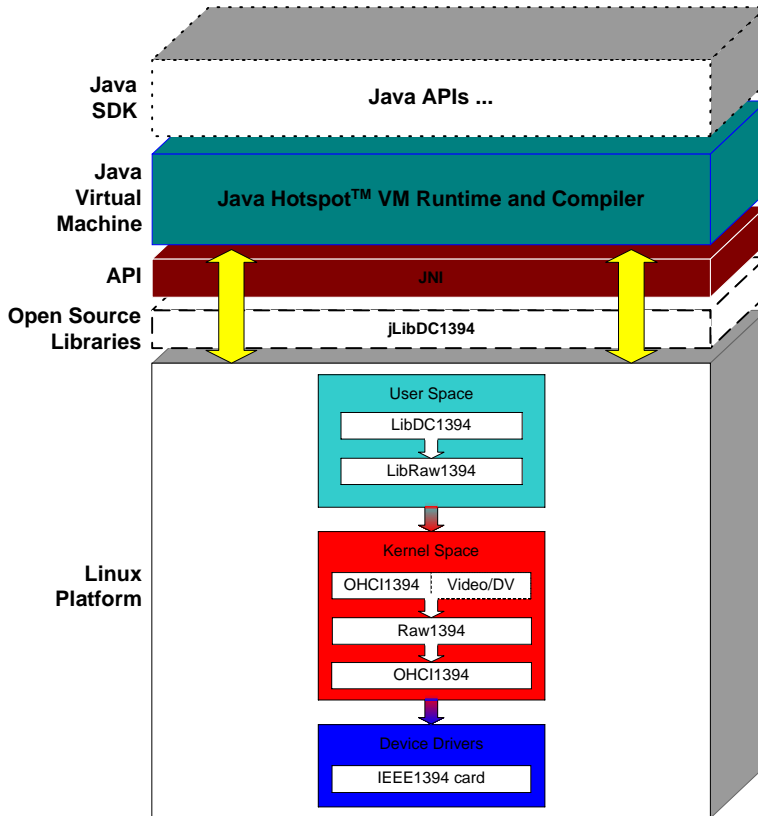


Fig. 50. Linux kernel libraries modules IEEE 1394 subsystem and Java platform for IEEE 1394 Digital Camera. (Linux platform schema based on drawing from (Vandewoude *et al.* 2002)).

The *IEEE1394* module handled the “device driver” part. This module could practically handle the device interrupts and bus activities.

On the Kernel level, there were several modules. The first module was the device driver (OHCI-1394 support). This module could take care of the hardware specifications for the bus settings, buffers and the DMA. On top of this module were the protocol modules. There were various protocol modules. The most popular one was the *Raw1394*. The raw mode allowed direct communication of user programs with the IEEE 1394 bus and thus with the attached peripherals. This mode was useful to set the speed of the bus or to instantiate a node (to get the bus “handle” or address).

Another protocol module was the *OHCI-1394 Video*. This module allowed the video device usage for OHCI-1394 cards. It was useful only when the IEEE1394 bus was used for communicating with digital cameras compatible with the protocol IEEE 1394 DC.

Before the use of this module, another video related module used was the *OHCI-1394 DV*. This specific one allowed the handling of the transmission and receiving of video streams on a frame-oriented interface (useful for Digital 8 Video Cameras).

The module *OHCI-1394 Video* was important since it allowed the grabbing of images directly from the memory (it used a DMA mechanism that delivered the isochronous data from the bus directly to the application memory address).

After discussing the kernel level modules, the next level was the user level or the library. In order to facilitate the programming and operating of IEEE 1394 devices, a library was used. The library (known by the name *libraw1394*) provided the direct access to the IEEE 1394 bus through the *raw1394* kernel's module discussed above. However, even *libraw1394* was awkward to use when handling the video stream as fast as possible. Using another library to handle the *libraw1394* (and hence the *raw1394* module of the kernel) and the *OHCI-1394 Video* module of the kernel was more reasonable. The library provided a high enough programming level to allow the access to the video stream coming from the bus while taking full advantage of the DMA access of the kernel module. Additionally, the library came with specialized functions to control the camera parameters (1394TA-IIDC_Spec_v1_30¹ from IEEE or <http://www.1394ta.org>).

The last level was to access the video frames within the Java environment. An important aspect considered was the way Java operated at the VM level. The Java VM, as the name suggests, is an abstract processing machine (Lindholm & Yellin 1999). Like a normal processor, it has VM code instructions and it manipulates the virtual memory. Only the VM's run time unit is aware of the correspondence between the real or physical memory of the system and the one of the VM.

The Java applications were available to the VM after compilation. They came in a particular binary format (pseudo-code) called the class file format (the class files). Such files contain the instructions to operate the VM (or bytecodes) and some symbol table, and other ancillary information (Lindholm & Yellin 1999).

For the sake of security, the Java VM contained many restrictions of how to run and how to access the class code data. A special mechanism was required so that a Java application could access resources outside of the VM. The mechanism operated through the JNI. The JNI allowed the programmer to run code written in another language and still operate under the Java platform. The interface operated both ways: it allowed Java applications to call native code and vice versa (Liang 1999).

8.2.2 *jLibDC1394 Architecture*

When working with a video stream, the application required a mechanism to enable a large amount of data transfer. Under Java VM, this mechanism was even more complicated due to the nature of the architecture. For example, consider an array of data available on the native memory. If the address of the variable is known, each value of the

¹ The document used to be online here:

http://www.1394ta.org/Technology/Specifications/Descriptions/IIDC_Spec_v1_30.htm

array can be accessed by adding to the address (pointer) the type increment of bytes. This rule does not apply to a Java application.

Under Java, the programmer has no access to the address of the variables (pointer) or objects. Because of that, a complete one-to-one mapping between the physical memory of a variable and the virtual memory used by the Java VM is impossible. A special mechanism of importing or transferring the data between the virtual and the real memory was provided. For example, in order to access the array, the programmer had to use the JNI calls: `Get<Type>ArrayRegion` and `Set<Type>ArrayRegion` where `<Type>` is the array primitive type (like `int`).

Because of this mechanism, the architecture of the `jLibDC1394` had to isolate the data from the methods. To achieve that, the library was split into several layers. One layer or java class took care of the exceptions (errors), another did the implementation and corresponding native calls, and the last one did the mapping for the Java application (Fig. 51).

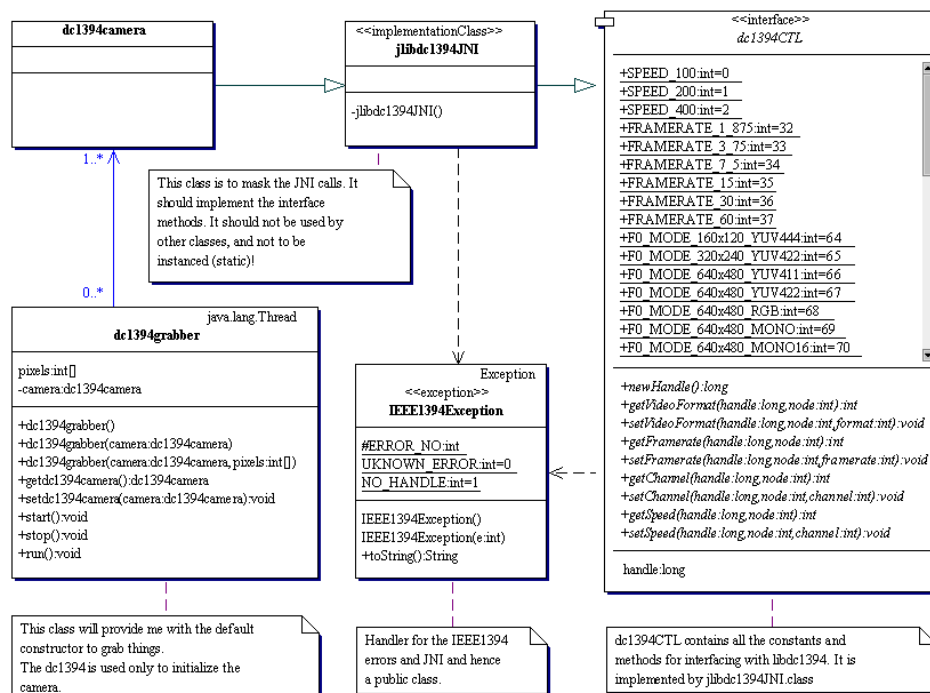


Fig. 51. Class diagram for jLibDC1394.

Following the development of the Java and its VM, the latest version has a more advanced mechanism of working with variables from within the JNI framework. The new mechanism allowed more easily methods like direct referencing (to quote from the Java2 overview web page: “Improvements to the Java Native Interface allow Java and C code to share certain kinds of data structures, which allows for much faster data access between Java and native code.”). The current implementation available online contains

only the *implementationClass*, and future developments would have to deal with a more standard access to the functions.

8.2.3 Finger Tracking and Hand Recognition Process

One of the heaviest in terms of processing and data load process is the tracking and hand recognition process. Due to these factors, the module was designed to reside at the native level, but in the future, when technology will allow more processing power at smaller sizes of the processor, this could become a fully compatible Java implementation. Another alternative could be, of course, the introduction of powerful Java microprocessors.

The tracking and recognition process had to deal with two tasks. The first was to find the hand and to extract its shape in order to overlay the interface. The second was to find the finger and follow the interaction of it. Fig. 52 shows the steps in the tracking process.

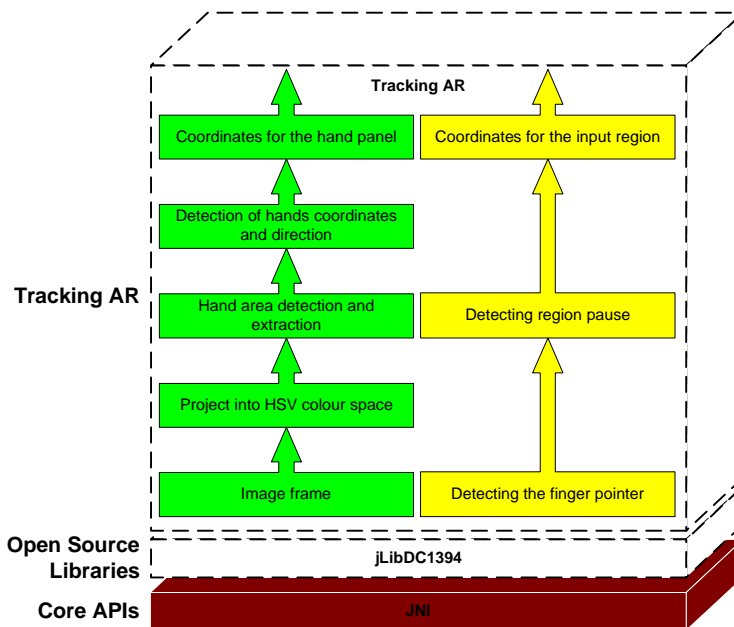


Fig. 52. Tracking and recognition process for AR based UI.

The trip of the image data from jLibDC1394 through the tracking and recognition model starts with grabbing the image frame. Usually this comes in an array but other types of objects or formats are available. The input is Red-Green-Blue (RGB) 640x480 pixels meaning $640 \times 480 \times 24 = 7372800$ bytes (around 7 MB of memory).

Once the image frame was accessible for processing, the software module would change the RGB colour base into the Hue-Saturation-Intensity (HSI) colour base. The

HSI system encodes colour information by separating the information about intensity (I) from the information encoding chromaticity – hue and saturation (HS).

Consider the normalized RGB colour cube from Fig. 53 (A). The diagonal axis of the cube (between Black and White corners) can be the grey axis. The grey values vary from 0 to 1 (since it is normalized. Usually, in practice the value varies from 0 to 255 in binary representation). When it is $\{0,0,0\}$ as RGB, the colour is Black. When it is $\{1,1,1\}$, the colour is White.

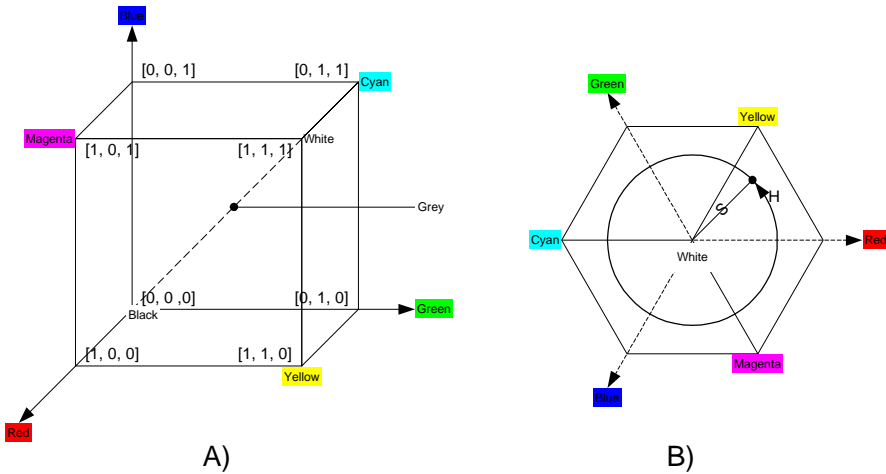


Fig. 53. (A) RGB colour cube with normalized coordinates. (B) RGB plan projection or colour hexagon for HSI representation.

If the colours are projected in a plan formed by the points $\{1, 0, 0\}$, $\{0, 1, 0\}$ and $\{0, 0, 1\}$ (basically the RGB colours) and the other colours are projected from the cube, what is called the HSI hexagon representation is formed (Fig. 53 (B) after horizontal flipping).

In the HSI representation (sometimes known as Hue-Saturation-Value (HSV)) the Hue is defined by an angle between 0 and 2π relative to the red axis. The Saturation is the value between 0 and 1 (sometimes expressed in percents) and the Intensity is the vertical axis perpendicular in the centre on the hexagon plane.

The choice to use of HSI space instead of RGB was done because it provided better support for image processing algorithms. When using this colour code the focus was only on the chromaticity. The source of lighting became obsolete, since in normal life it can vary much (between places, or in time, etc.).

In order to change the RGB colour scheme to this, the module used the RGB2HSI conversion algorithm (Fig. 54).

```

"the RGB values are normalized, meaning that for 0-255
values we have 0-1. To do that, you have to put the value
and multiply it with 100/255);"
RGB_to_HSI (in R,G,B; out H, S, I) {
    I:= max (R,G,B);
    Min:= min (R,G,B);
    Diff:=I-Min;
    if (I≥0) then S:=Diff/I else S:=0;
    if (S≤0) then {H:=-1; return;}
    "Compute the H (angle) based on the relative RGB"
    if (R=I) then H:=( $\pi/3$ )*(G-B)/Diff;
    else if (G=I) then H:=( $2\pi/3$ )+( $\pi/3$ )*(B-R)/Diff;
    else if (B=I) then H:=( $4\pi/3$ )+( $\pi/3$ )*(R-G)/Diff;
    if (H≤0) H:=H+ $2\pi$ ;
}

```

Fig. 54. RGB to HIS conversion algorithm written in pseudo-code.

Using the colour information (predefined) and the new encoding (HSI), the Tracking AR module separated the hand area from the background area. The new information was stored into a binary image in which the white area was the hand area (1) while the black area was the background (0). The new array occupied only $640 \times 480 = 307200$ pixels.

After separating the hand colour from the background image, the module applied a series of filters (median filters). This evened and filled the holes caused by normal noise produced by pixels latency or cables.

After the smoothing filters were applied, a template matching method did the recognition of the open hand. There were two given template images (Fig. 55, the top boxes) which were required to be located or found in the binary image. To do this, the method had to find the match of the template image to all (or as many) of the possible places it could be located in the binary image. Some distance function (in this case a simple Euclidean distance) was used to measure the similarity of the template and the specific binary image. The method extracted the co-ordinates with the smallest distance as the co-ordinates of the template image from the binary image.

In the next figure and in the open hand recognition method, there are two template images used. One for the thumb, that once detected is looking for the next template, i.e., the lower part of the palm. Based on the location of the two templates, the algorithm was able to recreate the hand volume and direction, and send the coordinates for building the hand panel (diagram from Fig. 52 on page 130).

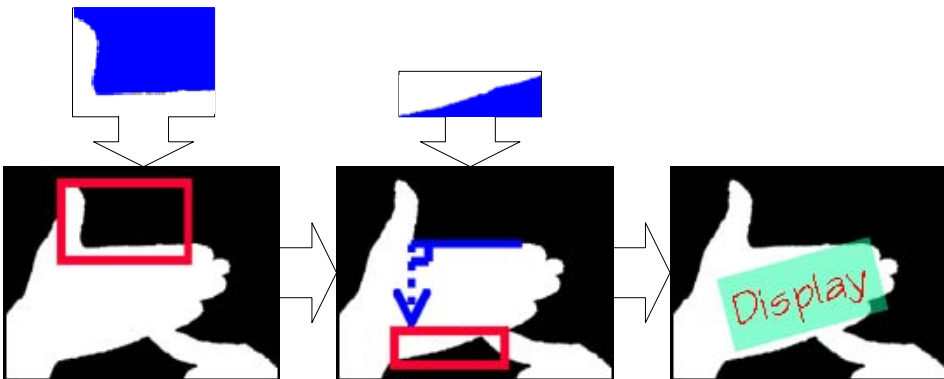


Fig. 55. Template matching technique used for recognition of open hand. Courtesy Dr. Hiroshi Sasaki.

Tracking of the selecting-finger (normally, the index finger from the other hand as described in Chapter 5.1 on page 80), this system used a marker (Fig. 56) and checked the image area to find it. The hand panel area, described in Fig. 55 as the green part from the far right box, split into several logical regions/segments (depending on the image area). Usually these segments were the building boxes for the virtual keys. If the tracked finger remained for a certain amount of time over a region, the system recognized that as the input from that region and sent the coordinate to the upper module (i.e., the MARISIL module, next section).

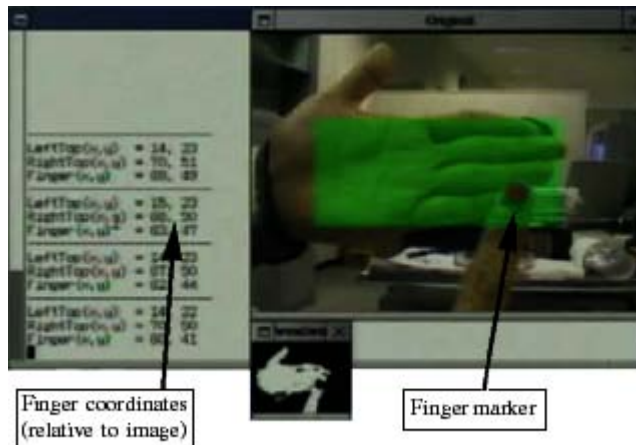


Fig. 56. Finger tracking based on marker (right) and coordinates displayed for debugging (left).

This module was partially written in C and C++ and used separately for speed reasons. A possible integration in the proposed Java framework was started and in future, if agreed by the developers, a possible open source project could be started. The main difficulties were to remap all the image operating functions and to optimise them for Java use.

8.2.4 MARISIL Module

Previous chapters presented how the image went from the camera until it became a set of coordinates for the UI. The MARISIL module handled the overlaying of the UI on the user's augmented view. Another task for this module was to decide about the input (meaning what type of action it was and how would it be possible to launch the corresponding task).

In order to build the interface, the module needed the coordinates of the hand and where to overlay the hand on the interface. This module also decided about the command associated with the user input based on the information overlaid on the user's hands. The Tracking AR module only provided the coordinates, while the MARISIL module associated the segment with the action and launched the application/method respectively. A short schema is presented in the next picture (Fig. 57).

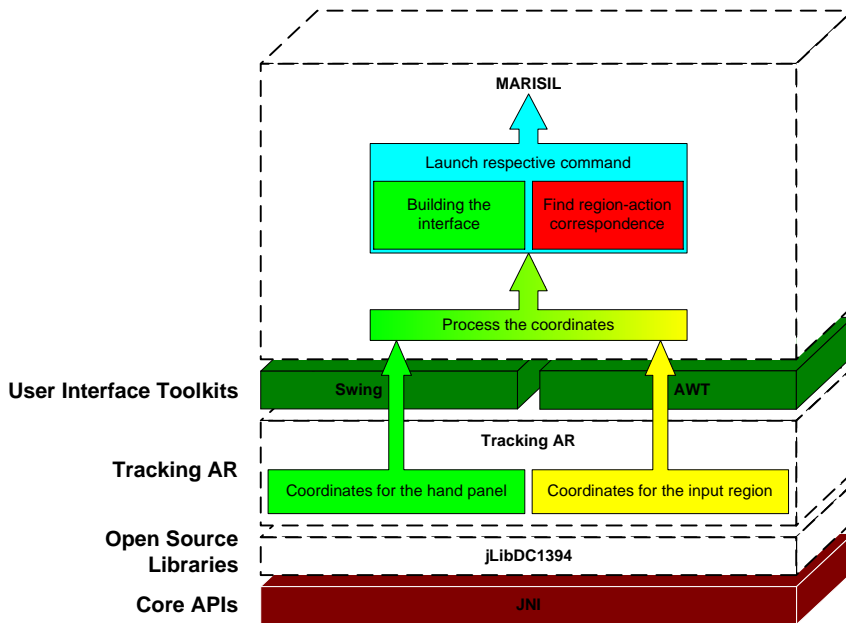


Fig. 57. MARISIL Module, processing the input from the Tracking and launching the command.

The current implementation of the module excluded the processing and calibration between the users's eyes, the display, the camera and the hands. This is the current status of the implementation and the decision not to implement these processes was because of the current state of technologies and processing power available.

In order to have a true overlay of the interface and a semi-automatic calibration process between the eyes and the see-through glasses, the system should also include an eye-tracking device. Once the position of the eyes is known, based on the position of the camera (in this hardware implementation the camera was attached on top of the glasses) the display could calculate the exact place where to overlay the interface. Moreover, in

order to have a stereoscopic overlay (both displays to show correctly the hand overlaid by the interface), another camera would be required that would provide distance measurements.

With the current hardware implementation, adding the second camera would increase the IEEE 1394 communication bandwidth by 50% causing a drop in quality of the video transmitted (from 400Mb/s to 200Mb/s). This, as a result, would affect the registration and tracking process.

Another aspect is that in order to have the stereoscopic view and distance detected correctly, at least one more cameras would be necessary to provide accurate data. Such a system (Shapiro & Stockman 2001) requires another processing unit that could be a powerful DSP device and hence the cost of implementation would rise without improving the quality sufficiently. A more optimal approach would be to consider the input from only one camera, the second one to be used with a lower resolution just for the small calibration of the second eye (since accurate information to calculate the distances is not required, like with the stereo-camera system described above).

In the current implementation this module was written in C and C++ and it did not deal with the calibration. The module could be easily implemented in Java and the future implementation should contain this module as a Java class.

8.3 Summary

This chapter described the constructs for the HandSmart system. The hardware and software involved in the construction of the prototypes were described. Future plans and improvements are still necessary. The work done so far provided the basis of how to build a working system. Also presented was the current state of the art of a MAR based UI built to operate an interface like the one proposed (based on the MARISIL). It also described the implementation problems faced by developing such a system as well as future improvements that could contribute to a better adoption by industry and the mass market. Future technologies and better integration of components already promise to deliver a better implementation. The next chapter evaluates the implications of the artefacts.

9 Evaluation of the HandSmart Implementation and some Implications

This chapter discusses the status of the prototype and evaluates the implementation of the prototypes discussed in the previous chapter. The focus of the evaluation was on the technical aspects of the implementation. A usability evaluation of the system was impossible due to it being at the early stage of the prototypes development and due to the lack of resources. The author believes that in order to better evaluate the usability of such a system, a more advanced prototype should be produced and in sufficient numbers so that users could experience the full interaction as with a fully functional system.

From the technical point of view, the evaluation discusses the achievements made both for the hardware and software implementation of the prototypes. The hardware evaluation criteria were the size, portability and availability of the components for the system. The software evaluation concerned the capacity to implement the basic functions specified by the proposed MARISIL interaction (described in Chapter 5) on the current hardware and the flexibility for future implementations and additional requirements.

Moreover, this chapter discusses the implications of the current capabilities of the implemented system. The chapter ends with a section defining the perspectives of the work.

9.1 Hardware

The current system was available in a small enough size to fit in a large pocket or a small backpack. The glasses and the camera were also available with a see-through option and they were small enough to wear them on the head without too much inconvenience.

Moreover, new surveys of the technology showed a faster move towards the mobility aspect of the hardware. There are now commercial products that target specifically the mobile aspect of the computers (Fig. 58).



Fig. 58. LittlePC's LPC-301 Small FootPrint comes with a Pentium 3, LAN, USB, Firewire 1394, Audio, Serial, Video at the price of a new computer and includes even a built-in CD-ROM.

The commercial products focused mainly on providing the same applications as a normal PC at the expense of being mobile. When designing a MAR, the designer should focus more on the AR aspects rather than on the desktop computer features (like the CD-ROM).

Another hardware component that was a part of the system was the see-through glasses. The prototype used the Sony Glasstron. Current advances in micro-optics have encouraged companies like MicroVision to market products like Nomad that are based on such technologies. These devices are lighter, have better contrast and a higher level of see-through (the amount of ambient light visible through the glasses is hence higher).

As for the calibration issue, the eye tracking devices have also progressed. Eye tracking devices are presently already used in small video cameras to assist focus setting. The technology has found other applications in the aerospace industry for simulations (what is called gaze control by the pilot). Another application developed has been by car manufacturers, to detect driver's fatigue based on eye movements. The devices are available from various companies (like iViewX HED – a head eye-tracking device) and in different sizes, at a relatively average cost and accuracy. Another use of the eye-tracking device could be for calibrating the preferences of the user based on the gaze control (or on which part of the interface the user is mostly looking at). This could add another input to the, let us say, a personalisation module that takes care of user preferences.

The power for the system could be supplied from a Lithium ion (LiIon) battery. An estimation of power consumption from the current devices is available in the next table (Table 12). Naturally that some values vary with the usage (between idle and maximum usage). Some estimations are available, based on present technologies, but these were not considered and integrated in the prototypes presented. There will be a large difference between the current prototype and a future one, including what some of the present technologies could provide in terms of power consumption.

Table 12. The Power consumption of the current and future MARSs.

Component	Voltage	Current power (mW)	Future power (mW)
Sony Glastron	12V	1100	
MicroOptics	≈5V		≤50
WL 100 net	12V	130	
4MB CF net			≤20
Current PC	24V	100000	
Mobile PC	12V		≤20000
iBot video	12V	1400	
Fire-I	12V		≤900
TOTAL		≥102630	≤20970

Integrating the newest technologies, the hardware specifications for the system could be met. There are already some MARSs available for commercial use. When more applications are available, future devices would probably emerge to integrate the new technologies. This would also depend on the social acceptance of wearing such devices as well as how the technologies will be able to provide them. While the prototypes did not meet the criteria of portability due to the off-the-shelf integration, there are commercial products that can now follow closer the specifications allowing the building of a smaller and portable mobile IAs that will support the proposed AR interface.

9.2 Software

At the beginning of the research, the main concern was whether a mobile information system could handle a large amount of image processing. The greatest problems were how to track in real-time the hands of the user with a high enough level of accuracy that would deliver a seamless virtual-real interface. One solution considered was to use a mobile passive system, in which the video stream would be transmitted to a powerful computer for processing and the results fed back (like in the Regenbrecht and Specht mPAR approach (Regenbrecht & Specht 2000)). Hence, the question was whether it could be available on a small, mobile platform and not on a high performance fixed server. The requirements to run in real-time were impossible to meet by this approach, due to network latency and the amount of data needed in the transmissions.

Another implicit requirement was to have the software modular so that it could handle the future changes in the hardware or in the software platforms. In order to have the portability as high as possible, the programming was done, as much as possible, in Java.

To be able to operate the high amount of processing, some of the code had to run on the native level. This meant that the portability of some sections were lost for the sake of speed. Another problem was to find the proper API to handle the IEEE 1394 protocol. Since the protocol is still young, the Java platform did not provide the APIs for handling the standard. Because of this, another part of the code had to also run at the native level and had to be included into the Java VM's libraries.

These drawbacks are minor compared with the achievements. In the future implementation of the Java platform, it is most likely that Sun Microsystem will include such libraries as provided by our jLibDC1394. Moreover, future technologies will provide better integration of Java code that would lead to faster execution of Java binaries. This could allow the future implementation to process the images under the Java VM and not outside it.

The system was able to track the user's hand and the pointing finger in real-time, allowing interaction with the system. A demo video made by Dr. Hiroshi Sasaki demonstrated the placing of the call using the system. In the future, more applications are likely to be implemented and demonstrated on the platform.

The author believes that the UI application could be fully portable to Java in a couple of years. This would mean that the whole code would be available in Java and it would be fully portable to any machine running a Java VM.

9.3 Implications

The major implications of the implementation and their impact are discussed in the following sections. The implications discussed here are concerned only with the issues that have not been discussed in previous chapters. These implications became evident during the development of the prototypes and later, during the use of the system.

9.3.1 *Security and privacy*

An important requirement of a mobile device's UI is the privacy and the security of the user. When using the current interfaces, like mobile phones, PDAs, or even speech based command interfaces, they all lack privacy when operated in a public space.

Considering that only the user of the AR based UI is able to see the interface, the level of privacy is even higher than the one provided by desktop computers (that are not generally used in public spaces). Moreover, one idea was to have a special secured version of interaction when using this interface. In this version, the interface's labels and numbers would no longer be laid out as usual, but they could be randomly distributed on the palm panel. Because of this, nobody except the user would be able to identify the real value introduced when placing the pointer finger on the palm panel.

In the light of this characteristic, a higher level of privacy and security could be achieved. Another example could be to use the technology on ATM machines on the street. An onlooker, even when watching very closely, would not be able to know what is the PIN code introduced by the customer using the proposed secure version of the system.

9.3.2 Ergonomics

The ergonomics of a UI involved many studies and experiments. Based on the current laboratory experiments, there were new specifications concerning the use and ergonomics of the interface.

During some tests, it appeared that while using the interface, the user should keep the panel hand up all the time, even if no input was being performed, but mainly to read some information. This action is futile, since the hands should serve only for interaction and not to just display some static data. Because of this, the new implementation considered that when the hand withdraws from view, the information should remain on the display, unless the user closes the interface with a gesture before removing the hand. This problem was first discovered during the production of the movie (Section 4.2.3 on page 74).

Another problem was the raw removal. The anatomy of users hands can differ from person to another. Some people can, while others are not able to, close just one finger of the hand (Fig. 30 on page 86). The new implementation considered that with the help of the pointing finger this should be an easy task for everyone.

Other ergonomic aspects that were also looked into were the feeling of wearing the glasses all the time. While the current system was quite awkward in terms of size and fashion, future system could have a lighter and nicer design for the see-through glasses as well as an alternative classical display that is physical and not virtual, allowing the device to operate also in a classical manner.

Psychologists have demonstrated that a person is able to remember easier the things done if the action to do them involves more than one of the senses (or a combination of senses). When using a tangible interface, the user is able to see and to touch in order to input the data. By using the hands, not only is the user able to see but also the touch helps in repeating the gesture more easily. This speeds up the learning process and the way the user remembers the actions produced while working with the interface.

The virtual desktop is another potential application for AR. Such an approach would use the advantages of accessing an unlimited size display surrounding the user when using the AR system. Handling a 360° area of desktop where applications can be available and a lot of objects are present could be confusing. This approach removed the confusion by setting the hand as the origin of the display. Because of this, the confusion of not knowing where the information was situated and the need to search for it by turning around disappeared.

While the system was not mature enough to undertake full scale ergonomic testing, the development and laboratory tests showed that the quality was good enough to allow the system to be used for several hours without causing fatigue. Of course, some people may experience specific negative reactions due to the see-through glasses, but the operation of the system proved to be ergonomic enough. Moreover, the development of future displays will probably improve the standard of the glasses (Sugihara *et al.* 1999) and any distracting symptoms should disappear.

9.3.3 Social Aspects

Even if for the researchers the appearance and comfort of wearing a MARS might have seemed unimportant, these attributes are of importance for normal users. In spite of latest developments and enhancements in terms of size, the FOV, brightness, colours and the resolution, there are still many requirements to be satisfied in order to improve the comfort and the appearance of the system. The current trend is to have the see-through display attached to normal glasses while the video stream is available through a wireless Bluetooth link (MicroOptical DV-1 Wireless Digital Viewer).

Other aspects that could contribute to the adoption of the system is by combining the novelty of the system with new services and applications that are not available on present platforms (like navigation assistance and digital agents). Such applications could create the need for the hardware and possible adoption of such devices to be worn even during the entire duration of a day. In the future, such a mobile lifestyle could spread to all social classes, from workers in the factory, to researchers in the aerospace industry.

When using an AR based UI, the system grabs the interaction via a video camera. This peculiarity could cause problems when using the interface in a museum or around military bases (as the access to a camera will become so ubiquitous). In the future, such restrictions may no longer exist or maybe there will be other technologies that would better detect the breaking of the security restrictions (since the access to a camera would become so ubiquitous).

Another problem arises when the devices are used in extreme climatic conditions. For example, LCD displays will not work at temperatures below zero. A video camera is also incapable of capturing gestures in dim light. Future systems should provide solutions to these problems in order to broaden the area of operation of the devices.

9.3.4 Mobile Services and Applications

The previous section mentioned that mass acceptance would change if more applications and services were to be available. Due to the nature of the system, the potential area of the applications is broader than any present system available on the market. The fact that the system can and is operating in a VE brings a new dimension to the development of applications and services for such devices (Section 3.2 and Chapter 1 for examples of applications).

The initial scope of the work was to provide the future mobile phones with an extended display that would allow better interaction for operating higher-level applications. Through research, quite a large number of applications are now available for the system, and they cover various research fields like medicine (surgery, x-rays, tele-medicine), engineering (maintenance, inspection, design, prototyping), architecture (site inspection, mobile computer aided support), tourism (guided tour, navigation assistant), business (teleconference, browsing, brokerage) and education (outdoor play, games, entertainment).

9.4 Perspectives and Conclusion

This chapter examined some of the implementation problems and provided some possible solutions for future problems using the proposed AR interaction. It presented the current state of the art of MAR based UIs operated by the specifications presented as the MARISIL. It also described the implementation problems faced during developing such a system as well as future improvements that could contribute to better adoption by industry and the mass market. Looking at the future, there appear to be many opportunities for the work on AR to be applied in order to improve the interaction or to widen the fields of applications of mobile devices.

While exhaustive research on social acceptance of such devices was unrealisable due to physical constraints, based on conference reviews and reaction noted in public appearances of the author, a reasonable conclusion can be made that the system will be accepted. Each conference and presentation of the idea caused much discussion and many questions were raised from the audience on how the system would work or how long it would be before a product would be available in the shops. An impressive achievement was during a presentation of a high ranked official of the European Commission on future research, where this interaction technique was regarded as very important for future development in Europe.

10 Conclusion

The overall contribution to the research field was the finding of a new interaction mode and the realization of a device to support it so that it could extend the display of future mobile IAs. The approach of this work was to virtualize the display so that physical restrictions, size and weight, of the mobile device would not affect the size of the display. Augmented reality was able to provide the best environment to realize the compact UI that was both virtual and yet allowed the user to see the real surroundings.

The new interaction mode proposed was based on a combination of gestures and augmented information. Based on the findings, this novel interaction mode (called the MARISIL) is an alternative for current interaction modes available for mobile phones. With the new method, the interface is overlaid on the user's hand by interpolating between the hand of the user and the eyes equipped with a special see-through glasses capable of displaying computer generated images. This interpolation required that the hand and the hand's natural partitions are recognized and correctly overlaid with artificial data representing buttons and icons needed to interact with the system. The AR method, combined with the advantages of the hardware, see-through glasses and video recognition are the root of the implementation. The user then, interacts with the device by selecting a desired partition of the hand that had been overlaid with the UI. The device would perform a specific action accordingly to the hand segment pointed to by the user (as in pressing a key on the keyboard).

The set of gestures specified for the interaction was based on common life actions that are used during various activities of operating IAs. However, one requirement was to keep the interface flexible to allow the assimilation of new gestures. If, for example, in the future the pattern of how users behave is changed, the interface should be able to assimilate (learn) the new language specifications easily. Defining the core set of gestures equipped the interface with a standard framework for interaction. During the research, this framework was evaluated to see if it was capable of operating a certain number of tasks (in the evaluation, the tasks selected were those commonly used by a user operating a media phone). Even so, the necessary core specifications should not stand in the way of developing new interactions. During the evaluation, the comparative studies showed that the new interaction technique had enough capabilities to cover all the interactions

required to operate a media phone device. Future extensions could enable newer and more advanced applications capable of operating devices that are more sophisticated.

Another contribution was the implementation of a platform, called HandSmart, to handle the interaction described by the MARISIL. The platform had three implementation phases (called prototypes) and they were all elaborated in collaboration with Dr. Hiroshi Sasaki from NAIST, Japan.

Once the platform was available and it demonstrated that an extended display for mobile devices was possible through virtualization of the interface, the interest shifted to how the new interaction technique would be capable of implementing and operating current applications and services. As a result, further effort was allocated to evaluate how the proposed interaction mode would work as a mobile phone interface. During the evaluation, the study looked deeper into how future and more advanced applications could be integrated on a mobile device using the proposed interaction mode. As a result of the exploration, various applications coming from both desktop computers and AR using MARISIL as the interaction mode were now capable of being deployed on mobile devices. This result confirmed that using this approach of the UI, the mobile devices could benefit from an increased number of applications and services that were otherwise hard to access with the current interfaces.

Through the implementation, the study investigated whether the technology could provide a system capable of handling the proposed MARISIL interaction technique. The evaluation showed that, with the current technologies, it was possible to construct such devices. Some mobility restrictions like the size, weight and power of the device could be met using current technologies.

The study was also interested to equip mobile devices with an interaction mode that would be easy and based only on a set of intuitive and simple gestures. Previous research in this area only emphasized the non-mobile aspects of the systems while the interaction required physical objects (i.e., markers, pointing devices, trackers and other input devices). When using this approach, the interface was capable of morphing into any interface for an IA, like a phone, a browser or even a video player. Another benefit of using this approach was the capability of redesigning and separating the interface. In the future, such an interface, if regarded as an “interface appliance”, could become a universal appliance interface for the TV, washing machine or other IAs.

In an effort to prove the novelty of the idea, the survey located several patents – of which Kazama and co-authors and Fukushima and co-authors. (Kazama *et al.* 2000, Fukushima *et al.* 2002) were the closest – related to the proposed interface. Even so, both of them failed to achieve the same functionality or operation as in this approach. Fukushima did not refer to how to display the real environment and was more concerned with static information retrieval. In Kazama, the user operated in the virtual world and, therefore, was not able to see or react in the real environment. This work led to the application for a US patent in 1999 that was later accepted in 2004 (Pulli & Antoniac 2004).

From the evaluation of the abstract language and prototypes, the work also identified several new benefits of using AR that were not initially recognized. One benefit was the level of privacy when operating the HandSmart device. The privacy increased when using the new device since the AR interaction required the user to use the NTE display, for which, by construction, the output was restricted only to the person wearing it. In

addition to this hardware restriction, the privacy could be further enhanced by the software by shuffling the locations of the buttons in the interface. This random rearrangement of the interface layout (seen only by the user) would obfuscate, for an onlooker, the meaning of the gestures executed by the person operating the interface. Another benefit was the flexibility and customisability of the interface. If the interface is virtual it allows easy adaptation and modification of its operations, unlike hardware keys that are hard to modify. Virtual also means a longer life, since the components of the interface are non-mechanical and, therefore, not subject to mechanical failures. The system should be able to “grow” with the user. This means that once the user is accustomed to the basic operations, he/she could teach the system to carry out advanced tasks by including new interactions (more like *learn once, use all life*).

Another benefit of using this interaction mode is that operating the interface is possible without needing to pull something out from the pocket, but only enabling the system. This is useful in circumstances where the user needs access to the interface without having to search for the device to operate it (i.e., a mobile phone has to be pulled out for use).

As the system, in order to operate, requires a certain separation between the device and the UI (at hardware and software level), tracking the user's actions is possible at any time when the owner is operating the interface. This permits the system to access broader interaction situations, as people are expected to use the interface in various conditions. Moreover, including artificial intelligence and more sensors or context-awareness in the system would enable the device to accommodate better experiences, particularly in the area of personalization and filtering, hence providing a more individualized interface.

The final contribution was to provide the large open source software community with a glimpse of the work. Parts of the work were made available on-line and they could be downloaded and tested. Future packages will be similarly deployed, including an AR based UI framework.

The beneficiaries from this research would be not only the users, but also the technology providers of PDAs, Mobile Phones, Tablet PCs, and other equipment. During the exploration and dissemination of the applications of the new system, for some specific areas, the interaction mode and system design proved to be particularly beneficial. For example, for the workers in sterile environments and clean rooms, the system could provide an interaction mode that did not involve physical contact of the hands with another object (usually a keyboard). Additionally, for people performing in conditions that force them to wear certain equipment (like scuba divers or astronauts), that limits their abilities to use a keyboard, this system could be easily adapted to operate under these conditions. A fascinating possibility that came up from the reviews of the system by some medical professionals would be the deployment of the HandSmart to aid medical workers (i.e., telepresence systems, medical examinations), surgical operations on some patients (i.e., memory support activities).

The novel design and the use of top technologies when building the artefact attracted much curiosity. This suggests that normal users are interested and they would be willing to try such an interface. Some critics commented on the cumbersome aspect of wearing the see-through glasses and the inability of users to have eye contact. The introduction of micro-display glasses removes these inconveniences as long as the wearing of some eyeglasses is not inconvenient for the user.

Playing with the hands to place a call was an easy task to achieve with the proposed system. Also browsing and viewing pages provided an enjoyable set of gestures. However, writing text remained a difficult task (similar in difficulty to writing text on a mobile phone), even when using shortcuts or an adaptive dictionary character completion (or T9 (Grover *et al.* 1998)). Several possible solutions could improve the interaction, but only a more advanced implementation of the recognition process would be able to really solve this problem.

The aim of this work was to assimilate, evaluate and promote AR for mobile devices. The motivation relied on the assumption that AR based interaction would enable mobile users to operate in a more natural way without having to worry about the size of their device keyboard. The interest of the work resides mainly in the field of mobility. As the method used is AR, the research area has a multidisciplinary as well as vast coverage of issues and applications. This, added to the number of mobile applications and services currently available, shows the potential. Moreover, this also means that further integration work and knowledge from various research areas will be necessary to improve the results presented.

The research implications mostly come from the multidisciplinary aspect of the work. The information processing science field and human-computer interactions when combined with technology surveys and integration, produces a novel user centric design on which it is possible to conduct further research and lead to novel contributions to the science. The significance of the work can be summarized as a complete survey, with implications, on how AR can provide a new interaction mode that is capable of enhancing future mobile IAs with a virtualized UI.

The MARISIL specifications and HandSmart artefact provide the designers and developers of mobile devices with extended knowledge on what issues are to be tackled and what could be already done with current technologies. Moreover, the sign language specification underlines the need for a more natural or human oriented approach in the interaction between the user and IA. The work, therefore, unveils a novel, simple and intuitive approach to help a user interact better with mobile IAs.

11 Future Work

The subject of UIs for mobile computing is at the beginning of a new era. Technologies to support such interfaces are still in the embryonic stage or they are being imported from different fields. Future research should focus on integrating media technologies with portable and communication devices. This should take advantage of the new technologies and simultaneously increase the value of the present one.

The work presented in this thesis demonstrated that such a system is realistic. Implementing a fully operational prototype is a matter of time. Future research should include usability tests and ergonomic studies in relation to specific applications.

Moreover, some more publications and a full operational implementation of a prototype media phone using the MARISIL should be used to promote the concept to industry. To progress, mobile phone technology needs a revolutionary interface in order to attract new users. This interface could overcome the present limitations of the media phone technology, and enable the expansion into new areas yet unaddressed.

Mobile AR is still uncharted, particularly the UIs of such systems. Powerful concepts have arisen and good applications and implementations are already available. Applying innovative UI techniques to this area would open new opportunities to integrate services and applications and, therefore, create a new momentum for developments and improvements in the field. Augmented reality supplemented by mobility could cover a wide range of services and applications that could intensify the personal experience and communication capabilities within the professional and other communities. Better mobility would also promote the communication skills of people. This would also have an impact on new fields like virtual product development and designing, increasing the quality and decreasing the time a product takes to reach commercialisation.

Sometimes, the level of abstraction used in UIs causes confusion in people using them. The idea behind the implementation of this interface and the way to use it looked at the intuitive process of moving the hands. Such processes vary between individuals and hence, they are hard to model or to define as a standard. The design of the interface should be more open to changes so that a person operating it could redefine it based on the individual background coupled with physical and mental capabilities. The way to set or to automatically adapt the interface for individual preferences would have to be the

subject of deeper and further research. Computer vision is another field that would need to advance in order to sustain such requirements.

The ideal AR based UI for mobile applications and services should be non-intrusive, self-adapting to the user's preferences and very ergonomic. These competences are already studied by both industry and academia. It is only a matter of time before this or similar devices will become a reality, enhancing the life style of future generations.

During coming years, the study will continue the exploration of MARS. Starting from 2005, the academic curriculum will include two courses related to the subject of this study. In these courses, the students will learn about the current advances and relations between mobile information systems and AR. Moreover, the research work will also continue in the laboratory. Several project funding applications to advance the present work were submitted last year. New proposals are planned for submission this year. In the future, the research focus will be on improving the prototype implementation and finding new application areas for the HandSmart devices.

References

- Abowd GD & Mynatt ED (2000) Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7(1): 29-58.
- Akao Yo (1990) *Quality function deployment: integrating customer requirements into product design*. Productivity Press, Cambridge, MA, USA.
- Antoniatic P (2002) HandSmart-Enhancing Mobile Engineering User Interfaces with MARISIL. 8th International Conference on Concurrent Enterprising (ICE 2002), Roma, Italy, Center for Concurrent Enterprising, 331-338.
- Antoniatic P, Bendas D & Pulli P (2000), User Interface for Location Support. Cyphone Project Report. Oulu, Finland, University of Oulu.
- Antoniatic P, Hickey S, Manninen T, Pallot M & Pulli P (2004) Rich Interaction for Training Collaboration Work in Future Mobile Virtual Enterprise. 10th International Conference on Concurrent Enterprising (ICE 2004), Sevilla, Spain, Center for Concurrent Enterprising, 119-126.
- Antoniatic P, Kuroda T & Pulli P (2001a) User Interface Appliance for Mobile Devices. 2nd WWRF Workshop Meeting, Helsinki, Finland.
- Antoniatic P & Pulli P (2000) The Trust Factor in the Virtual Enterprise. A Mobile Broadband Service Approach. 6th International Conference on Concurrent Enterprising (ICE 2000), Toulouse, France, Center for Concurrent Enterprising, 287-294.
- Antoniatic P & Pulli P (2001) Marisil-Mobile User Interface Framework for Virtual Enterprise. 7th International Conference on Concurrent Enterprising (ICE 2001), Bremen, Germany, 171-180.
- Antoniatic P, Pulli P, Kuroda T, Bendas D, Hickey S & Sasaki H (2001b) HandSmart Mediaphone, Advanced Interface for Mobile Services. World Multi-Conference on Systemics, Cybernetics and Informatics, Orlando, FL, USA.
- Antoniatic P, Pulli P, Kuroda T, Bendas D, Hickey S & Sasaki H (2002) Wireless User Perspectives in Europe: HandSmart Mediaphone Interface. *Wireless Personal Communications* 22(2): 161-174.
- Azuma R, Baillot Y, Behringer R, Feiner S, Julier S & MacIntyre B (2001) Recent advances in augmented reality. *IEEE Computer Graphics and Applications* 21(6): 34-47.
- Azuma RT (1997) *A Survey of Augmented Reality*. Presence: Teleoperators and Virtual Environments, 355-385.
- Bannon LJ (1997) Problems in human-machine interaction and communication. 7th International Conference on Human Computer Interaction jointly with 13th Symposium on Human Interface, San Francisco, CA, USA, Elsevier, Amsterdam, Netherlands, 47-50.
- Barfield W, Baird K, Shewchuk J & Ioannou G (eds) (2001). *Applications of Wearable Computers and Augmented Reality to Manufacturing*. Fundamentals of wearable computers and augmented reality. Mahwah, NJ, Lawrence Erlbaum Associates.

- Barrilleaux J (2001) 3D User Interfaces with Java 3D: A guide to computer-human interaction in three dimensions. Manning Publications Company, Greenwich.
- Behringer R, Tam C, McGee J, Sundareswaran S & Vassiliou M (2000) A wearable augmented reality testbed for navigation and control, built solely with commercial-off-the-shelf (COTS) hardware. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 12-19.
- Bendas D & Myllyaho M (2002) Wireless Games-Review and Experiment. Proceedings of the 4th International Conference on Product Focused Software Process Improvement, PROFES 2002, Rovaniemi, Finland, Springer-Verlag, 589-600.
- Bergman E (ed) (2000a). Information appliances and beyond: interaction design for consumer products. Morgan Kaufmann series in interactive technologies. San Francisco, Morgan Kaufmann Publishers.
- Bergman E (2000b) Introduction. In: Bergman E (ed) Information appliances and beyond: interaction design for consumer products. Morgan Kaufmann Publishers, San Francisco: 1-8.
- Bertelsen OW & Nielsen C (2000) Augmented reality as a design tool for mobile interfaces. Symposium on Designing Interactive Systems (DIS '00), Brooklyn, New York, ACM Press, 185-192.
- Bier EA, Stone MC, Pier K, Buxton W & DeRose TD (1993) Toolglass and magic lenses: the see-through interface. Proceedings of the 20th annual conference on Computer graphics and interactive techniques, ACM Press, 73-80.
- Billinghurst M, Bowskill J, Jessop M & Morphet J (1998) A Wearable Spatial Conferencing Space. Second International Symposium on Wearable Computers (ISWC'98), Pittsburg, PA, USA, 76-83.
- Billinghurst M & Kato H (1999) Collaborative Mixed Reality. In: Ohta Y and Tamura H (eds) Mixed reality: merging real and virtual worlds. Ohmsha, Tokyo, Japan: 261-284.
- Birkfellner W, Huber K, Watzinger F, Figl M, Wanschitz F, Hanel R, Rafolt D, Ewers R & Bergmann H (2000) Development of the Varioscope AR. A see-through HMD for computer-aided surgery. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 54-59.
- Brewster S (2002) Overcoming the Lack of Screen Space on Mobile Computers. Personal Ubiquitous Computing 6(3): 188-205.
- Broll W, Schäfer L, Höllerer T & Bowman D (2001) Interface with angels: the future of VR and AR interfaces. IEEE Computer Graphics and Applications 21(6): 14-17.
- Brown T & Thomas RC (1999) Finger tracking for the Digital Desk. User Interface Conference, AUIC, 11-16.
- Burdea G (1996a) Force and touch feedback for virtual reality. John Wiley & Sons, New York, NY, USA.
- Burdea G (1996b) Preface. Force and touch feedback for virtual reality. John Wiley & Sons, New York, NY, USA: xiii-xiv.
- Burdea G & Coiffet P (1994) Virtual reality technology. J. Wiley & Sons, New York, NY, USA.
- Butz A, Baus J & Kruger A (2000) Augmenting buildings with infrared information. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 93-96.
- Butz A, Hollerer T, Feiner S, MacIntyre B & Beshers C (1999) Enveloping users and computers in a collaborative 3D augmented reality. Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99), 35-44.
- Caudell TP & Mizell DW (1992) Augmented Reality: An Application of Heads-Up Display Technology to Manual Manufacturing Processes. Proceedings of Hawaii International Conference on System Sciences, Hawaii, HA, USA, 659-669.
- Cheok AD, Weihua W, Yang X, Prince S, Wan FS, Billinghurst M & Kato H (2002) Interactive theatre experience in embodied + wearable mixed reality space. Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2002), 59-68.
- Cho Y, Lee J & Neumann U (1998) A Multi-ring Fiducial System and an Intensity-Invariant Detection Method for Scalable Augmented Reality. Proceedings of 1st IEEE and ACM International Workshop on Augmented Reality (IWAR '98), 147-165.

- Company HM (1994) American Heritage Dictionary of the English Language, [Electronic Version], [cited 1994]. Available from: CD-ROM.
- Cruz-Neira C, Sandin DJ & DeFanti TA (1993) Surround-screen projection-based virtual reality: the design and implementation of the CAVE. Proceedings of the 20th annual conference on Computer graphics and interactive techniques, ACM Press, 135-142.
- Dahne P & Karigiannis JN (2002) Archeoguide: system architecture of a mobile outdoor augmented reality system. Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2002), 263-264.
- Dam Av (1997) Post-WIMP user interfaces. Communications of the ACM 40(2): 63-67.
- Day RG (1993) Quality Function Deployment: Linking a Company with Its Customers. ASQ Quality Press.
- Dempski KL (1999) Augmented Workspace: the world as your desktop. International Symposium on Handheld and Ubiquitous Computing, Karlsruhe, Germany, 356-8.
- Dodsworth C (1998) Digital illusion: entertaining the future with high technology. ACM Press; Addison-Wesley, New York, NY, USA.
- Doil F, Schreiber W, Alt T & Patron C (2003) Augmented reality for manufacturing planning. Workshop on Virtual Environments, Zurich, Switzerland, ACM Press, 71-76.
- ESA (2004) Essential Facts about the Computer and Video Game Industry, [online], [cited 15/12/2004]. Available from: <http://www.theesa.com/pressroom.html>.
- Eustice KF, Lehman TJ, Morales A, Munson MC, Edlund S & Guillen M (1999) A universal information appliance. IBM Systems 38(4): 575-601.
- Feiner SK (1999) The importance of being mobile: some social consequences of wearable augmented reality systems. Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99), 145-148.
- Feiner SK, Macintyre B, Höllerer T & Webster A (1997) A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. The Proceedings of the First International Symposium on Wearable Computers (ISWC '97), Cambridge, MA, USA, 74-81.
- Feiner SK, Macintyre B & Seligmann D (1993) Knowledge-based augmented reality. Communications of the ACM 36(7): 53-62.
- Feiner SK, Webster A, MacIntyre B & Höllerer T (2004) Augmented Reality for Construction, [online], [cited 16.04.2004]. Available from: <http://www1.cs.columbia.edu/graphics/projects/arc/arc.html>.
- Figl M, Birkfellner W, Hummel J, Hanel R, Homolka P, Watzinger F, Wanshit F, Ewers R & Bergmann H (2001) Current status of the Varioscope AR, a head-mounted operating microscope for computer-aided surgery. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001), 20-29.
- Fiorentino M, de Amicis R, Monno G & Stork A (2002) Spacedesign: a mixed reality workspace for aesthetic industrial design. Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2002), 86-95.
- Flickner MD, Lu Q & Morimoto CH (2001) International Business Machines Corporation, assignee. Gaze-based secure keypad entry system. US patent 6,282,553.
- Foxlin E & Harrington M (2000) WearTrack: a self-referenced head and hand tracker for wearable computers and portable VR. The Fourth International Symposium on Wearable Computers, 155-162.
- Fuggetta A, Picco GP & Vigna G (1998) Understanding code mobility. IEEE Transactions on Software Engineering 24(5): 342-361.
- Fukushima N, Muramoto T & Sekine M (2002) Canon Kabushiki Kaisha, Tokyo (JP), assignee. Display Apparatus Which Detects an Observer Body Part Motion in Correspondence to a Displayed Element Used to Input Operation Instructions to Start A Process. US patent 6,346,929 B1.
- Fukuyama F (1996) Trust: the social virtues and the creation of prosperity. Penguin, London.
- Gemperle F, Kasabach C, Stivoric J, Bauer M & Martin R (1998) Design for wearability. Second International Symposium on Wearable Computers, Pittsburgh, PA USA, 116-122.

- Genc Y, Sauer F, Wenzel F, Tuceryan M & Navab N (2000) Optical see-through HMD calibration: a stereo method validated with a video see-through system. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, 165-174.
- Giaglis GM, Kourouthanassis P & Tsamakos A (2003) Mobile commerce: technology, theory, and applications. In: Mennecke B and Strader TJ (eds) *Mobile commerce: technology, theory, and applications*. Idea Group Publishing: 67-85.
- Gross M, Würmlin S, Naef M, Lamboray E, Spagno C, Kunz A, Koller-Meier E, Svoboda T, Gool LV, Lang S, Strehlke K, Moere AV & Staadt O (2003) Blue-C: A Spatially Immersive Display and 3D Video Portal for Telepresence. *ACM Trans. Graph.* 22(3): 819-827.
- Grover DL, King MT & Kushler CA (1998) Tegic Communications, Inc, Seattle, WA, USA, assignee. Reduced keyboard disambiguating computer. US patent 5,818,437.
- Halttunen V & Tuikka T (2000) Augmenting Virtual Prototypes with Physical Objects, [online], [cited 11.07.2002]. Available from: <http://www.hci.oulu.fi/digiloop/publications.html>.
- Harrison BL, Ishii H, Vicente KJ & Buxton WAS (1995) Transparent layered user interfaces: an evaluation of a display design to enhance focused and divided attention. *Proceedings of the SIGCHI conference on Human factors in computing systems (SIGCHI 95)*, Denver, CO, USA, ACM Press/Addison-Wesley Publishing Co, 317-324.
- Heuer A & Lubinski A (1996) Database Access in Mobile Environments. 7th International Conference on Database and Expert Systems Application, Zurich, Switzerland, Springer-Verlag, 544-553.
- Hickey S, Manninen T & Pulli P (2000) TeleReality-The Next Step to Telepresence. World Multiconference on Systemics, Cybernetics & Informatics, Orlando, FL, USA, 65-70.
- Hinckley K, Pierce J, Sinclair M & Horvitz E (2000) Sensing techniques for mobile interaction. The 13th annual ACM symposium on User interface software and technology, San Diego, CA, USA, ACM Press, 91-100.
- Hoff B & Azuma R (2000) Autocalibration of an electronic compass in an outdoor augmented reality system. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, 159-164.
- Höllerer T, Feiner SK, Hallaway D, Bell B, Lanzagorta M, Brown D & Julier S (2001) User interface management techniques for collaborative mobile augmented reality. *Computers-&Graphics*, 25: 799-810.
- Hoover D (2001) Arcot Systems, Inc, assignee. Method and apparatus for secure entry of access codes in a computer environment. US patent 6,209,102.
- IIS (2003) Interactive Imaging System's Second Sight M1100, [online], [cited 26/03/2003]. Available from: http://www.iisvr.com/products_mobility_ssm1100_specs.html.
- Ishii H (2002) Tangible bits: designing the seamless interface between people, bits, and atoms. *Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, 199.
- Ishii H & Ullmer B (1997) Tangible bits: towards seamless interfaces between people, bits and atoms. *SIGCHI conference on Human factors in computing systems*, Atlanta, GA, USA, ACM Press, 234-241.
- Ishii H, Wisneski C, Orbanes J, Chun B & Paradiso J (1999) PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play. *The SIGCHI conference on Human factors in computing systems*, Pittsburgh, PA, USA, ACM Press, 394-401.
- ITU-T (1995) Standard No. E.161, Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network, International Telecommunications Union.
- Järvinen P (2001) Research questions guiding selection of an appropriate research method. *Proceedings of ECIS2000*, Wien: Vienna University of Economics and Business Administration, Austria, 124-131.
- Jiang B & Neumann U (2001) Extendible tracking by line auto-calibration. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001)*, 97-103.
- Johnson P (1998) Usability and mobility: Interactions on the move. *First Workshop on Human Computer Interaction with Mobile Devices (EPSRC/BCS 98)*, Glasgow, UK.

- Julier S, Lanzagorta M, Bailiot Y, Rosenblum L, Feiner S, Hollerer T & Sestito S (2000) Information filtering for mobile augmented reality. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 3-11.
- Kalawsky RS (1993a) Knowlton's virtual push button system (1975). The science of virtual reality and virtual environments: a technical, scientific and engineering reference on virtual environments. Addison-Wesley, Wokingham, England: 23-24.
- Kalawsky RS (1993b) Philco Corporation's Headsight television surveillance system (1961). The science of virtual reality and virtual environments: a technical, scientific and engineering reference on virtual environments. Addison-Wesley, Wokingham, England: 20-21.
- Kalawsky RS (1993c) The science of virtual reality and virtual environments: a technical, scientific and engineering reference on virtual environments. Addison-Wesley, Wokingham, England.
- Kalawsky RS (1993d) What are virtual environments? The science of virtual reality and virtual environments: a technical, scientific and engineering reference on virtual environments. Addison-Wesley, Wokingham, England: 4-7.
- Kamba T, Elson SA, Harpold T, Stamper T & Sukaviriya P (1996) Using small screen space more efficiently. Proceedings of the SIGCHI conference on Human factors in computing systems: common ground (SIGCHI 96), Vancouver, British Columbia, Canada, ACM Press, 383-390.
- Kasai I, Tanijiri Y, Endo T & Ueda H (2000a) Actually wearable see-through display using HOE (Holographic Optical Element). 2nd International Conference on Optical Design and Fabrication (ODF2000), Tokyo, 117-120.
- Kasai I, Tanijiri Y, Endo T & Ueda H (2000b) A forgettable near eye display. The Fourth International Symposium on Wearable Computers, Opt. Technol. Div., Minolta Company Limited, Japan, 115-118.
- Kato H, Billinghurst M, Poupyrev I, Imamoto K & Tachibana K (2000) Virtual object manipulation on a table-top AR environment. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 111-119.
- Kazama H, Onoguchi K, Yuasa M & Fukui K (2000) Kabushiki Kaisha Toshiba, Kawasaki, Japan, assignee. Apparatus and Method for Controlling and Electronic Device with User Action. US patent 6,111,580.
- Kendon A (1986) Current issues in the study of gesture. In: Nespoulous J, Perron P and Lecours A (eds) The Biological Foundation of Gestures: Motor and Semiotic Aspects (Neuropsychology and Neurolinguistics). Lea: 23-47.
- Kerttula M, Pulli P, Kuosmanen K & Antoniac P (1998) Deliverable No. CE-NET/VTT/WP4/v1.0/150998, CE Tools and Techniques. ESPRIT Project No 25946. Oulu, Finland, VTT Electronics and CCC Software Professionals.
- Kiyokawa K, Kurata Y & Ohno H (2000) An optical see-through display for mutual occlusion of real and virtual environments. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 60-67.
- Kiyokawa K, Niimi M, Ebina T & Ohno H (2001) MR² (MR Square): a mixed-reality meeting room. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001), 169-170.
- Kjeldsen R & Kender J (1996) Toward the use of gesture in traditional user interfaces. Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Killington, VT, USA, 151-156.
- Klinker G, Creighton O, Dutoit AH, Kobylinski R, Vilsmeier C & Brugge B (2001) Augmented maintenance of powerplants: a prototyping case study of a mobile AR system. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001), 124-133.
- Klinker G, Reicher R & Brugge B (2000) Distributed user tracking concepts for augmented reality applications. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 37-44.
- Klockar T, Carr DA, Hedman A, Johansson T & Bengtsson F (2003) Usability of Mobile Phones. Proceedings of the 19th International Symposium on Human Factors in Telecommunications, Berlin, Germany, 197-204.

- Kuroda T, Sato K & Chihara K (1998) S-TEL: An avatar based sign language telecommunication system. 2nd European Conference on Disability, Virtual Reality and Associated Technologies, Skövde, Sweden, ECDVRAT and University of Reading, 159-167.
- Kutulakos KN & Vallino JR (1998) Calibration-free augmented reality. *IEEE Transactions on Visualization and Computer Graphics* 4(1): 1-20.
- Kuutti K, Pulli P, Pyssysalo T, Hickey S & Antoniac P (1999) CyPhone Mediaphone Project-Taxi Trip Scenario. 9th International Conference on Artificial Reality and Telexistence (ICAT '99), Tokyo, Japan, 50-52.
- Latva-Aho M (2002) Personal communication.
- LaViola JJ (2000) A Discussion of Cybersickness in Virtual Environments. *SIGCHI Bulletin* 32(1): 47-56.
- Lepetit V & Berger M-O (2000) Handling occlusion in augmented reality systems: a semi-automatic method. *International Symposium on Augmented Reality (ISAR 2000)*, 137-146.
- Leppälä K, Kerttula M & Tuikka T (2003) User Interfaces. *Virtual Design of Smart Products*. IT Press, Oulu, Finland: 72-74.
- Liang S (1999) *Java Native Interface: Programmer's Guide and Specification*. Addison-Wesley.
- Lindholm T & Yellin F (1999) *The Java Virtual Machine Specification*. Addison-Wesley.
- Luff P & Heath C (1998) Mobility in collaboration. *The ACM Conference on Computer Supported Cooperative Work*, Seattle, WA, USA, 305-314.
- Lyytinen K & Yoo Y (2002) Research Commentary: The Next Wave of Nomadic Computing. *Information Systems Research* 13(4): 377-388.
- MacIntyre B & Machado Coelho E (2000) Adapting to dynamic registration errors using level of error (LOE) filtering. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, 85-88.
- March ST & Smith GF (1995) Design and natural science research on information technology. *Decision Support Systems*, 15: 251-266.
- Milgram P & Kishino F (1994) A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems* E77-D(12): 1321-1329.
- Mizell D (2001) Boeing's Wire Bundle Assembly Project. In: Barfield W and Caudell T (eds) *Fundamentals of wearable computers and augmented reality*. Lawrence Erlbaum Associates, Mahwah, NJ: 447-468.
- Mueller F, Agamanolis S & Picard R (2003) Exertion interfaces: sports over a distance for social bonding and fun. *Human factors in computing systems*, Ft. Lauderdale, FL, USA, ACM Press, 561-568.
- Naimark L & Foxlin E (2002) Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker. *Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, 27-36.
- Nam Y & Wohn KY (1996) Recognition of Space-Time Hand-Gestures using Hidden Markov Model. *ACM Symposium on Virtual Reality Software and Technology*, Hong Kong, ACM Press, 51-58.
- Navab N, Bascle B, Appel M & Cubillo E (1999) Scene augmentation via the fusion of industrial drawings and uncalibrated images with a view to marker-less calibration. *Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99)*, 125-133.
- Neumann U & Majoros A (1998) Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. *Virtual Reality Annual International Symposium*, Atlanta, GA, USA, 4-11.
- Newman NJ & Clark AF (1999) An Intelligent User Interface Framework for Ubiquitous Mobile Computing. *International Symposium on Handheld and Ubiquitous Computing (HUC 99)*, Karlsruhe, Germany.
- Nielsen J (1993) *Usability engineering*. Academic Press, Boston, MA, USA.
- Nielsen J (1995) Scenarios in discount usability engineering. In: Carroll JM (ed) *Scenario-based design: envisioning work and technology in system development*. John Wiley & Sons, Inc, New York, NY, USA: 59-83.
- Norman DA (1998) The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution. MIT Press, Cambridge, Mass.: 62-64.

- Pallot M & Sandoval V (1998) Concurrent enterprising: toward the concurrent enterprise in the era of the internet and electronic commerce. Kluwer Academic Publishers, Boston, MA, USA.
- Park J, Jiang B & Neumann U (1999) Vision-based pose computation: robust and accurate augmented reality tracking. Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99), 3-12.
- Pascoe J, Ryan N & Morse D (2000) Using while moving: HCI issues in fieldwork environments. ACM Transactions on Computer-Human Interaction (TOCHI) 7(3): 417-437.
- Pasman W & Jansen FW (2001) Distributed low-latency rendering for mobile AR. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001), 107-113.
- Perry M, O'hara K, Sellen A, Brown B & Harper R (2001) Dealing with mobility: understanding access anytime, anywhere. ACM Transactions on Computer-Human Interaction (TOCHI) 8(4): 323-347.
- Piekarski W, Gunther B & Thomas B (1999) Integrating virtual and augmented realities in an outdoor application. Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99), 45-54.
- Pouwelse J, Langendoen K & Sips H (1999) A feasible low-power augmented-reality terminal. Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99), 55-63.
- Pulli P & Antoniac P (2000) MARISIL-User Interface Framework for Future Mobile MediaPhone. 2nd International Symposium on Mobile Multimedia Systems & Applications (MMSA 2000), Delft, Netherlands, 1-5.
- Pulli P & Antoniac P (2004) Petri Pulli & Peter Antoniac, assignee. User interface. US patent 6,771,294 B1.
- Pyssysalo T, Repo T, Turunen T, Lankila T & Röning J (2000) CyPhone — bringing augmented reality to next generation mobile phones. Proceedings of DARE 2000 on Designing augmented reality environments, Elsinore, Denmark, ACM Press.
- Quek FKH (1996) Unencumbered gestural interaction. IEEE Multimedia 3(3): 36-47.
- Rakkolainen I (2002) Novel Applications and Methods for Virtual Reality. Doctoral Thesis thesis. Tampere University of Technology, Computer Science.
- Rao R & Card SK (1994) The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence (SIGCHI 94). ACM Press, Boston, MA, USA: 318-322.
- Raskar R, Welch G & Chen W-C (1999) Table-top spatially-augmented realty: bringing physical models to life with projected imagery. Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99), 64-71.
- Regenbrecht HT & Specht R (2000) A mobile Passive Augmented Reality Device-mPAR. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000), 81-84.
- Reitmayr G & Schmalstieg D (2001) Mobile collaborative augmented reality. Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001), 114-123.
- Rekimoto J (2001) NaviCam: A palmtop Device Approach to Augmented Reality. In: Barfield W and Caudell T (eds) Fundamentals of wearable computers and augmented reality. Lawrence Erlbaum Associates, Mahwah, NJ: 353-377.
- Rekimoto J & Nagao K (1995) The world through the computer: computer augmented interaction with real world environments. 8th annual ACM symposium on User interface and software technology, Pittsburgh, PA, USA, ACM Press, 29-36.
- Rhodes BJ, Minar N & Weaver J (1999) Wearable Computing Meets Ubiquitous Computing: Reaping the Best of Both Worlds. International Symposium on Wearable Computers (ISWC 2000), 141-149.
- Rodden T, Cheverst K, Davies N & Dix A (1998) Exploiting Context in HCI design for Mobile Systems. Workshop on Human Computer Interaction with Mobile Devices, Glasgow, UK.
- Sasaki H, Kuroda T, Manabe Y & Chihara K (1999) HIT-Wear: A Menu System Superimposing on a Human Hand for Wearable Computers. 9th International Conference on Artificial Reality and Telexistence (ICAT '99), Tokyo, Japan, 146-153.

- Sasaki H, Kuroda T, Manabe Y & Chihara K (2000) Augmented Reality Based Input Interface for Wearable Computers. In: Heudin J-C (ed) *Virtual World*. Springer-Verlag, 1834: 294-302.
- Sasaki H, Kuroda T, Manabe Y & Chihara K (2001) Hand-Area Extraction by Sensor Fusion Using Two Cameras for Input Interface of Wearable Computers. *Scandinavian Conference on Image Analysis (SCIA 2001)*, Bergen, Norway, 779-784.
- Sato Y, Kobayashi Y & Koike H (2000) Fast tracking of hands and fingertips in infrared images for augmented desk interface. *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 462-467.
- Satoh K, Anabuki M, Yamamoto H & Tamura H (2001) A hybrid registration method for outdoor augmented reality. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001)*, 67-76.
- Sauer F, Wenzel F, Vogt S, Tao Y, Genc Y & Bani-Hashemi A (2000) Augmented workspace: designing an AR testbed. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, 47-53.
- Seo Y & Hong KS (2000a) Calibration-free augmented reality in perspective. *IEEE Transactions on Visualization and Computer Graphics* 6(4): 346-359.
- Seo Y & Hong K-S (2000b) Weakly calibrated video-based augmented reality: embedding and rendering through virtual camera. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, 129-136.
- Sexton I & Surman P (1999) Stereoscopic and autostereoscopic display systems. *IEEE Signal Processing Magazine* 16(3): 85-99.
- Shapiro LG & Stockman GC (2001) *3D Object Reconstruction*. Computer vision. Prentice Hall, Upper Saddle River, NJ: 460-468.
- Simon G, Fitzgibbon AW & Zisserman A (2000) Markerless tracking using planar structures in the scene. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, 120-128.
- Smith RD (1998) Essential techniques for military modeling and simulation. *Proceedings of the 30th conference on Winter simulation*, Washington, D.C., USA, IEEE Computer Society Press, 805-812.
- Smith RD (1999) *Military Simulation: Techniques & Technology*. Information & Security: An International Journal 3.
- Starner T (2002) Wearable computers: No longer science fiction. *IEEE Pervasive Computing* 1(1): 86-88.
- Starner T, Auxier J, Ashbrook D & Gandy M (2000a) The Gesture Pendant: A Self-illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring. *International Symposium on Wearable Computing*, Atlanta, GA, USA, 87-94.
- Starner T, Leibe B, Singletary B & Pair J (2000b) MIND-WARPING: towards creating a compelling collaborative augmented reality game. *Proceedings of the 5th international conference on Intelligent user interfaces (IUI2000)*, New Orleans, LA, USA, ACM Press, 256-259.
- Starner T, Mann S, Rhodes B, Healey J, Russell KB, Levine J & Pentland A (1995) Technical Report No. 355, *Wearable Computing and Augmented Reality*, MIT Media Lab Vision and Modelling Group.
- Starner T, Mann S, Rhodes B, Levine J, Healey J, Kirsch D, Picard R & Pentland A (1997) Augmented Reality through Wearable Computing. *Teleoperators and Virtual Environments* 6(4): 386-398.
- State A, Ackerman J, Hirota G, Lee J & Fuchs H (2001) Dynamic virtual convergence for video see-through head-mounted displays: maintaining maximum stereo overlap throughout a close-range work space. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001)*, 137-146.
- State A, Hirota G, Chen DT, Garrett WF & Livingston MA (1996) Superior augmented reality registration by integrating landmark tracking and magnetic tracking. *International Conference on Computer Graphics and Interactive Techniques*, ACM Press, 429-438.

- Stetten G, Chib V, Hildebrand D & Bursee J (2001) Real time tomographic reflection: phantoms for calibration and biopsy. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001)*, 11-19.
- Stratton GM (1896) Some preliminary experiments on vision. *Psychological Review* 3: 611-616.
- Stricker D & Kettenbach T (2001) Real-time and markerless vision-based tracking for outdoor augmented reality applications. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001)*, 189-190.
- Stricker D & Navab N (1999) Calibration propagation for image augmentation. *Proceedings of 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99)*, 95-102.
- Sugihara T, Miyasato T & Nakatsu R (1999) Evaluation of visual fatigue in 3-D displays: Focusing on the mismatching of convergence and accommodation. *IEICE Transactions on Electronics*, Tokyo, Japan, 1814-1822.
- Sutherland I (1965) The ultimate display. *Proceedings of International Federation for Information Processing (IFIP) Congress*, New York, NY, USA, 506-508.
- Sutherland I (1968) A Head-Mounted Three-Dimensional Display. *American Federation of Information Processing Society (AFIPS)*, Washington D.C., Thomson Books, 757-764.
- Szalavári Z, Eckstein E & Gervautz M (1998) Collaborative gaming in augmented reality. *Virtual Reality Software and Technology*, Taipei, Taiwan, ACM Press, 195-204.
- Takagi A, Yamazaki S, Saito Y & Taniguchi N (2000) Development of a stereo video see-through HMD for AR systems. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, 68-77.
- Tang A, Owen C, Biocca F & Mou W (2003) Comparative effectiveness of augmented reality in object assembly. *Human factors in computing systems*, Ft. Lauderdale, FL, USA, ACM Press, 73-80.
- Taylor CJ (1999) Technical Report, Virtual Keyboards, Department of Computer and Information Science, University of Pennsylvania.
- Terveen L, McMackin J, Amento B & Hill W (2002) Specifying preferences based on user history. *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves (SIGCHI 2002)*. ACM Press, Minneapolis, MN, USA: 315-322.
- Tripathi A (2000) *Augmented Reality: An Application for Architecture*. Unpublished Masters thesis. University of Southern California, Faculty of the School of Architecture.
- Tuikka T & Kuutti K (2001) Thinking Together in Concept Design for Future Products-Emergent Features for Computer Support. *The 4th International Conference of Cognitive Technology: Instruments of Mind (CT 2001)*, Coventry, UK, Springer-Verlag Berlin, 40-54.
- Uchiyama S, Takemoto K, Satoh K, Yamamoto H & Tamura H (2002) MR platform: a basic body on which mixed reality applications are built. *Proceedings of International Symposium on Mixed and Augmented Reality (ISMAR 2002)*, 246-256.
- Umeda T, Suenaga T, Kuroda T, Oshiro O & Chihara K (2000) A Real-Time Telemedicine System Using a Medical Ultrasound Image Sequence on a Low Bit-Rate Network. *Japanese Journal of Applied Physics* 39(5B): 3236-3241.
- Väänänen-Vainio-Mattila K & Ruuska S (2000) Designing Mobile Phones and Communicators for Consumers' Needs at Nokia. In: Bergman E (ed) *Information appliances and beyond: interaction design for consumer products*. Morgan Kaufmann Publishers, San Francisco, CA, USA: 169-204.
- Vandewoude Y, Urting D, Pelckmans K & Berbers Y (2002) A Java-interface to digital cameras. *Proceedings of the 20th IASTED International Multi-Conference Applied Informatics: Software Engineering*, Innsbruck, Austria, 113-118.
- Webster A, Feiner S, MacIntyre B, Massie W & Krueger T (1996) *Augmented Reality in Architectural Construction*. *Proceedings of the Third ASCE Congress for Computing in Civil Engineering*.
- Weiser M (1993) Hot topics-ubiquitous computing. *Computer* 26(10): 71-72.
- Weiser M (1994) *Creating the invisible interface*. *Symposium on User Interface Software and Technology*, Marina del Rey, CA, USA, ACM Press.
- Wu Y & Huang TS (1999a) Human hand modeling, analysis and animation in the context of HCI. *Proceedings of the International Conference on Image Processing (ICIP 99)*, 6-10.

- Wu Y & Huang TS (1999b) Vision-Based Gesture Recognition: A Review. In: Braffort A, Gherbi R, Gibet S, Richardson J and Teil D (eds) *Gesture-Based Communication in Human-Computer Interaction: Proceedings of the International Gesture Workshop, GW '99*, Gif-sur-Yvette, France, March 17-19, 1999. Springer-Verlag: 103-115.
- Yin Z, Ding H, Tso SK & Xiong Y (1999) A virtual prototyping approach to mold design. *IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC '99)*, 463 -468.
- Zhang X, Genc Y & Navab N (2001) Taking AR into large scale industrial environments: navigation and information access with mobile computers. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR 2001)*, 179-180.

Appendix



US006771294B1

(12) **United States Patent**
Pulli et al.

(10) **Patent No.:** **US 6,771,294 B1**
(45) **Date of Patent:** **Aug. 3, 2004**

(54) **USER INTERFACE**

(76) Inventors: **Petri Pulli**, Impolankuja 2 A 52,
FIN-90570 Oulu (FI); **Peter Antoniac**,
Peltolantie 8 D 45, FIN-90210 Oulu
(FI)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/473,709**

(22) Filed: **Dec. 29, 1999**

(51) **Int. Cl.**⁷ **G09G 5/00**

(52) **U.S. Cl.** **345/863**; 345/173; 345/835;
345/764; 345/810

(58) **Field of Search** 345/863, 835,
345/764, 810, 173, 155, 156

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,319,747 A * 6/1994 Gerrissen et al. 345/155
5,381,158 A * 1/1995 Takahara et al. 345/156
5,677,700 A 10/1997 Schwalba et al.
5,790,104 A * 8/1998 Shieh 345/173
6,111,580 A * 8/2000 Kazama et al. 345/764
6,346,929 B1 * 2/2002 Fukushima et al. 345/156

OTHER PUBLICATIONS

Sasaki, H., Kuroda, T., Chihara, K., "Hit-Wear: an Input
Interface for Wearable Computers Utilizing Visual Feedback
and Physical Sense," Correspondence on Human Interface,
vol. 1, No. 4, no date.

Kalawsky, Roy S., "The Science of Virtual Reality and
Virtual Environments," Addison-Wesley Publishers, Ltd,
ISBN 0-201-63171-7, pp. 88-135, 188-197, 21-22 and
135-163, no date.

Akao, Yoji, "Quality Function Deployment: Integrating Cust-
omer Requirements into Product Deisgn," Productivity
Press, Cambridge, MA, 1990, ISBN: 0-915299-41-0, pp.
6-7, 20 and 22.

* cited by examiner

Primary Examiner—Kristine Kincaid

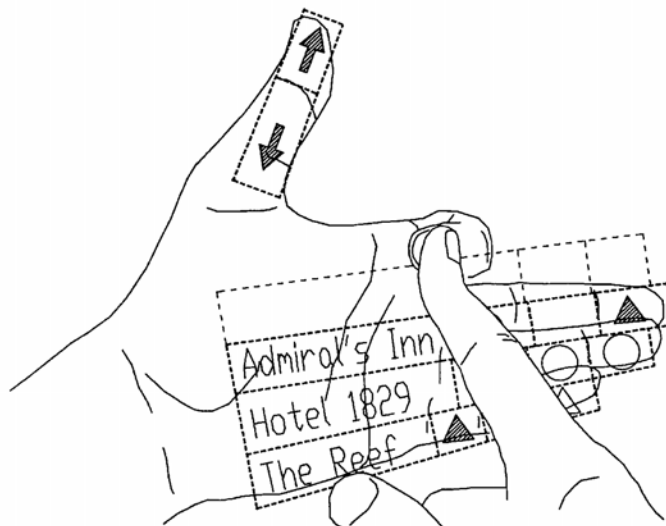
Assistant Examiner—Mylinh Tran

(74) *Attorney, Agent, or Firm*—Pillsbury Winthrop LLP

(57) **ABSTRACT**

The invention relates to a method of using a portable mobile
communication device such as mediaphone. In the method,
an image of a hand of the user is made and the natural
partition of the hand is recognized. An augmented reality
image of a user interface of the device is made by laying an
image of a desired user interface comprising input segments
onto the image of the user's hand in such a way that the
segments of the user interface are separated from each other
by the natural partition of the hand. The augmented reality
image of the user interface is seen by the user, who selects
a desirable segment by touching a partition on the hand. The
selection on the hand is recognized and a function related to
the selected segment is performed by the device. MARISIL
is defined and used to operate this class of devices.

39 Claims, 11 Drawing Sheets



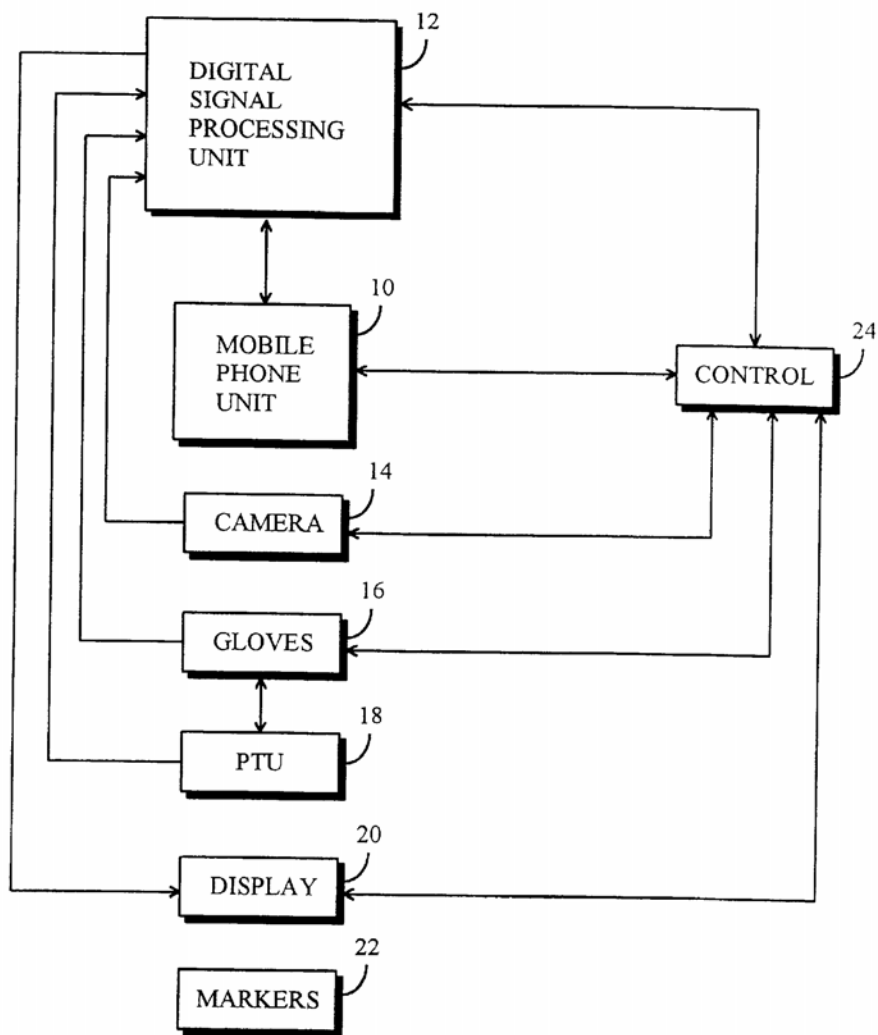


FIG. 1

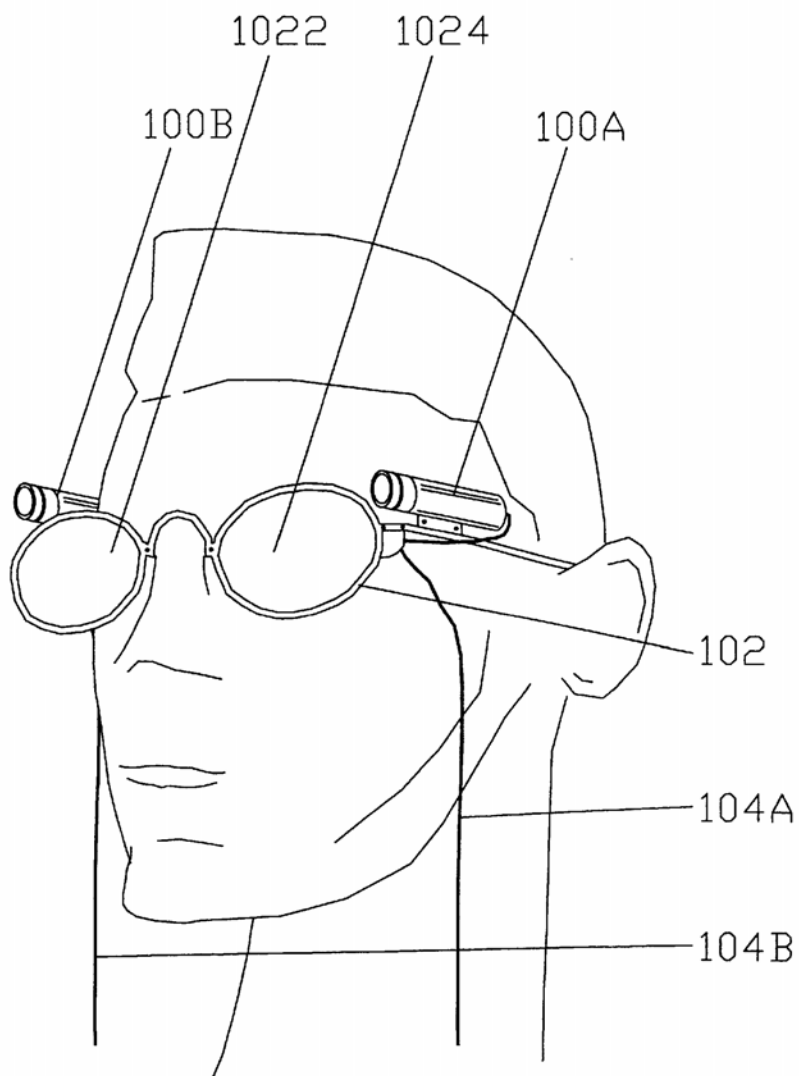


Figure 2A

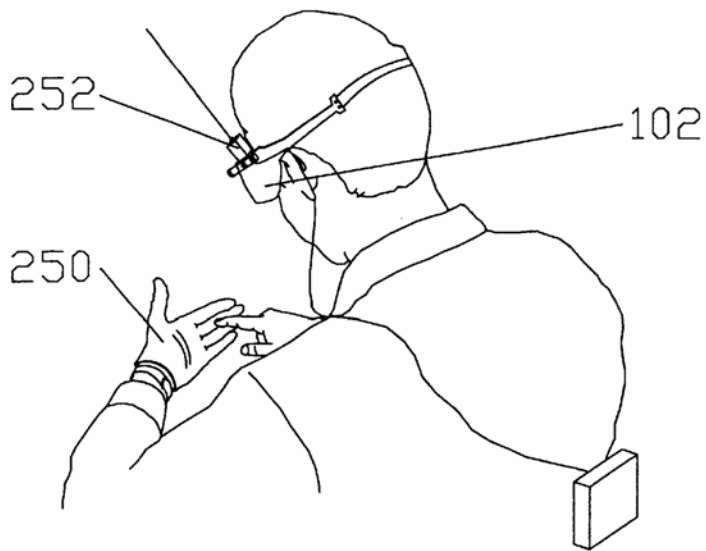


Figure 2B

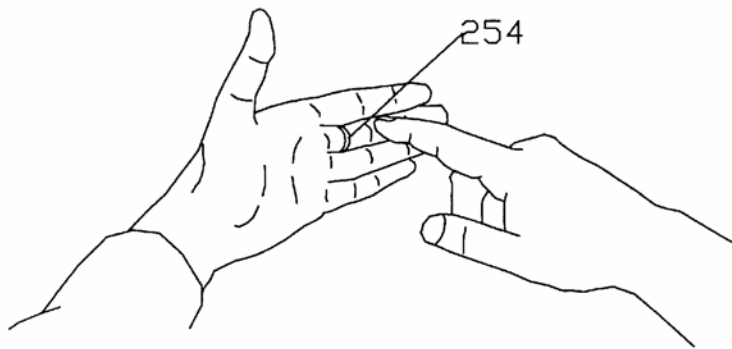
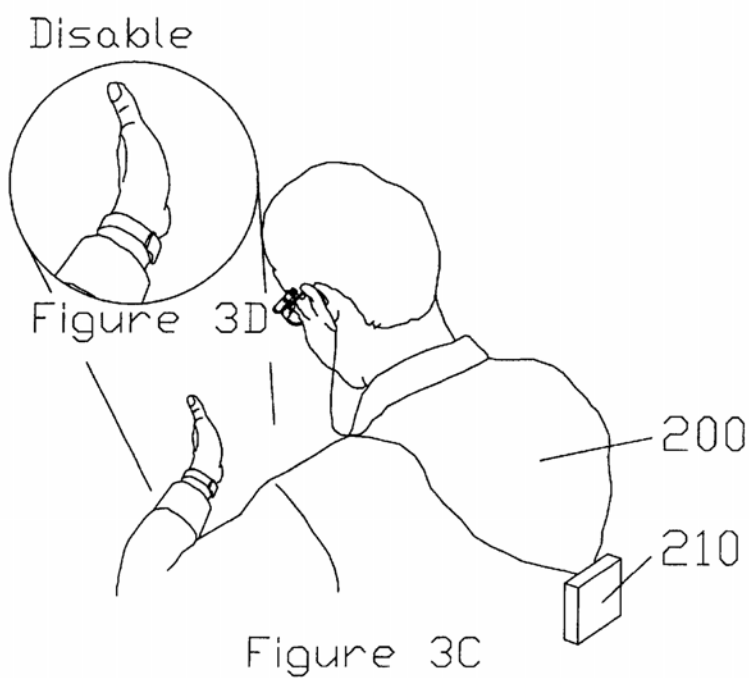
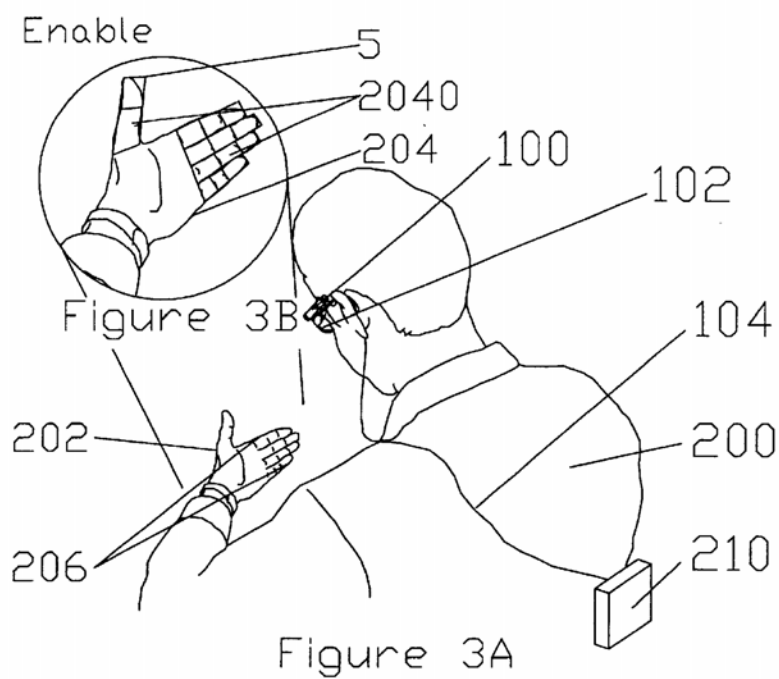


Figure 2C



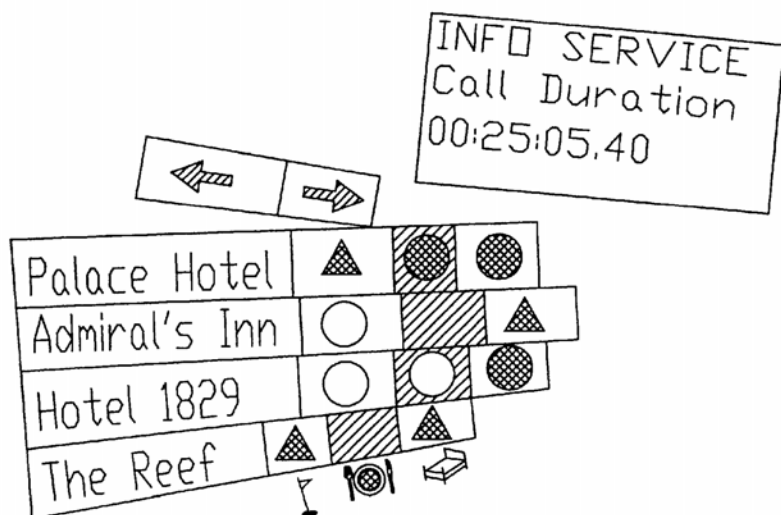


Figure 3E

Figure 4B

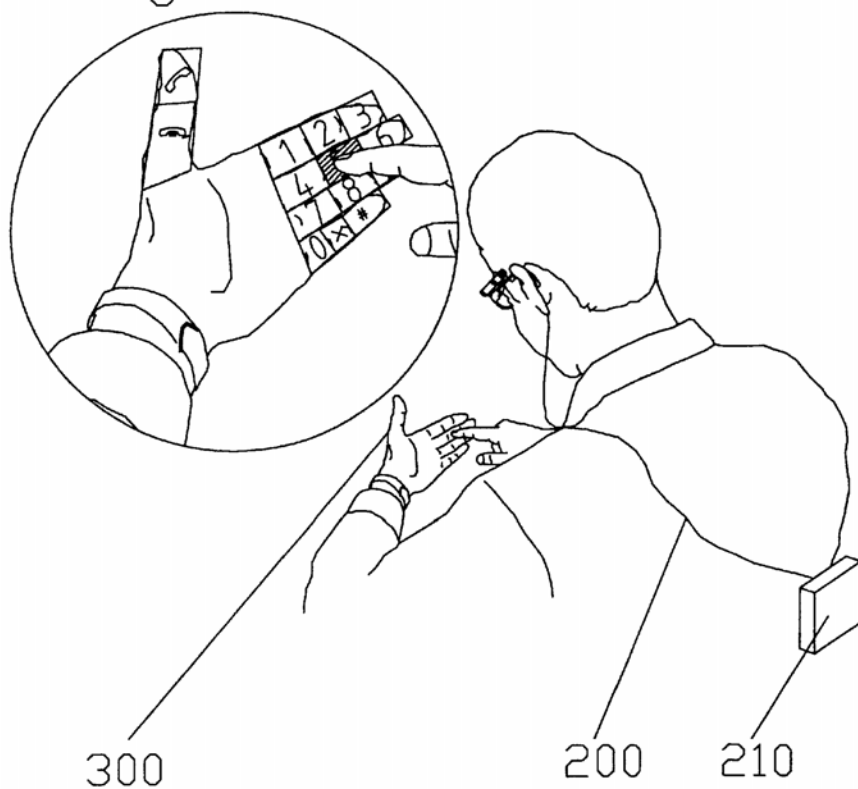


Figure 4A

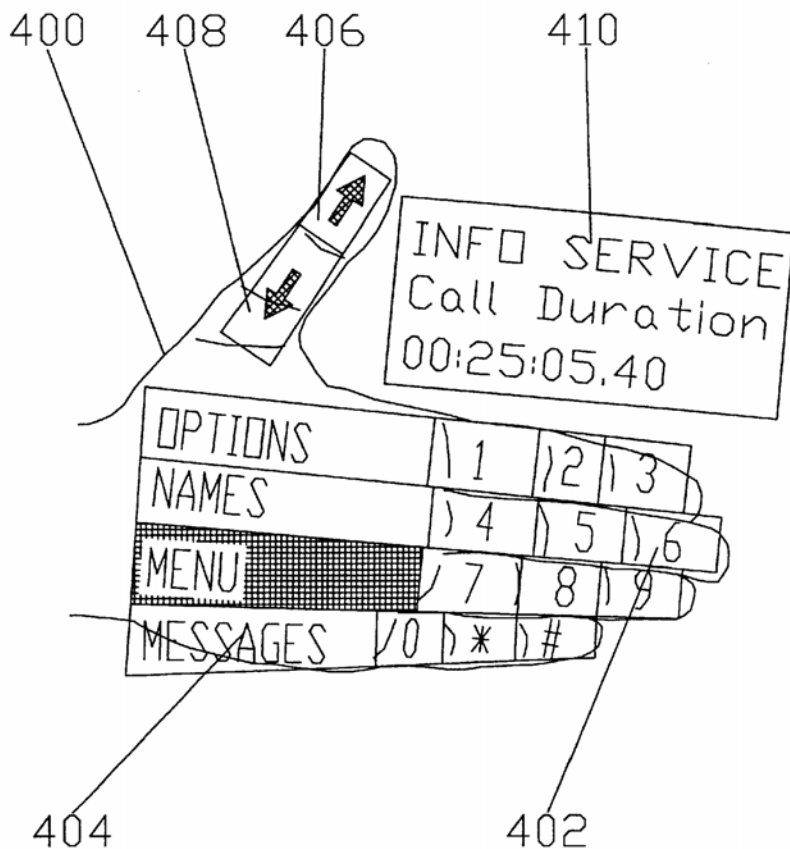


Figure 5

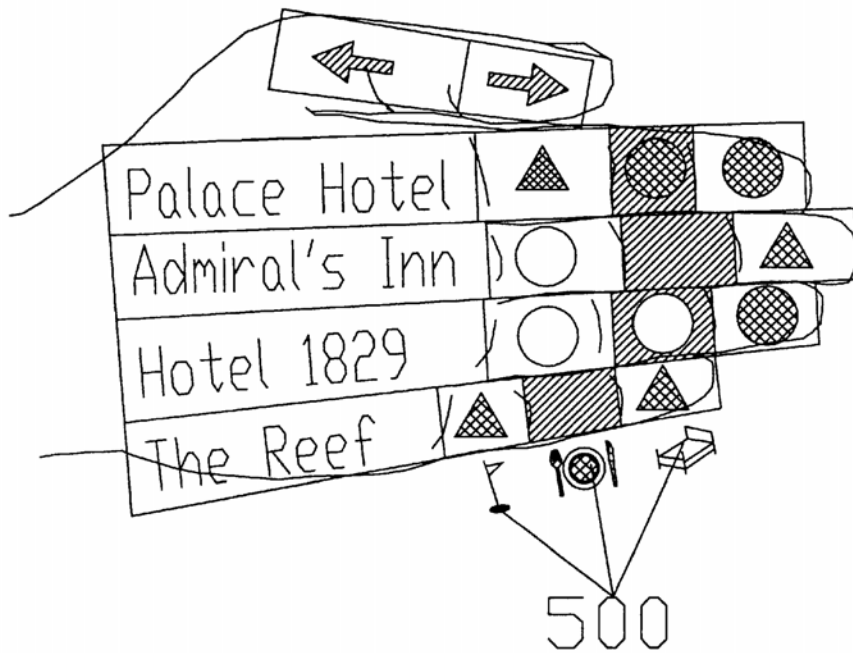


Figure 6

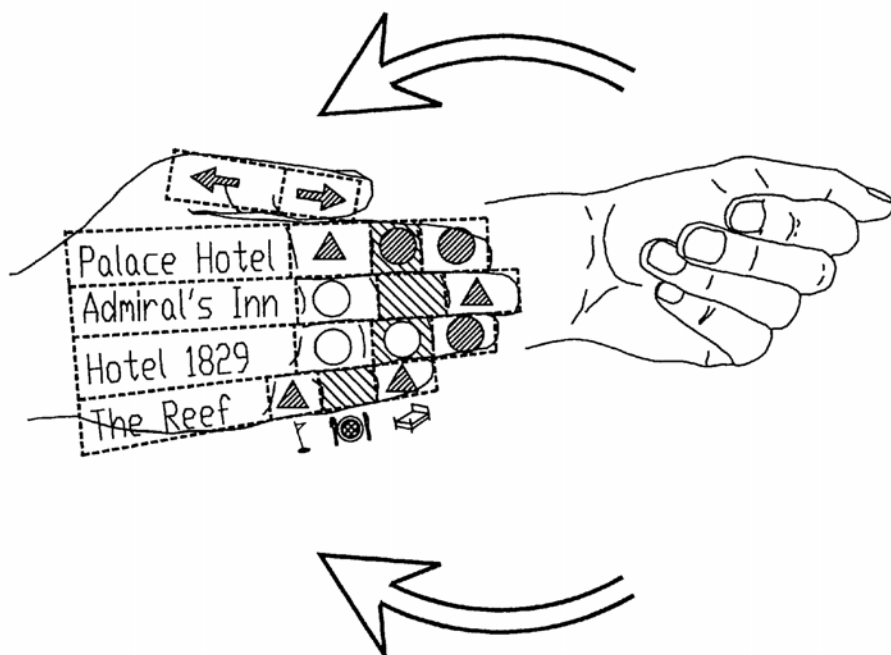


Figure 7

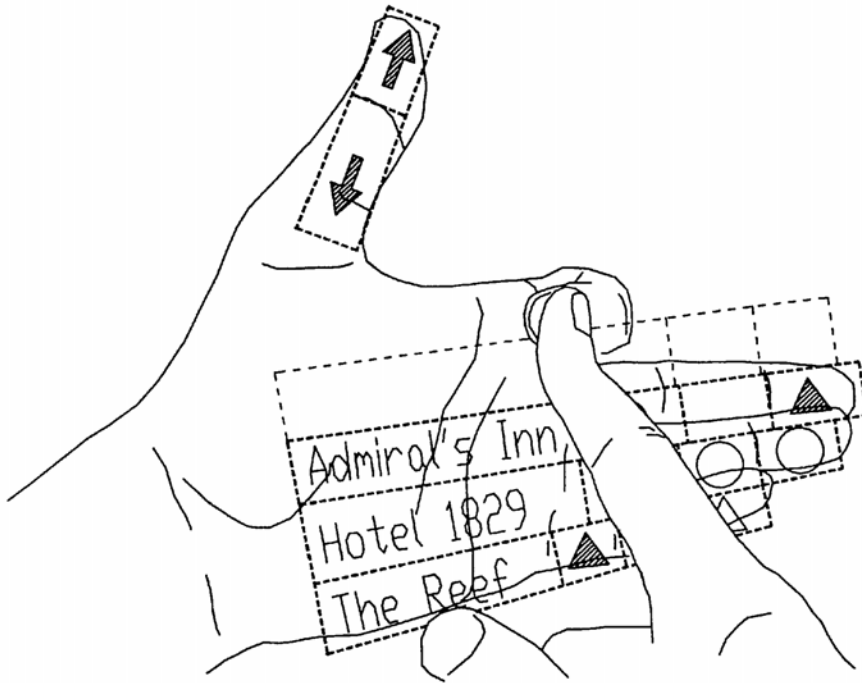


Figure 8

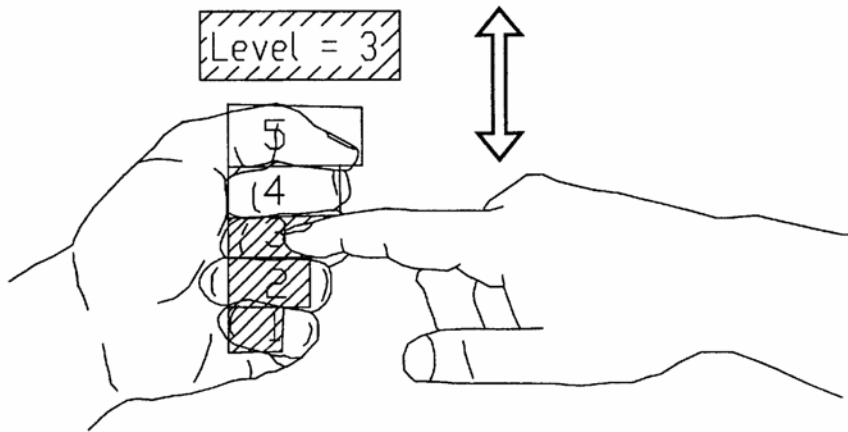


Figure 9

1

USER INTERFACE

BACKGROUND OF THE INVENTION

One of the most important factors contributing to the effective use of a computer and phone devices is the interface between a human and a device. The most popular human-computer interface is a keyboard. A commonly used computer keyboard has nearly 100 keys in order to facilitate the user input, and it is based on a matrix of electrical switches that are called keys. A keyboard for small devices has a reduced size and the keys have a special property. Pressing a key more than once produces a different action on the interface. Although such keyboards are commonly used and widely accepted, they do not often provide a very efficient or easy way to use electronic devices.

Mobile devices, such as cellular phones or palm top computers, have become very popular in the last few years. These devices are hand-held and they are provided with a keyboard for user input and with a display for showing the results of the inputs. However, users do not always want to carry mobile devices in their hands, nor is it always easy to input data into the devices.

A prior art arrangement to avoid the use of a keyboard is disclosed in "HIT-Wear": Input Interface for Wearable Computers Utilizing Visual Feedback and Physical Sense by Sasaki, H., Kuroda, T., Chihara, K. in Correspondence on Human Interface, Vol. 1, No. 4, which is incorporated herein by reference. In this arrangement, a hand is placed in front of the face and the device identifies the fingertips. The user is wearing a head-mounted display, which shows him a menu related to the use of the device, such that one function key of the menu is displayed outwards from each fingertip. When the user touches a particular key, the corresponding function is carried out. However, a drawback of the arrangement is that even though it enables rapid selection of a desired command or operation, only a few commands or operations can be shown at a time to the user. Another drawback is that the arrangement is not really suitable for inputting alphanumeric or other symbolic data.

SUMMARY OF THE INVENTION

An object of the invention is to provide a method and a device implementing the method, such that augmented reality can be used to realize a compact user interface in a matrix form, with which interface commands and symbolic data can be input into the device in a simple manner. This is achieved with a method disclosed in the introduction of using a portable mobile communication device, the method comprising: forming an image of a hand of the user and recognizing the natural partition of the hand; forming an augmented reality image by overlaying an image of a desired user interface on the image of the hand recognized visually by the user; selecting a desired segment by touching a suitable partition on the hand; recognizing the selection on the hand; and performing a function related to the segment in the device.

The invention also relates to a method of using a portable mobile communication device, the method comprising: forming an image of a hand of the user and recognizing the natural partition of the hand; forming an augmented reality image of a user interface of the device by overlaying an image of a desired user interface on the image of the hand; showing a user the augmented reality image of the user interface; selecting a desired segment by touching a suitable partition on the hand; recognizing the selection on the hand; and performing a function related to the segment in the device.

2

The invention further relates to a portable mobile communication device comprising a video camera unit, a display unit, a transceiver unit and a digital signal processing unit that is arranged to control the operation of the portable mobile communication device: wherein the video camera is arranged to form a video image of a user's hand and to feed the image into the digital signal processing unit, which is arranged to recognize the natural partition of the hand in the video image; the digital signal processing unit is arranged to feed an image of a user interface into the display and to form an augmented reality image of the user interface of the device by overlaying an image of a desired user interface on the image of the user's hand seen by the user; a desired segment is selected by the user by touching a suitable partition on the hand; the digital signal processing unit is arranged to recognize the selection on the hand; and the device is arranged to perform a function related to the segment.

The invention also relates to a portable mobile communication device comprising a video camera unit, a display unit, a transceiver unit and a digital signal processing unit that is arranged to control the operation of the portable mobile communication device, wherein: the video camera is arranged to form a video image of a user's hand and to feed the image into the digital signal processing unit, which is arranged to recognize the natural partition of the hand in the video image; the digital signal processing unit is arranged to form an augmented reality image of a user interface of the device by overlaying an image of a desired user interface on the image of the user's hand; the digital signal processing unit is arranged to feed the augmented reality image of the user interface into the display, which is arranged to show the image to the user; a desired segment is selected by the user by touching a suitable partition on the hand; the digital signal processing unit is arranged to recognize the selection on the hand; and the device is arranged to perform a function related to the segment.

The invention further relates to a portable mobile communication device comprising a virtual reality gloves unit, a position tracking unit, a display unit, a transceiver unit and a digital signal processing unit that is arranged to control the operation of the portable mobile communication device, wherein: the virtual reality gloves are arranged to feed information on the user's hand into the digital signal processing unit, which is arranged to form an image of the hand and to recognize the natural partition of the hand; the digital signal processing unit is arranged to form an augmented reality image of a user interface of the device by overlaying an image of a desired user interface on the image of the user's hand; the digital signal processing unit is arranged to feed the augmented reality image of the user interface into the display, which is arranged to show the image to the user; a desired segment is selected by the user by touching a suitable partition on the hand; the digital signal processing unit is arranged to recognize the selection on the hand; and the device is arranged to perform a function related to the segment.

The preferred embodiments of the invention are disclosed in the dependent claims.

An essential feature of the invention is that the user interface is implemented without any actual physical element. The actual physical part of the user interface is replaced by the user's own hand. The portable electronic device is operated by pointing at input segments which are shown on the user's hand. This operation is implemented by means of augmented reality where an image of the hand seen by the user is supplemented with a display for the informa-

tion of the user interface. The user interface information includes the menu and the results, which are visible to the user in the form of segments. The segments are preferably displayed on the hand according to the symbolic partition of the hand, i.e. how the contours of the hand and the joints of the fingers divide the hand into symbolic sectors.

The method and the system according to the invention provide several advantages. Even though a physical user interface is not needed to carry out the functions of the device, it is possible to utilize the input of symbolic or alphanumerical data or data based on other characters. When numerals or symbols are used, the operation of the device is independent of the user's language.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described below in more detail in connection with preferred embodiments with reference to the attached drawings, in which

FIG. 1 shows a block diagram of the device,

FIG. 2A shows camera and display units of the device,

FIG. 2B shows image forming unit, gloves and display unit of the device

FIG. 2C shows a marker,

FIG. 3A shows a user looking at his hand,

FIG. 3B shows a user interface that is visible to the user,

FIG. 3C shows a situation where the user has disabled the device,

FIG. 3D shows a user interface disabled by the user,

FIG. 3E shows a situation where a hand is not a basis of a user interface,

FIG. 4A shows a user selecting numbers,

FIG. 4B shows a situation seen by the user when he selects the numbers,

FIG. 5 shows selection segments visible on the hand,

FIG. 6 shows the use of QFD symbols in the user interface,

FIG. 7 shows one gesture of sign language,

FIG. 8 shows one gesture of sign language,

FIG. 9 shows one gesture of sign language.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention can be applied as a user interface between mobile hand-held devices and a user. The arrangement provides easier access to information and facilitates browsing. The arrangement according to the invention is based on augmented reality (AR), which can be implemented by means of a video camera, a head-mounted display (HMD) (such as AR glasses) and a pointer.

The inventive interface is based on augmented reality. The basic idea is that the user sees a matrix of information segments overlayed on the image of his hand and the user is able to interact with the contents of the matrix by touching the matrix segments that are superimposed on the image of the hand.

Examine at first a block diagram of the device. The device comprises a mobile phone unit 10, which comprises a transceiver, a digital signal processing unit 12, a camera unit 14, a unit of virtual reality gloves 16, a position tracking unit 18, a display 20, marker(s) 22 and a control unit 24. The mobile phone unit 10 represents a conventional mobile phone comprising a digital signal processing unit 12. In the arrangement according to the invention, the camera 14, the

virtual reality gloves 16 and the position tracking unit 18 are alternative components of the device. The camera unit 14, which comprises at least one camera, or units 16 and 18 produce information about the user's hand for the formation of an image of the user interface and the hand. The display 20 shows the user at least the image of the user interface, which is visible to the user as superimposed on the hand. There can be one marker 22 or many markers 22 that can be active or passive. Markers are used to help to mark and detect a certain part of the hand. A marker can be passive and then it is detected by the camera. A marker can also be active when it transmits a signal (i.e. electromagnetic signal or acoustic signal) to the position tracking unit. The entire device is controlled by the control unit 24.

FIG. 2A shows the units of the device which are related to the invention and which form an image of the user's hand. Cameras 100A and 100B film the space in front of the user's face where the hands are usually situated. Instead of two cameras it is also possible to use only one camera. The cameras 100A and 100B are preferably conventional CCD cameras or the like, where the detecting surface is a matrix consisting of pixels. The cameras are preferably colour cameras, but it is also possible to use monochrome cameras in the arrangement according to the invention. The location of the user's hand is monitored in the digital signal processing unit from an image formed by the camera(s). The user forms an image of his hand with his own eyes after the rays of light originating from the hand have passed through a display 102. This viewed image of the hand is supplemented with augmented reality by means of the display 102, the augmented reality comprising an image of the user interface formed by the signal processing unit. Another alternative is that the images which have been formed by the cameras and supplemented with augmented reality in the digital signal processing unit are shown to the user by means of a head-mounted display 102. When the hand moves, the digital signal processing unit follows the movement in the image formed by the camera(s) and shifts the augmented reality similarly in order that the user sees the hand and the augmented reality moving at all times at the same relation to each another. Both the camera 100 and the display 102 are connected by a cable 104 to a mobile station, which comprises a digital signal processing unit for the images formed by the camera(s) and for the computer graphics of the augmented reality.

The interface comprises several options. Examine first full augmented reality interaction, where the interface comprises two small cameras 100A and 100B, which enable formation of a stereo image. The cameras are preferably installed in the user augmented reality eyeglasses 102. Low-resolution cameras can be used as the video cameras 100A and 100B. The augmented reality glasses 102 are a device that is based on laser technology and light reflection on a partially transparent surface. An image of the user interface is reflected as an augmented reality image from a reflector surface 1022 into the right eye, and the augmented reality, which consists of the image of the user interface, is reflected from a reflector surface 1024 into the left eye. Another alternative is that an image of the hand formed by camera 100B and supplemented with augmented reality is reflected from reflector surface 1022 into the right eye, and the image of the hand formed by camera 100A and similarly supplemented with augmented reality is reflected from reflector surface 1024 into the left eye. An image can be shown either in stereo (to both eyes) or mono (to one eye only). If the user cannot see straight through the AR glasses, the image that is shown to the user by the AR glasses 102 of

the hand, the user interface and the surroundings preferably corresponds rather accurately to the image the user would see with his own eyes in the same situation. The eyeglasses for augmented reality are described in greater detail in *The Science of Virtual Reality and Virtual Environments*, Roy S. Kalawsky, Addison-Wesley Publishers Ltd., ISBN 0-201-63171-7, pages: 88-135, which is incorporated herein by reference. Since there are two cameras **100A** and **100B**, triangulation can be used in the digital signal processing to determine the distance to the hand when the angles of view of the cameras, the hand's location in the image of the cameras and the distance between the cameras are known. The contours of the hand are also determined in the digital signal processing in a manner known per se. This makes it easier to monitor the hand's movements and to adapt the image of the user interface to the image of the hand. The formation of the contours of the hand is described in greater detail in FIG. 6. Instead of the AR glasses **102**, it is also possible to use an HMD (a visor display).

Instead of the full augmented reality interaction it is possible to use partial augmented reality interaction. FIG. 2B shows a device, which comprises a head-mounted display **102**, virtual reality gloves **250** and a position tracking device **252**. These are described in more detail in *The Science of Virtual Reality and Virtual Environments*, Roy S. Kalawsky, Addison-Wesley Publishers Ltd., ISBN 0-201-63171-7, pages: 88-135, pages: 188-197 and pages: 21-22 and 135-163, which are incorporated herein by reference. The position tracking device monitors the hand's movement and location, on the basis of which the image of the user interface is placed on the image of the hand. The HMD forms the augmented reality in the tracked positions of the hands with gloves.

As shown in FIG. 2C, another alternative of implementing the partial augmented reality interaction is that the interface comprises a head-mounted display, similarly as the previous alternative, but instead of gloves the user wears a camera and has some location or position tracking markers **254** on the hands. The markers are not necessarily seen by the naked eye. The camera, however, detects the markers and the digital signal processing unit builds the interface over the image of the hand as in the previous cases. For example, the nail of the forefinger on the pointing hand can be painted with a strong colour. The colour is preferably such that it is clearly visible in the infrared zone, but it cannot be detected in visible light or by the naked eye. It is also possible to use a separate additional element or an ornament in the fingertip. Furthermore, a separate pointer element (such as a signal bar which resembles a pencil) can be used. Also, the hand which is the basis for the user interface can be made more noticeable for example by means of a ring or some other sign or ornament, since the pattern recognition process of the digital signal processing unit can identify it without problems. The left hand is useful as a basis of the user interface since most people already wear a ring on the ring finger and the pattern recognition process identifies the ring as a location marker.

Examine below in greater detail the method according to the invention with reference to FIGS. 3A, 3B, 3C and 3D. FIGS. 3A and 3C show how an outside observer sees the user **200** and the device. FIGS. 3B and 3D show how the user **200** sees his hand. The operation of the device is activated when the hand **202** is moved in front of the user **200** such that the perpendicular of the palm approximately faces the eyes of the user **200**. In such a case the digital signal processing unit identifies the hand from an image formed by the camera. When the user **200** looks at his hand **202**, he sees that the image **204** of his hand **202** formed

either by his eyes or by the video camera **100** is superimposed by a menu or some other segmented information produced by the portable electronic device. Other people cannot see the information overlaid to the image **204** of the hand, since the information (menu) is not actually reflected in the hand **202** but it is formed on top of the image **204** of the hand produced either by the user's eye or by the video camera **100**. When the user **200** can see the menu, he is able to point at a desired segment of the menu in his hand **202** for example with his finger in order to produce a desired function.

Moving the hand **202** sideways towards the video camera **100** erases the menu from view in FIG. 3D and prevents the use of the portable electronic device. FIGS. 3C and 3D show an example of how the operation of the device is prevented by turning the hand sideways so that the graphical image **2040** of a desired user interface cannot be laid over the hand. When the hand is positioned sideways, the result of the operation (or the menu) can be shown to the user **200** in the form of a matrix or table, and the information (information segments) are visible to the user **200** even if the hand **202** is positioned sideways. However, the portable electronic device cannot be used before the hand **202** is turned such that it forms the basis of the information to be shown. Removing the hand from the visual range of the video camera **100** stops the display operation of the portable electronic device. In such a case the digital signal processing unit no longer identifies the hand from the image formed by the camera.

In the method a video image **204** of a hand **202** of the user **200** is formed and the natural partition **206** of the hand is recognized. The natural partition of the hand means that the hand is divided mainly into parts **206** that are defined by the joints of the fingers and the outer contours of the hand. The forefinger, the middle finger, the ring finger and the little finger are each divided into three parts by the joints. The thumb is divided into two parts. Since the skin is creased at the joints, the locations of the joints can be identified as the borders of the parts by a graphics programme in the digital signal processing unit **210** of the device. The partition and the adaptation of the matrix segments are shown in greater detail in FIGS. 4A and 4B. An augmented reality image of a user interface of the device is formed by overlaying an image **2040** of a desired user interface on the image of the user's hand. The segments of the user interface are placed on the image of the hand in such a way that the segments of the user interface are separated from each other by the natural partition of the hand. In an embodiment of the invention, the user sees his hand through the display, which forms an image of the user interface over the image of the hand in the user's eye. In another embodiment of the invention the user does not see through the display, but before the image of the hand is shown to the user it is supplemented with the image of the user interface, and the augmented reality image of the hand and the user interface formed in this manner is shown to the user. The user can select any segment according to his will by touching a suitable partition on the hand. The signal processing unit **210** of the device recognizes the selection on the hand and performs the function related to the segment in the device.

The user can also remove the hand and still be able to see the display as shown in FIG. 3E. In this way the user can see the interface in case he wants to see static or dynamic information that does not need interaction (i.e. pictures, movies, etc.).

The improved user interface is for a mobile device like a cellular telephone subscriber terminal or a palm top com-

puter or any other handy device which has a processor and is light enough to be carried by a user. One example of these devices is mediaphone that is a portable device capable of transmitting and receiving speech, data and video images over wireless connection. The interface includes the following features: (1) a table browser interface based on a table-driven browser that enables a high information content in a small space; (2) an extended virtual display for use by hand, enabling a display of information for the user outside the area of the hand; (3) an unfixed panel based on the user's hands that enable input without making any distinction in handedness; (4) a sign language providing different browsing extensions for use in the interface in order to speed up the input and to make the use more ergonomic; (5) easy pointers for browsing the interface; (6) secrecy of user input; (7) easiness of updating for use in a flammable environment.

The table browser interface is based on a table-driven browser and the interface has a high information content in a small space. Examine this arrangement below with reference to FIGS. 4A and 4B. FIG. 4A shows how an outside observer sees the user 200 using the device, and FIG. 4B shows an image of the hand 300 seen by the user 200. When the four fingers of the hand 300 are together, an image of the fingers can be superimposed by an augmented reality matrix of 4x3 segments, which is adapted to the three parts of each finger. The matrix segments can be, for example, the number buttons and buttons * and # of the phone. The image of the thumb can be superimposed by an augmented reality image of start and end buttons for a call. Since the matrix rows are of different lengths similarly as the fingers, the different columns can be shown with different colours for the sake of clarity.

In the arrangement according to the invention, the user sees a segmented menu over the image of the hand, and he can select from the menu a desired function for example by pointing with a finger of the other hand. The segmentation follows the contours of the hand and is divided into parts similarly as the fingers are divided into parts by the joints. If the user wants to make a call, he can use one finger of his other hand or even the thumb of the same hand to point at the numbers shown over his four fingers. The numbers are visible between the joints in a similar manner as the numbers in a conventional phone can be selected from a 4x3 matrix. A call is started and ended by touching the symbols shown on the thumb. The numbers and the other information are visible on the hand as described above, since the graphics programme of the portable electronic device, which is known per se and therefore obvious to those skilled in the art, searches for the characteristic features of the hand's contours, such as the spaces between the fingers and the fingertips (the alignment marks are shown in the figure), and places the menu over the image of the hand. For example, when a finger of the other hand points at a function in the menu, the video camera detects the pointing finger and the graphics programme activates the desired function in the portable electronic device. The desired function is selected for example similarly as in a conventional telephone, i.e. for instance number 3 is pressed with a finger (the tip of the forefinger is pressed with a finger of the other hand or with the thumb of the same hand). Similarly, if the user wants to browse through messages, he presses the lower part of his palm (messages) with his finger, whereafter a message processing programme is activated in the same manner as in the prior art mobile phones. The portable electronic device identifies the location of the pointing finger on the hand, and the operation of the device can thus be controlled. Additional information relating to the call or some other operation can

also be shown outside the field of image of the hand (e.g. the number to be dialled, the duration of the call, etc.).

FIG. 5 shows an arrangement where an image 400 of the hand is superimposed by an augmented reality image consisting of not only the number buttons 402 of the phone but also of selection option segments 404, which are restricted to the part of the fingers on the palm side and placed on the image of the palm. On the image of the thumb there are arrow keys 406 and 408, which can be used to select a suitable selection option in the up/down direction. In such a case the thumb is in an upright position. Other possible selection options include a person's name list, the general menu of the device, and the messages. The image seen by the user also comprises an extended virtual display 410. A virtual display 410 is linked to the hand so that it can be placed anywhere in space except on the hand. The display location is preferably situated between the thumb and the other fingers. The display will be enabled when the user requests for it or the application (device) needs it (like error, call etc).

FIG. 6 shows a different menu of the user interface. The menu is a table that shows different attributes for the columns. The attributes are standard symbols. Clicking on a column reorders the column with the specified column content. Different QFD symbols (Quality Function Deployment) are used. The QFD symbols show the extent to which a specified row is related to the column symbol (which represents a classification). In this way the user can browse the information accordingly. That is explained more in Quality function deployment: integrating customer requirements into product design, Akao, Yoji, Productivity Press, Cambridge, Mass. 1990, ISBN: 0-915299-41-0, Page: 6-7, 20 and 22, which is incorporated herein by reference.

The arrangement of FIG. 6 can be used, for example, when a user travels to a new place. The user can be interested in travelling, food and night life, as shown by a row of symbols 500 below the forefinger in FIG. 6. In such a case it is possible to utilize for example a QFD-based (Quality Function Deployment) micro browser and a symbol display. When the device is a cellular phone according to the invention, the phone retrieves information about the place for example from the Internet, and thereafter compares the obtained information to the interests of the user and indicates the result of the comparison with a few symbols. The best result is denoted for example by a white circle, the second best by a circle marked with lines and a weak result is denoted by a triangle. In other words, when the correlation between the desired service and the interests of the user is high, the display shows a white circle, as seen in the intersection of the first row (Palace Hotel) and the second column (food). In this embodiment the thumb is in a horizontal position, which means that augmented reality arrows situated on the image of the thumb can be used to browse the columns left or right.

The efficiency of the arrangement according to the invention is highly dependent on the Mobile Augmented Reality Interface Sign Interpretation Language (MARISIL) that provides different browsing extensions for use in the interface. MARISIL sign language is defined here in this patent application. The sign language speeds up the input. In order to make the use more ergonomic, it is possible to use a combination of gestures that are detected by the computer as extra inputs. In the following, the most important gestures of the sign language will be discussed.

Enabling the interface is defined by a gesture of placing the palm panel in front of the face at a distance from which

it is possible to normally read a text. The palm is defined as a palm panel on which the interface is laid using augmented reality techniques. The area is determined by the cameras or by augmented reality equipment. Disabling the interface is defined by removing the palm panel from the sight of the camera.

The undo command is implemented when the user is doing a fast disable and enable gesture, i.e. when he removes the palm panel hand and replaces it in the area. Doing faster than the camera frame grabbing speed the gesture might not be detected.

Selecting the items to be displayed in the interface is defined by placing the pointer on the item. The item is defined as a virtual object that has an augmented reality image superimposed on the palm panel. The "item" exists only in the augmented reality world. Only the user sees the interface. Other users cannot see the contents of the interface unless they have access to the interface.

An easy pointer for browsing the interface is for example the forefinger of the opposite hand. The opposite hand refers to the user's hand that is other than the palm panel hand. The pointer is not restricted to the forefinger.

Deselecting the items displayed as selected in the interface is defined by placing the pointer over the selected items.

Multiple selection of the items displayed in the interface is defined by the gesture of placing the forefinger and thumb, which are linked at their fingertips, on the item. The "item" exists only in the Augmented World and will be selected and included into the list of selected items. Only the user sees the interface. Other users can not see the content of the interface unless they have access to the interface.

Multiple deselection of the items selected in the interface is defined by the gesture of placing the forefinger and the thumb, which are linked at their fingertips, on the item that is selected and will be deselected from the list after this gesture.

Opening the items displayed in the interface is defined by the gesture of opening and closing the palm panel in a predefined time (FIG. 7). The predefined time has to be shorter than the predefined time for closing items.

Closing an open item displayed in the interface is defined by the gesture of opening and closing the palm panel in a predefined time. The predefined time has to be longer than the predefined time for opening items.

Starting a phone call in the interface is defined by the gesture of placing the pointer over the lower part of the thumb of the panel palm. The thumb is in a phone mode. The thumb is divided into segments as shown in 4A and 6, which means that lower segment means start the call and upper segment means close the call. The call will be directed to the selected number.

Closing a phone call in the interface is defined by the gesture of placing the pointer over the lower part of the thumb of the palm panel. The interface is in phone mode. Therefore the thumb has to be overlaid with segments indicating the phone mode (FIG. 4A).

Up and down commands are shown in FIG. 5. The up command is defined by the gesture of placing the pointer over the upper part of the thumb that is perpendicular to the palm. The thumb is the one corresponding to the palm panel. The thumb has to be overlaid with segments indicating a browsing up/down mode (i.e. the thumb is perpendicular to the palm). The down command is defined by the gesture of placing the pointer over the lower part of the thumb that is perpendicular to the palm and the down command is used for

moving a list of items downward. The thumb corresponds to the palm panel. The thumb has to be in the browsing up/down mode (i.e. the thumb is perpendicular to the palm).

Right and left commands are shown in FIG. 6. The right command is defined by the gesture of placing the pointer over the upper part of the thumb that is parallel to the palm, and it is used for moving a list of items to the right. The thumb is the one corresponding to the palm panel. The thumb has to be in a browsing left/right mode (i.e. the thumb is parallel to the palm). The left command is defined by the gesture of placing the pointer over the lower part of the thumb that is parallel to the palm, and it is used for moving a list of items to the left. The thumb is the one corresponding to the palm panel. The thumb has to be in the browsing left/right mode (i.e. the thumb is parallel to the palm).

Fast key selection is defined by the use of the thumb of the palm panel to point certain segments on the other fingers of the same hand. If the thumb is overlaid with segments indicating the phone mode, the user will make a speed call by using a well-known speed-dial method using one number to index table to contain full phone number as it is commonly used in mobile phones. If the thumb is overlaid with segments indicating the browse mode, the user is able to select a specific symbol to obtain details or to get help. If it is in the writing mode, the user is able to select letters by using this method.

Removing a row is defined by contracting the corresponding finger of the row matrix (FIG. 8). The operation could be difficult for some users, therefore we recommend the use of the pointer instead. With this command the user can deselect one row or put it in some other table. The result of this action depends on the interface that the user is using (i.e. if the user is in a phonebooktype interface, the action could mean the selection of the preferred phone to make a call, etc.).

The command of setting up a level is defined by making a fist with the palm panel so that the outside part of the finger is perpendicular to the user's face (FIG. 9). This action enables a level interface. The user can select a proper level by choosing one of the four backsides of the fingers from the palm panel.

By definition the interface is not viewable by any other person than the user. The user can select the secured version of the interface and the interface can have the numbers or letters place randomly on the palm panel such that other people do not know the real value introduced by placing the pointer finger on the palm panel. That is described in a U.S. Pat. No. 5,677,700, which is incorporated herein by reference.

The device is easy to update for use in a flammable environment. If the parts of the interface are encased in a plastic cover and closed hermetically, the interface can be used under water. Or the interface can be used in other environments, such as gas stations, where there is an imminent danger of a shortcut causing an explosion. The interface can also be used in environments where the user's hands cannot be uncovered (i.e. astronauts or deep-sea scuba divers).

What is claimed is:

1. A method of using a portable mobile communication device, the method comprising:

forming an image of a physical hand of a user and recognizing a natural partition of the user's physical hand;

forming an augmented reality image by overlaying an image of a desired user interface on an image of the user's physical hand formed by rays of light originating from the user's physical hand and entering the user's eyes;

selecting a desired segment by touching a suitable partition on the user's physical hand, the segment being recognized as a natural partition of the user's physical hand; and

performing a function related to the segment in the device. 5

2. A method according to claim 1, wherein the desired segment is selected by the user.

3. A method according to claim 1, wherein the image of the desired user interface, overlaid on the image of the user's physical hand recognized by the user with his eyes, is formed by means of digital signal processing from data provided by virtual reality gloves. 10

4. A method according to claim 1, wherein the user interface is an alphanumerical user interface.

5. A method according to claim 1, wherein the user interface is an interface based on QFD symbols. 15

6. A method according to claim 1, wherein the image of the user's physical hand is formed as a video image.

7. A method according to claim 1, wherein the image of a desired user interface consists of computer graphics. 20

8. A method according to claim 1, wherein the user interface comprises input segments and is situated on the image of the user's physical hand in such a way that the segments of the user interface are separated from each other by the natural partition of the user's physical hand.

9. A method according to claim 1, wherein the device is operated by means of MARISIL sign language.

10. A method of using a portable mobile communication device, the method comprising:

forming an image of a physical hand of a user and recognizing the a natural partition of the user's physical hand;

forming an augmented reality image of a user interface of the device by overlaying an image of a desired user interface on the image of the user's physical hand; 35

showing a user the augmented reality image of the user interface;

selecting a desired segment by touching a suitable partition on the user's physical hand, the segment being recognized as a natural partition of the user's physical hand; 40

recognizing the selection on the user's physical hand; and performing a function related to the segment in the device.

11. A method according to claim 10, wherein the desired segment is selected by the user. 45

12. A method according to claim 10, wherein the image of the desired user interface, overlaid on the image of the user's physical hand, is formed by means of digital signal processing from data provided by virtual reality gloves. 50

13. A method according to claim 10, wherein the user interface is an alphanumerical user interface.

14. A method according to claim 10, wherein the user interface is an interface based on QFD symbols.

15. A method according to claim 10, wherein the image of the user's physical hand is formed as a video image. 55

16. A method according to claim 10, wherein the image of a desired user interface consists of computer graphics.

17. A method according to claim 10, wherein the user interface comprises input segments and is situated on the image of the user's physical hand in such a way that the segments of the user interface are separated from each other by the natural partition of the user's physical hand. 60

18. A method according to claim 10, wherein the device is operated by means of MARISIL sign language. 65

19. A portable mobile communication device comprising a video camera unit, a display unit, a transceiver unit and a

digital signal processing unit that is arranged to control the operation of the portable mobile communication device, wherein:

the video camera is arranged to form a video image of a user's physical hand and to feed the image into the digital signal processing unit, which is arranged to recognize a natural partition of the user's physical hand in the video image;

the digital signal processing unit is arranged to feed an image of a user interface into the display and to form an augmented reality image of the user interface of the device by overlaying an image of a desired user interface on an image of the user's physical hand formed by rays of light originating from the user's physical hand, passing through the display, and entering the user's eyes;

a desired segment, recognized as a natural partition of the user's physical hand, is selected by the user by touching a suitable partition on the user's physical hand;

the digital signal processing unit is arranged to recognize the selection on the user's physical hand; and

the device is arranged to perform a function related to the segment.

20. A device according to claim 19, wherein the user selects a desired segment.

21. A device according to claim 19, wherein the user interface is an alphanumerical user interface.

22. A device according to claim 19, wherein the user interface is an interface based on QFD symbols.

23. A device according to claim 19, wherein the image of a desired user interface consists of computer graphics.

24. A device according to claim 19, wherein the user interface comprises input segments and is situated on the image of the user's physical hand in such a way that the segments of the user interface are separated from each other by the natural partition of the user's physical hand.

25. A device according to claim 19, wherein the device is arranged to operate by means of MARISIL sign language.

26. A portable mobile communication device comprising a video camera unit, a display unit, a transceiver unit and a digital signal processing unit that is arranged to control the operation of the portable mobile communication device, wherein:

the video camera is arranged to form a video image of a user's physical hand and to feed the image into the digital signal processing unit, which is arranged to recognize a natural partition of the user's physical hand in the video image;

the digital signal processing unit is arranged to form an augmented reality image of a user interface of the device by overlaying an image of a desired user interface on the image of the user's physical hand;

the digital signal processing unit is arranged to feed the augmented reality image of the user interface into the display, which is arranged to show the image to the user;

a desired segment, recognized as a natural partition of the user's physical hand, is selected by the user by touching a suitable partition on the user's physical hand;

the digital signal processing unit is arranged to recognize the selection on the user's physical hand; and

the device is arranged to perform a function related to the segment.

27. A device according to claim 26, wherein the user selects a desired segment.

13

28. A device according to claim 26, wherein the user interface is an alphanumeric user interface.

29. A device according to claim 26, wherein the user interface is an interface based on QFD symbols.

30. A device according to claim 26, wherein the image of a desired user interface consists of computer graphics.

31. A device according to claim 26, wherein the user interface comprises input segments and is situated on the image of the user's physical hand in such a way that the segments of the user interface are separated from each other by the natural partition of the user's physical hand.

32. A device according to claim 26, wherein the device is arranged to operate by means of MARISIL sign language.

33. A portable mobile communication device comprising a virtual reality gloves unit, a position tracking unit, a display unit, a transceiver unit and a digital signal processing unit that is arranged to control the operation of the portable mobile communication device, wherein:

the virtual reality gloves are arranged to feed information on a user's physical hand into the digital signal processing unit, which is arranged to form an image of the user's physical hand and to recognize a natural partition of the user's physical hand;

the digital signal processing unit is arranged to form an augmented reality image of a user interface of the device by overlaying an image of a desired user interface on the image of the user's physical hand;

14

the digital signal processing unit is arranged to feed the augmented reality image of the user interface into the display, which is arranged to show the image to the user;

a desired segment, recognized as a natural partition of the user's physical hand, is selected by the user by touching a suitable partition on the user's physical hand;

the digital signal processing unit is arranged to recognize the selection on the user's physical hand; and

the device is arranged to perform a function related to the segment.

34. A device according to claim 33, wherein the user selects a desired segment.

35. A device according to claim 33, wherein the user interface is an alphanumeric user interface.

36. A device according to claim 33, wherein the user interface is an interface based on QFD symbols.

37. A device according to claim 33, wherein the image of a desired user interface consists of computer graphics.

38. A device according to claim 33, wherein the user interface comprises input segments and is situated on the image of the user's physical hand in such a way that the segments of the user interface are separated from each other by the natural partition of the user's physical hand.

39. A device according to claim 33, wherein the device is arranged to operate by means of MARISIL sign language.

* * * * *