FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Personal Assistant for Improving the Social Life of Mobility-Impaired Citizens

Carlos Miguel Nascimento Tonim Galinho Pires

Master in Informatics and Computing Engineering

Supervisor: Eduarda Mendes Rodrigues (Ph.D. and Professor at the Faculty of Engineering, University of Porto)
Second Supervisor: Miguel Sales Dias (Ph.D. and Director of Microsoft Language Development Center and Professor at ISCTE)

28th June, 2010

Personal Assistant for Improving the Social Life of Mobility-Impaired Citizens

Carlos Miguel Nascimento Tonim Galinho Pires

Master in Informatics and Computing Engineering

Approved in oral examination by the committee:

Chair: Maria Cristina de Carvalho Alves Ribeiro (Assistant Professor)

External Examiner: António Joaquim S. Teixeira (Assistant Professor)

Supervisor: Maria Eduarda Silva Mendes Rodrigues (Assistant Professor)

31th July, 2010

Abstract

We began to interact with personal computers using keyboards and mouse. But nowadays we can speak to computers, we can touch them and even transmit commands using gestures. These alternative modalities can however have some problems when working alone, and so, by combining them, into multimodal systems, we are removing weaknesses and possibly improving usability, especially for disabled users.

The purpose of this thesis is to investigate whether multimodal interfaces offer advantages in enhancing mobility-impaired users' experience with communication and entertainment applications. This was achieved by studying alternative methods of human-computer interaction, encompassing multiple input and output interaction modalities, specially targeted for mobility-impaired users. A user study was made in order to study their limitations with current software and hardware interfaces, regarding services like, email, agenda, conference and audio-visual information access (media center). After that study, guidelines and alternatives were proposed, based on a clear set of user requirements, and based on that, a multimodal prototype was made to verify or not those assumptions.

After evaluation tests with the prototype, we have seen how multimodal interaction, can in fact improve mobility-impaired users' interaction with the above mentioned services, as they can choose the best modality for them or for each task situation. The prototype, offers communication services which are available not only on the desktop, but also on mobility. We believe that the developed prototype and the carried usability evaluation, have demonstrated that, it can help fighting isolation and therefore improve mobility-impaired citizens' social life.

Resumo

Quando os primeiros computadores pessoais surgiram, os periféricos de entrada disponíveis eram apenas os tradicionais teclado e rato. No entanto hoje em dia já é possível falar com os computadores, tocar-lhes e até transmitir-lhes comandos usando simplesmente os gestos. Estas formas de interacção alternativas, podem contudo ter alguns problemas quando estão disponíveis sozinhas. E assim, ao agregar várias modalidades num sistema, ou seja, fazendo-o multimodal, estamos a retirar fraquezas e possivelmente a melhorar a usabilidade, especialmente para utilizadores com mobilidade reduzida. Isto porque eles podem escolher a melhor modalidade tendo em conta as suas limitações, bem como a situação. O objectivo desta dissertação é investigar se as interfaces multimodais, podem melhorar a experiência de utilização de aplicações de comunicação e entretenimento, por parte de utilizadores com mobilidade reduzida .

Primeiro, começou-se por fazer um estudo de usabilidade, o qual abrangeu interfaces de *hardware* e *software*, no que diz respeito a serviços de email, agenda, conferência e informação audio-visual (media center), e do qual se retiraram directivas e linhas de orientação. Seguidamente foi desenvolvido um protótipo multimodal, tendo em conta os dados do estudo de usabilidade e necessidades do utilizador, a fim de se validarem ou não os resultados do estudo inicial.

Após avaliação do protótipo com um painel de utilizadores com mobilidade reduzida, concluiu-se que de facto as interfaces multimodais podem facilitar a interacção deste grupo de utilizadores com os serviços acima mencionados, já que é possível escolher a modalidade mais adequada à pessoa e à situação. O protótipo oferece serviços de comunicação disponíveis não só na plataforma *desktop* como na plataforma *mobile*. Assim, o protótipo desenvolvido pode ajudar a combater o isolamento e consequentemente a melhorar a vida social dos cidadãos com mobilidade reduzida.

Acknowledgements

I would like to thank to all those that helped me and contributed to the elaboration of this thesis.

I would like to thank:

My supervisors Professor Eduarda Mendes Rodrigues of FEUP and Professor Miguel Sales Dias of Microsoft/MLDC and ISCTE for their valuable orientation and supervision and for selecting me to this project;

António Calado, João Freitas and Pedro Silva of the MLDC, for all their technical support and friendship;
Ana Judice, Daan Baldewijn and Difda Monterde for all their help and support during testing stage;
Marta Castanheira, Nuno Costa and Tiago Barbosa for all the valuable feedback regarding design improvements on the final prototype;
All the mobility impaired participants from Associação Salvador, which took a very important role in this study, and without whom it would be impossible to do this study;
Salvador Mendes de Almeida, founder and CEO of Associação Salvador, for all his support and for mediating the recruitment of the study participants;
Fernando Pinto, my colleague, who helped me during requirements analysis and development stages, and whose his thesis was did in parallel with mine, under this Microsoft's project;
All Microsoft colleagues that helped me during my stay at Microsoft Portugal;
My family that always supported and helped me.

Thank you all!

Carlos Miguel Nascimento Tonim Galinho Pires

This work was co-funded by Microsoft under QREN 9700 Living Usability lab (http://www.livinglab.pt/).

Contents

List of Figures

List of tables

Abbreviations

List of abbreviations (in alphabetical order)

AAL	Ambient Assisted Living
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
ASR	Automatic Speech Recognition
EMMA	Extensible MultiModal Annotation markup language
GPS	Global Positioning System
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HD	High Definition
ICT	Information and Communication Technologies
OCS	Office Communications Server
PC	Personal Computer
PDA	Personal Digital Assistant
PLA	Personal Life Assistant
PTT	Push to Talk
SAPI	Speech Application Programming Interface
SDK	Standard Development Kit
SIP	Session Initiation Protocol
SOAP	Simple Object Access Protocol
TTS	Text-to-Speech
UCCA	Unified Communications Client API
UCMA	Unified Communications Managed API
W3C	World Wide Web Consortium
WPF	Windows Presentation Foundation

Chapter 1

Introduction

1.1 Introduction

Much work has been done in the area of Human-computer Interaction (HCI) over the past three decades to improve user experience. Means of communication have also evolved from rudimentary text-based services to email, audio and video conferencing or even socially-enabled means of communication. Now we can even interact with computers just using gestures to them, as in augmented reality applications. These evolutions are, however, taking more time to reach impaired users. These delays, coupled with real-world physical obstacles can severely limit mobility impaired users' interaction with other people and their integration in society, leading to forced social isolation.

And so there is room to explore and perceive how current interfaces and applications limit mobility-impaired people interaction. Can we enhance mobility-impaired users' experience with communication and entertainment applications, using multimodal interfaces? This is the main question beyond this thesis.

1.2 Work Context

This dissertation work was proposed as a collaboration between the Faculty of Engineering of University of Porto and the Microsoft Language Development Center (MLDC), under QREN Living Usability Lab[1], located in the Portuguese subsidiary of Microsoft, in Porto Salvo, Oeiras. Approximately nine weeks were spent at MLDC's installations in Porto Salvo, dedicated to the specification and development of a prototype, as well as the execution of the user studies and the writing of a scientific paper.

Fernando Pinto [68] (from Faculty of Engineering of University of Porto) developed his dissertation ("Multimodal Access to Social Networks") in parallel with this. Regarding the

[1] http://www.livinglab.pt/

current dissertation, the author contributed with email, agenda and conference service functionalities. Both authors contributed equally regarding common features.

1.3 Problem and Motivation

Mobility impaired individuals are people whose disabilities affect their ability to move, manipulate objects or, in any other way, interact with the physical world. These limitations usually result from genetic abnormalities, accidents or excessive muscle strain. These impairments can also severely limit individuals' interaction with the digital world, due to their inability to control regular computer input peripherals such as keyboards, mice or tactile devices. The barriers for online social interaction, coupled with mobility difficulties in real-world environments can severely limit these individuals' independence, as well as leading to social isolation.

Over the past 30 years, user interfaces have evolved from typical keyboard and mouse to more natural means of interaction, allowing touch and speech interaction, as well as some support for gesture interaction.

Although speech interfaces are relatively seamless to use, some robustness issues make them inappropriate to use in noisy environments, such as input misinterpretation. Also, usage when users are in public environments must be taken into consideration, mainly when accessing private data (i.e.: authentication, personal data insertion, etc). Usage for long periods of time can also cause fatigue, making them not very appropriate for constant usage.

Touch interfaces allow users to interact with devices by means of a bi-dimensional touch screen. Recent advances in these types of interfaces have made it possible to interact in more natural ways with these interfaces, through the usage of two-dimensional gestures on the screen, as well as multi-touch gestures, emulating in some cases common human gestures. Since these interfaces require some hand coordination, some caution must be taken when attempting to use these interfaces with mobility impaired users.

With three-dimensional gesture-based interfaces, a user can interact with a computer just by making natural gestures (hand, body or motion gestures) to a camera, or using special sensory attached to the body, such as holding a smartphone equipped with accelerometers and gyroscopes. However, these systems also have some issues that may limit their adoption: image noise, in image-based processing systems, can in some scenarios create incorrect interpretations. Due to the amount of required physical activity, user stress and physical limitations are also other issues that must be taken into account when using these types of 3D interfaces.

One attempt to make interfaces suitable for impaired people relies on gaze detection. This type of interface works by identifying where the user is looking at, either through image processing, captured by a webcam, where the image is scanned to track eye position, or through infrared cameras that track special markers placed, for example, on glasses or in a hat, or even through electrooculography, where by measuring the measuring the resting potential of the

retina, it is possible to derive eye movements. The latter options, although more invasive, are still considered better options to highly noisy environments, such as in unfavorable light conditions, since many image processing and computer vision problems can be avoided. Issues, such as equipment placement by users with mobility or dexterity limitations can, however, make it more difficult to use this kind of interface.

One way to overcome some of the problems seen above, as well as reducing the impact of some of the limitations disabled people face, is by using multimodal interaction systems. These allow users to interact through one or more means of input/output, be it concurrently or not, according to users' interaction environment, personal preferences or even disabilities. Thus, with these interfaces, should users be unable to speak, they could instead use a gesture interface or, in situations where they cannot properly coordinate their arms, a speech interface could be used. The advantage of these types of interfaces is not only the ability to allow multiple means of interaction, but the ability to use them in a seamless way, without explicitly requiring users to specify which types of interfaces to use.

Giving this, there is room to explore how multimodal systems can improve mobility-impaired people experience with communication and entertainment applications, and possible improving their everyday lives as well. As by raising communication barriers imposed by current interfaces, mobility-impaired individuals can be closer to society.

1.4 Main thesis goals

The main goal of this thesis is to study how alternative HCI modalities such as, speech or touch, can improve mobility-impaired users interaction, especially in the email, agenda, conference and media center applications.

The first objective is to perceive in which way current interfaces, considering both software and hardware, limit and affect mobility-impaired people interaction.

Another objective of this thesis is to apprehend if and in which way, alternative modalities (such as speech or touch) can improve mobility-impaired user's interaction. This will be investigated through user studies with mobility-impaired users (paraplegics and quadriplegics).

A third objective is to develop a prototype, taking into account requirements collected in the user studies. The prototype will consist on a Personal Life Assistant (PLA) offering email, agenda, conference and media center service capabilities, and it will be deployed on desktop and mobility platforms. The prototype will enable us to test our hypothesis and draw conclusions about the use of multimodal interfaces for communication and entertainment services.

1.5 Thesis Organization

This thesis is organized as follows:

The current chapter introduced the motivation and objectives of this thesis. The remaining chapters are organized as follows.

Chapter 2 presents some background of alternative ways of interaction beyond keyboard and mouse, such as: speech-based interfaces, gaze-detection interfaces and gesture-based interfaces. Definitions and features of multimodal systems, which combine more than one modality, are also presented in this chapter.

Chapter 3 presents two user studies aimed out to gather user requirements. The first study is a preliminary requirements analysis interview and the second one is a more structured requirements analysis session. Recommendations for the design of user interfaces geared towards mobility-impaired users, derived from the user requirements analysis, are covered in this chapter too.

Chapter 4 presents the prototype specification of a Personal Life Assistant, specially targeted for motor-impaired users, that will take care of accessing such ICT services, as email, agenda, conference and audio-visual information management (media center), encompassing its functionalities, interfaces and architecture. Additionally, the technologies used and some details about the prototype development are described.

Chapter 5 describes a prototype evaluation study, presenting the results and conclusions regarding the use of multimodal interfaces in an integrated Personal Life Assistant.

On the final chapter (Chapter 6), thesis conclusions and final remarks are presented. Possible future work and lines of research are also discussed.

Chapter 2

Literature Review

2.1 Introduction

The ways we interact with computers evolved a lot. In order to communicate with the first computers we needed to use punched card readers. Writing and reading punched cards took too long and these operations were error prone and completely unnatural. Then keyboards and screens appeared, at this point users could see what was happening, just by look at the screens. They could finally insert *human* text and avoiding handling binary code – the computer language. Later a new piece of hardware appeared: the mouse. Using a mouse allowed users to interact with computers in a different way, not only text insertion was available but also a graphical ambient. Regarding to personal computers, the set mouse and keyboard can be considered the primary input methods and the screen display the primary output method.

Until recently, a keyboard/keypad and a screen were sufficient to interact with machines. But new kinds of hardware appeared, and the human-machine interactions needed to be rethought. Home entertainment systems (set top boxes), PDAs (cell phones without any keypad) and vehicle on-board computers are examples of devices that are changing the way we interact with computers. They are a critical motivator for the development of multimodal interfaces [44], which we will talk about later.

Not only the advent of new devices make the current interfaces useless. If an interface cannot be used by everyone, automatically that interface is useless regarding users that cannot use it. Designing for everyone is something that is becoming more important. But why not everyone can use all interfaces? Well, we can argue that it is because of bad interface design and consequently it is difficult to use an application which makes that users simply don't bother in trying to use it. But probably the principal reason is that people that are excluded have some

kind of physical and/or mentally impairment, that is, they are disabled. In this demand of trying to integrate disabled people in the machine era a lot of work has been made, as we will see in this chapter. It is interesting that some inventions like the telephone, cassette tape and ballpoint were originally products whose objective was aid disabled people. We talked about designing for everyone and helping disabled people, and so there are 2 concepts that have to be distinguished: access and assistance. Access deals with making a current interface that a disabled user cannot use, accessible to him. Assistance is the activity of creating an interface in order to improve the life of disabled people [45]. One principle in order to make interfaces accessible for disabled, and consequently accessible to everybody, could be let the users interact with interfaces using any input that they choose. That is, if a disabled user cannot use his hands to write in a keyboard, why not using his voice? This leads us to multimodal applications. But firstly: what is multimodal? As the name implies, it is something that uses more than one modality. And what is a modality? A modality is defined as being a mode of communication corresponding to human senses or type of computer input devices. And so, we have cameras, haptic sensors, microphones, olfactory and taste; there are also inputs that do not map directly any human sense: keyboard, mouse, motion input and others (as defined in [1]).

As we have seen, Human Computer Interaction is becoming more ubiquitous, meaning that people can find computers in almost everything. A tough challenge is making the interaction with devices more natural and more intuitive, and consequently making applications available for everyone. This can be achieved by making software multimodal, that is, if we can make applications that can use various communicative modalities, we can therefore develop software that could be used by more people including older, impaired and handicapped users. Considering that disabled people, mainly handicapped, stay more at home, communicating with the exterior world is one of the major requests. They need to be in touch with other people and sometimes work at home is the only way. So all this can be made using technology even for users that normally do not use computers.

In this chapter we will talk about new ways of interaction between humans and machines that are more natural, as the user can use something that he uses everyday with other humans, like his voice or his body (gestures). Of course these new ways of interaction bring some problems, because they cannot be used as direct inputs like mapping a scancode from a keyboard to an ASCII or Unicode character. These inputs need to be recognized and this operation is many times ambiguous, which leads to errors. Finally we will see what multimodal applications are, how they can combine their inputs (modalities) and how these systems can be advantageous.

2.2 Speech-based interfaces

Regarding alternative modalities, speech is probably the most common modality. Text-to-speech and automatic speech recognition are technologies that are available in many applications and operating systems. Speech is used in many multimodal applications, probably because it is cheaper and more convenient for some situations, as argued in [2]. Although these technologies have been available for a while, there is still a lot work to do. The School of Science and Engineering of Waseda University, developed a robot who could communicate with multi-users, but in order to identify who was talking, the robot used face detection and recognition and not only with speech analysis [3]. As we can see, this type of interface still has some problems.

But if we want to dialog, for example, with older or impaired users, we have to be more cautious, as they could have some limitations. Christian Müller and Rainer Wasinger developed a mobile pedestrian navigation system (GPS) that could recognize a type of user (elderly or middle aged adults) by his voice, adapting the interface accordingly [4]. Meaning that if the user is elderly, the interface becomes simpler, if the speaker is newer (e.g. son or daughter), now the interface can be more complete, with more functionalities and therefore more complex. This has the advantage of reducing cognitive load that is something that has to be considered with disabled and elderly users, as these kinds of users could have cognitive disabilities (e.g. age degenerative process, short-term memory problems) or physical impairments (e.g. reduced visual or auditory capabilities) that could limit their HCI interaction. This system can recognize the group in which the user is, but not the user, and which raises issues with user authentication.

One of the biggest problems in speech-based systems is maintaining some privacy, mainly if the objective is making authentication, for example using a password. Probably dictating a password to a computer is not a good idea. But there is an interesting work on making robust authentication using ASR (Automatic Speech Recognition) [5]. Basically the computer asks the user private information that must be combined with random data provided by the system. For example, the computer asks “what is the sum of your password’s second number plus 10”. This approach has some problems: the user must use phones in order to hear what is asked, the password cannot be stored in a secure manner (encrypted) and this could lead to cognitive problems. For example, asking the user for the 20th character of his password could be a bad idea. Despite this, the objective of making authentication using ASR is accomplished and it is proved as a secure system.

Another problem of ASR is the possibility of misrecognizing a word. This occurs mainly with words or letters that have similar pronunciation. One way to solve this problem is to choose commands that have very distinct pronunciation from other commands. For example, the letters *a* and *b* could be identified as the same by the ASR, but if we use the Greek alphabet (alpha and beta respectively), the system is capable of identifying correctly what is being said [6]. The misrecognition problem can also occur in noisy environments, as other sounds can be

mixed with commands and a wrong one is selected, or simply the sound system cannot capture any user voice. There are some algorithms and hardware that try to solve this problem, as for example microphone array which is used in [7].

Another limitation that could occur is that after a long time using a speech-only interface, the user becomes tired. If we imagine that a person had to speak all day without having a break, it's easy to find out that at the end of the day, that person will be very tired. Equally, there are actions that require more speech commands than using another kind of modality. For example, in order to map mouse movements, if the user wants that the cursor goes to a point on the screen, he have to firstly select the destiny point and then give the command. To select a point on the screen, user can use grids in which he selects an area, and these grids become smaller and more accurate, later he can select the coordinates (this is an example using Windows Vista's Speech Recognizer). Therefore, after a long time of use, speech-only based interfaces can be stressing. VoiceDraw [49] is a very interesting approach in order to control a cursor, it is a vocal joystick. And why is it different? VoiceDraw does not recognize words or phrases but sounds. The big advantage is that by using sounds like "ch" or "uh", we are reducing the stress problem as we have to speak less. A vocal joystick works by mapping a different sound to each direction. It is possible too to map a level of loudness or pitch in order to insert multidimensional properties, like velocity. So a user can control a cursor just by producing sounds with different properties. But this approach has one major disadvantage: in order to learn how to use this system, the user must spend some time, although the sounds can be adapted to each user.

As we have seen, there are some issues that we have to consider in using speech-only based interfaces. And so a solution consists in adding more modalities and redundancy in order to reduce these problems, like we will see in the next section.

In summary we have seen that speech is a very common interface but has some problems like:

- limitations of speech-only interfaces;
- the privacy problem and a solution on how to make authentication in speech-based systems;
- the bad word recognition problem, that can occur due to ambient noise or similarities between words;
- stress and overhead that some speech-only interfaces could lead to, and how to minimize it.

2.3 From speech to multimodal

As we have seen, speech is very important in multimodal systems, however, speech-only applications are not enough to handle all inputs, as they could have some faults. We have

seen that ASR applications can lead to some problems such as authentication and bad word recognition. So, on the one hand it is not advised to speak or hear private information, and on the other hand it is probably inconvenient too to dictate a lot of text to a computer. By adding another modality, we are inserting some redundancy and even giving the user some freedom of choice regarding which modality to use [8].

Turunen and Hakulinen ([9] and [10]) presented a multimodal Media Center application controlled by speech and gestures using a mobile phone. In this case a mobile phone was used to detect command gestures, using its accelerometer, in order to reduce the number of commands that the user would have to dictate. Another interesting application is [11], describing tabletop games that are controlled using speech complemented with gestures, using a touch table. Users can then use a voice command, for example “move”, and complemented it pointing to a specific point on the map.

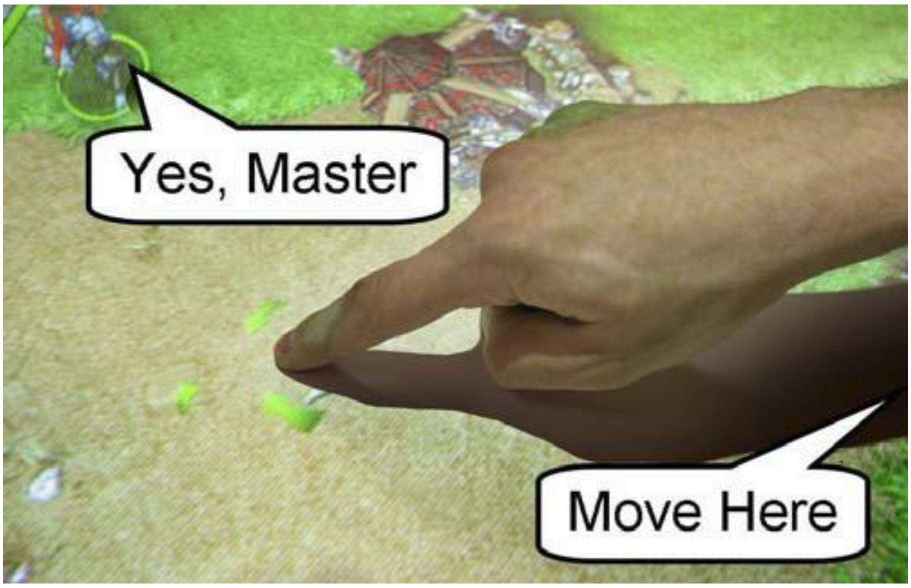


Figure 2.3.1 – Example of use of [11].

Another example of good cooperation between modalities is [12], in which a user could say “call this person” (speech) and at same time pointing to a picture of the person on a touch-screen display or handwriting the recipient’s name on the screen. Similarly, MATCH (Multimodal Access To City Help) [50] uses a touch and speech interface in order to find locations and information on a map. A user can say “show cheap Italian restaurants in Chelsea” or draw a circle on the map and then say “show cheap Italian restaurants in this neighborhood”. User can even draw a circle on the map and inside that circle write some search items like “cheap Italian”. This is another example of good cooperation between modalities and in this case with completely redundancy between them: speech-only, touch-only and speech with touch.

Applications that use speech have to maintain a database of available words/phrases that can be recognized. Normally this kind of data is static and so the user is restricted regarding to customize the vocabulary. [51] uses a multimodal approach in order to enable the edition of the rule grammar and the semantic database. Although this is interesting, there are some issues that we must consider, probably the most important is that it is still hard and complex for a regular user to customize a rule grammar and a semantic database. Clearly, this area needs more investigation.

Thus, speech is very interesting in an application, but does not increase interaction efficiency just by adding TTS (text-to-speech) and ASR on existing software and could even bring to more erroneous interaction, as pointed out in [13].

In summary we have seen that, following what was said in the previous section, speech by itself is not sufficient to solve interaction problems and could even lead to new ones. A good approach is to integrate speech with other modalities, leading to multimodal systems that avoid speech-only interfaces problems.

2.4 Gesture-based interfaces

So far we have seen multimodal applications using speech and touch recognition, but other modalities can be used. Gesture recognition is another kind of interface that tries to make communication more natural. In this group we have controller-based gesture and image/video based gesture recognition.

In image/video-based gesture recognition, movement tracking is more complicated, because image processing techniques need to be applied in order to separate worthwhile components from those that are not. This process is difficult because the retrieved image can have noise, the environmental lights can affect image processing and background objects or distinct features can also lead to bad recognition. Computer vision and image analysis are areas that are being investigated and there is still a lot work to do [67].

Nielsen [14] presented a gesture interface, in which a hand-based vocabulary was defined. This had some disadvantages such as: (1) fatigue, (2) only a limited number of inputs were possible and (3) some people had difficulty in performing the gestures. Nevertheless it had the advantage of easy recognition, as it is easier recognizing static images than dynamic ones. Image/video-based gesture recognition is being increasingly used nowadays in many entertainment systems such as: Microsoft's Project Natal [15], Sony EyeToy [16] and YDreams' Audience Entertainment [17]. Many of these systems incorporate augmented reality capabilities, giving a more immersive experience.

Controller-based gesture recognition relies on specific hardware. For example Nintendo's Wii [18] makes use of specific controllers as *Wii MotionPlus* or *Wii Balance Board*, in order to make movement tracking. Another example already seen is the use of accelerometers

on cell phones. Even so the common hardware for this kind of interfaces is touch screens, available in tablet PCs and smartphones. A tendency in this kind of applications is multi-touch that is the ability of recognizing multiple gestures. For example, this kind of technology is available on iPhone [19] and Microsoft Surface [20].

In summary we have seen that gesture-based interfaces are an interesting approach essentially in entertainment systems. Making gestures has some problems like stressing, limited number of inputs available and inability by some people to perform some gestures.

2.5 Multimodal applications as distributed systems

As we have seen, the advantages of multimodal applications are enormous, as one modality could suppress other modality weaknesses. The main objective is making interfaces more natural, by making them more similar to natural interactions. We believe that multimodal applications are the future and are here to stay. But multimodality by itself is not enough in a ubiquitous computing world. Is required something more, and distributed computing is the way, as systems become available more than in one device. If we can have a network of sensors connected, each one recognizing a different modality, then we can have a complete ubiquitous and multimodal system. MONA [21] stands for Mobile Multimodal Next-generation Applications, and describes a middleware for mobile-based multimodal applications. The *Mona@Play* application was a quiz game that could be played using cell phones. One of the interesting features was the fact that it had TTS and ASR engines, and a multimodal chat, that is, it was possible to speak and those spoken messages were translated to written text and vice-versa. The users were very pleased with this application, because they could choice between using written or spoken messages. And this is a good principle of multimodal applications.

The concept of distributed systems bring new problems, as developing for mobile phones or set top boxes is different than developing for a "standard" device as a desktop or laptop. *TravelMan* [46] is a multimodal mobile application for serving public transport information that uses a centralized ASR, that is, the speech recognition function is placed on a remote server. Then, the sound is transmitted to the server, it is processed and a response containing the recognized word/phrase is received. Considering that the vocabulary is considerably large, the speech recognition process would be impracticable on the device because of memory and processing speed issues, so a distributed architecture is better. Another advantage of this approach is that mobile applications could use the server beyond speech recognition, for example operations that could take too long to do on a mobile device, can be made in the server, and so the client receives only the result. Still considering the speech processing on mobile devices, there are more problems that developers must take into account. Regarding to ASR and TTS engines, there are many limitations because there are no standardized SAPIs for each operating system, device neither programming language. One

option could be purchase third party software, but this will increase the development costs. Even if an ASR/TTS engine is available probably only one language could be handled [47]. So, considering this, if we are developing a mobile application that uses ASR and/or TTS engine(s), it is a good option to prefer a server/client architecture, in which all the speech processing is made on the server. Beyond hardware limitations, some interface constraints need to be also considered. Considering the small mobile displays, menus, submenus, icons and available information should be reduced to minimum. Also the design of the interface must take into account the hardware (device with keypad or touch screen) and the operating system [48].

Normally being distributed imply that a system is available anytime everywhere. And so, mobility is one of the requirements of 21st century applications. But, as we have seen, there are hardware-related issues that have to be considered, which normally affect the interface itself (e.g. small screen displays). Despite these problems, new ways of interaction for mobile ambient are appearing. Gestures and touch are already available on smartphones. SixthSense [55] is a revolutionary gestural interface that expands how a user can interact in a mobile environment. Images from the real world and gestures using user's fingers are the system inputs. For output, an image is projected on every surface (e.g. wall or hands), making this an augmented reality interaction that will change mobile interaction. Of course being available everywhere, does not imply that a system exist in terms of mobile devices. It could simply imply that a system exists through various devices. But we have to consider that devices besides PCs or laptops have hardware limitations and consequently interface issues.

Besides being distributed, multimodal applications need to be intelligent too, in order to make distribute computing more effective. PMO [22] is a multimodal presentation planner that uses distributed agent architecture, and is part of EMBASSI project. One advantage of using agents, is that they can cooperate to accomplish a task, as they did in EMBASSI. INHOME [23] is an AAL (Ambient Assisted Living) system whose architecture is composed by: (1) white goods (refrigerators, washing machines, etc), (2) audiovisual and entertainment equipment, (3) health care devices and (4) home automated devices. Considering that modern entertainment systems are complex, by using agents, make possible to reduce some complexity, build smarter systems and even implement new features without affecting the existent ones.

In summary, using multimodal systems only in one device can be restrictive since the full potential of a multimodal interface might not being used, and so making a multimodal application with a distributed architecture could be very advantageous. In using a distributed architecture there are issues that have to be considered such as memory and processing speed of some devices (e.g. mobile phones). Also each component of the system needs to be intelligent enough to reduce complexity to the central component. So that's a tradeoff problem that developers need to handle.

2.6 The main problem of multimodal systems: the fusion engine

2.6.1 Introduction

A multimodal application must be capable of dealing with different kinds of inputs. As we have seen ubiquity is more and more common, and so multimodal applications developers not only have to deal with different modalities but also have to deal with different hardware, operating systems, APIs, etc. In order to deal with this, an abstraction architecture of the multimodal system has to be created.

HephaisTK [53] is a toolkit for rapid prototyping of multimodal interfaces. As we can see in Figure 1.2, HephaisTK uses a distributed architecture in which each modality is treated by a recognizer. Each recognizer is associated with an agent, which encapsulates and sends input data to a postman. The postman centralizes all incoming data and sends it to all interested agents. The integration committee is composed by a fusion engine that fuses all input data, a fission engine that encapsulates the fused data and sends it to the application, and a dialog manager that helps the fusion engine.

In terms of multimodal architectures, there already some standardizations. W3C Multimodal Interaction Framework [65] defines a comprehensive architecture that includes input components (recognizers), output components, an interaction manager and a session component among others. W3C also defines a language capable of encapsulating modality data, named EMMA [66].

In this section we will focus our attention in the fusion engine that is the heart of every multimodal application.

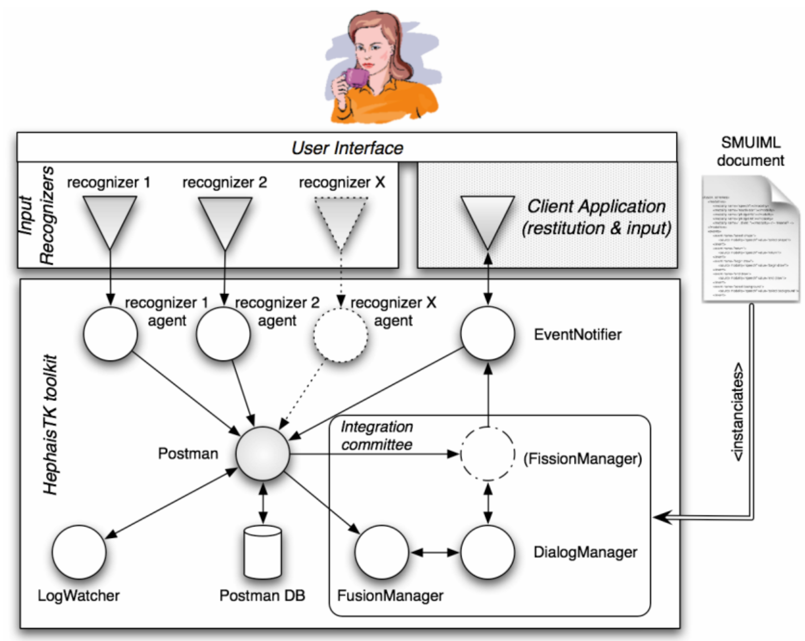


Figure 2.6.1 – HephaisTK [53] toolkit architecture.

2.6.2 What is a fusion engine?

We have seen different types of multimodal systems combine many different types of modalities. But the combination of input and output modalities can lead to errors and problems. Thus we have the fusion problem.

Basically multimodal applications have to deal with more than one input, and eventually more than one output[2]. And there must be some mechanism capable of joining different signals into one, and this operation is called *fusion* which is made by a *fusion engine*. This operation called *fusion* can occur at different levels: data level, feature level and decision level. The first one is made directly on the input streams, the second one is made by analyzing data characteristics and patterns, and the final one occurs at the application level. But what are the main concerns of fusion engines? The fusion engine must be capable of dealing with the following problems:

- **Probabilistic inputs:** if the input data is non-deterministic, then it has to be converted into a deterministic one, in order to make it readable and interpretable by the application;
- **Multiple and temporal combinations:** the problem is input/output generation delays, from either fault of the user or software/hardware limitations;
- **Adaptation to context, task and user:** for example, dealing with elderly users or with children is completely different, from dealing with experienced users. A multimodal command given by a child or an older adult can and must be differently interpreted. Also operating on a vehicle or at home is differently. Therefore, a fusion engine must be capable of adapting according to context, task and user.
- **Error handling:** dealing with all this could lead to errors, and fusion engines must be capable of error handling and error avoidance.

The input signal integration can be made at feature level, also called early fusion, and at semantic level, also known as late fusion. The first one is more appropriate to synchronous modalities (speech and lips movements) as it uses structures like Hidden Markov Models and neural networks. Consequently they are more complex, they have to handle with a lot of training data and they are very computationally heavy. On the other hand, late fusion architectures are more scalable and could handle asynchronous inputs. Another advantage is the easy combination with various modalities, meaning that we can add or remove modalities without re-training the entire system, so semantic (late) fusion architectures rely on independent low-level handlers for each modality.

[2] Some authors consider that a fusion engine includes the fission engine.

2.6.3 Fusion engine terminology

Fusion engines must combine different inputs and interpret them in order to perceive what user is trying to say. But multimodal systems must be capable of dealing with different modalities at same time and being capable of combining them. And so, we can simply define a multimodal system as an application that accepts various redundant modalities.

Nigay & Coutaz [56] created a classification that defines types of multimodal interfaces. According to Nigay & Coutaz we can have two different ways of use modalities: sequential (one modality at a time) and parallel (various simultaneous modalities). In terms of the fusion, we can have: combined fusion (different modalities are combined into a single action) or independent fusion (each modality is mapped as a single action). For each fusion combination and use of modalities, Nigay & Coutaz defined the following multimodal interfaces types:

Fusion combination \ Use of modalities	**Sequential**	**Parallel**
Combined	Alternative	Synergistic
Independent	Exclusive	Concurrent

Table 1 – Types of multimodal interfaces according to Nigay & Coutaz [56]

2.6.4 Fusion engine architectures

Multimodal interfaces born in 1980 with Bolt's *Put-That-There* [40] prototype, in which speech and gesture were used in order to do object manipulation. Unfortunately no details about the fusion engine used were published. Later, some work on how to engineering a fusion engine was made with Xtra and CUBRICON, and the concept of frame-based was introduced. This means that the information received must be time-stamped in order to be interpreted. Later, unification mechanisms were introduced. The unification process consists on the combination of normally two data streams to a single stream, using heuristics. Nowadays the way is to use hybrid solutions combining all previous methods. For example, a system could receive time stamped-data using frames, and then 2 or more frames could be joined (unified), to a single frame that specifies a high-level command.

One interesting approach on hybrid architectures was Quickset which used Associative Maps and Members-Teams-Committee techniques. An Associative Map is a component that defines semantic relations between all modalities sets, for those actions that cannot be unified. MTC (Members-Teams-Committee) is a technique that measures the contribution (prediction of an intended action) of each modality, and then selects the most probably action. For example, Quickset's gesture recognizer had 190 possible gestures in its vocabulary, and so there were 190

members. Each member scored each received gesture, and then the team-leader applied various weighting parameters, which were transmitted to the committee that determined the final results.

Time-based information has some problems. For some interactions it is difficult to deal with delays. For example, how long takes a specific gesture to complete? Well, we can arbitrate a value at least for the major people, but that's not a good approach. A gesture can be defined as a sequence of actions that have to fit in a time-interval with arbitrated time values. Equally, a command involving speech and touch, must deal with delays between speech commands and touch commands. These delays could be different from person to person and even from situation to situation, and so time-based approaches could lead to unpredictable errors. QuickFusion [41] is a grammar-based approach and it's a good response to solve time-based problems. So, we have to define a grammar representing all possible actions from multiple modalities. And then, commands are recognized just by analyzing the grammar. For example, a command is defined as a speech command followed by a touch command. If we use a time-based approach, a threshold has to be defined and we have the problems discussed above. In this approach, after the speech command has been received, we only know that we have to wait for a specific touch command. Another advantage is that having a grammar, we can avoid ambiguities since all commands are explicit defined. In a time-based approach it is possible to have many time-stamped commands (from each modality) that could be translated into multiple application commands, using heuristics. In conclusion, whenever possible it is better to use grammar-based fusion engines than time-based ones.

2.6.5 Fusion engine classification

As we have seen, there are many fusion techniques they could be time-based or grammar-based. The fusion process can occur at low or high level, and so there are a lot of possibilities on how to make a fusion engine. Considering the types of fusion engines available, [24] proposed a fusion engine classification based on 6 characteristics:

1. **Notation:** which is the language used in the fusion engine. For example it could be based on XML or stream stamped.
2. **Fusion type:** it could be frame-based, unification, procedural or hybrid.
3. **Level:** is fusion made at raw level (low data level) or at dialog level (fusion events can immediately trigger application commands)?
4. **Input devices:** it defines the set of devices used in a multimodal system (pen, speech, touch, etc).
5. **Ambiguity resolution:** in case of conflicted commands have been identified, which is selected? This resolution can be made using: priorities (the chosen command is from the predominant modality) and iterative tests. Other policies can be used.
6. **Time representation:** it could be quantitative (each data item is tagged with a timestamp), qualitative (each data item is ordered according creation time) or both.

The fusion problem discussed here is based in [24] and [39].

2.6.6 Fusion engine evaluation

If we want to test and evaluate a fusion engine, what should we do? Well, we can simply make unitary tests, that is, we can compare the real results of a fusion engine with the expected ones. But there are more points that have to be evaluated.

Dumas [54] proposed a set of qualitative and quantitative metrics in order to evaluate a fusion engine. For each multimodal event they proposed to measure in a quantitative way the following:

- **Response time:** time that fusion engine takes to return a result after receiving inputs;
- **Confidence: l**evel of confidence of the machine response (this is useful for some modalities like speech, that have a probability associated with each recognized word);
- **Efficiency:** comparison between current and expected results.

In terms of qualitative evaluation they proposed the following:

- **Adaptability:** capacity of the fusion engine in adapting to a context or user;
- **Extensibility:** can a fusion engine deal with new or different inputs? In this case, a distributed architecture that uses a recognizer for each modality is probably more extensible than other kinds of systems.

In summary we have seen:

- what is a fusion engine and its terminology;
- the main problems that fusion engines must be capable of dealing with;
- the evolution of fusion engines' architectures and which are the features of each one;
- a fusion engine classification;
- how to evaluate a fusion engine.

2.7 Multimodal systems for disabled

2.7.1 Introduction

In relation to general multimodal systems, we have concluded our analysis. Now it's time to focus on multimodal systems for elderly, impaired and handicapped users. But why are multimodal applications so important and so common for this kind of user groups? The reason

is that using some modalities at same time instead of one, makes it is possible to cover more human senses, and then reach more users. And so users like elderly, impaired or handicapped, that have reduced or even nonexistent some type of communication channel, can use another available channel, in a multimodal application. And that's why too, that applications specifically designed for handicapped users, may also be applied to elderly, as there are some common restrictions in both.

Many multimodal applications for elderly, impaired or handicapped users ([9], [10], [7], [12], [25], [26], [27], [23], [28]), focus their functionalities on the following aspects:

- **Cognitive support:** This can be supporting interaction, as giving at all time information about what is possible to do. Or, it can also be life support, as for example, remembering when to take a medicament or even when relative's birthday is. The first depends on a good interface, while the second one consists in an agenda.
- **Socialization support:** Being in touch with relatives and friends is something that helps preventing isolation. And then it is an essential requirement in multimodal personal assistants. This communication can be made using emails, instant messaging, audio and video-conferencing and even phone calls using VoIP.
- **Entertainment support:** This point consists on using media centers that can present multimedia components such as: video, photos, TV, radio, etc.
- **Care support:** Considering that users stay at home by themselves, in case of an emergency, it is required that applications inform relatives and even authorities of what is going on, and eventually take action. So, some applications have the option of monitoring in real-time the health of users, using non-intrusive sensors. Others have the option of maintaining contact with users' caregivers. But this kind of support may rely only on informing the user when to take a medicine and which quantities to take.

So many applications specifically designed for elderly, impaired and handicapped users consist on personal assistants that have the objective of improving their everyday life.

2.7.2 Gaze-based Interfaces

Aspects seen previously could also be applied to all audiences, including non-impaired users. And so, what differentiate user groups are ways of interaction and special concerns of design. Designers must adapt and create modalities depending on the type of impairments of the target users. For example, for quadriplegic users that only can move head and eyes, the most common interfaces are speech recognition and gaze detection.

Regarding to gaze-based interfaces, that is, the process of identifying for where we are looking at, we have some available products in the market. MyTobii [29], SmartNav [42], Magic Key and Magic Eye [30] are examples of gaze detection applications. In this kind of software, a webcam is used to identify user's face and eyes or just the user's eyes, and then makes it possible to know where the user is looking, and it is possible too to select options just

by closing eyes (clicking) or using alternate hardware (pedals, buttons, etc). SmartNav, instead of uses a regular webcam and image-processing techniques, it uses an infrared camera, which is capable of identifying special markers that can be in a hat or in the glasses, and then the user can control the computer cursor. This system has the advantage of operating at almost all light conditions, since the infrared beams are not affected by visible light.

Figure 2.7.1 – SmartNav's infrared camera.

Using these technologies makes not only possible to control a computer, but also other hardware, like a wheelchair. Another example of gaze detection application is EyeWritter [31], which makes possible to draw just using the eyes and it is a low cost project.

2.7.3 Other Interfaces

Other applications for disabled users rely on speech and touch. With EasyVoice [32] is possible to insert text without using a keyboard, and then a text-to-speech engine speaks by the user in a phone call (using Skype). Another interesting interface addressed in [25], consists of a simple board paper with RFID tags. The user selects an option pointing an IDBlue pen to an icon drawn in the board, which reads a RFID tag hidden in it and transmits a response via Bluetooth. This is an example of a simply interface for elderly. NavTouch and NavTap [33] are examples of techniques that use mobile phone keypad and touch screens respectively, in order to enable blind users to insert text. Using simple gestures, it is possible to select a letter from the alphabet and then writing text. Other software designed for blind users include Windows' Narrator and JAWS [34], which are screen readers (TTS).

El-E [52] is an assistive robot that lets impaired people manipulating objects. It uses a touch screen and laser devices. El-E is an interesting project as it combines assistive technology (robot) with an adapted interface that controls the robot. So, for example, a user can pick up a pen using the arm of the robot. To do that, El-E is equipped with a camera that lets the user see what the robot is seeing, and then user can send the "grab" order by pointing a laser device to the object on the screen. These laser devices are available in an ear-mounted laser pointer and in a hand-held laser pointer.

2.7.4 Guidelines

Now we present some guidelines regarding designing multimodal applications. Reeves [35], argues that:

- multimodal interfaces must be simple, in order to avoid cognitive load;
- applications must adapt to the needs and abilities of the users and to the context;
- users must be aware if they are being “listened” by the system, that is, which modalities are being detected and interpreted;
- in order to prevent errors, users must be capable to choose which modality to use, and should be also capable of undo or exit some functionalities.

Designing applications for special groups like elderly or impaired must have special concerns. The objective is to provide alternative ways of communication, making these groups more active and more included in society. But ironically technology can lead to more isolation, as new ways of communication and interaction with society can make users more isolated.

There are several approaches in order to design for disabled. Abascal and Nicolle [36] proposed the following:

- **Adaptation of existing systems -** In this approach we try to adapt existing systems that some people cannot use. For example for blind people, the way is making screen readers. Of course this has the problem of lack of generality (in the previous example it only works for blind users). Another problem raised is that they are technology-dependent (for some operating systems and applications this approach could be impossible and/or non-applicable for following versions).
- **Adapting HCI paradigms to assistive technology -** there are HCI techniques that could be applicable in the following points: (1) independence between the interface and the application (e.g. HTML-based screen readers could operate on any browser and operating system); (2) advanced user interface design (specific interfaces for some cases); (3) inclusive design (this is probably the easiest approach, since the objective is not develop for specific groups but for everybody, considering of course, everyone’s limitations).

There are also some ethical concerns that designers must have into account. Considering that these kinds of systems are *attached* to users, the privacy must be carefully considered. For example, the way how the information is stored (it should be encrypted) and for how long (just for strictly necessary time), must be specified. Also if the system is mobile, and can locate the user, he or she must be aware of that. There are also the cost problem, that is, if a system is too expensive, it can be a barrier.

2.7.5 Design Methodologies

We have seen some guidelines, now we will present some methodologies. One methodology proposed [37] consists on: (1) defining preliminary requirements, (2) users survey, (3) systems specification using refined requirements with user's point of view from (2), (4) system design, (5) apply test methodology *Wizard of Oz* with iterative design returning to (3). This can be an useful design technique, as the preliminary study is easier and faster than developing a demo application. It is useful also to understand which are the problems and difficulties that users face, before building a functional model. Another interesting advantage is that new features that designers never thought of can be proposed by the users, as they are completely included in the design process. So, this technique can be used in the beginning, since it consumes relatively less time than building a prototype and it is very effective in order to perceive what the application requirements are.

Another interesting methodology is 5-level Design Approach which is a technique for inclusive design, that is, designing for all, and is referred in [38]. IDC (Inclusive Design Cube) represents graphically the amount (volume) of public that is reachable, and therefore excluded, by the application (see Figure 2.7.2).

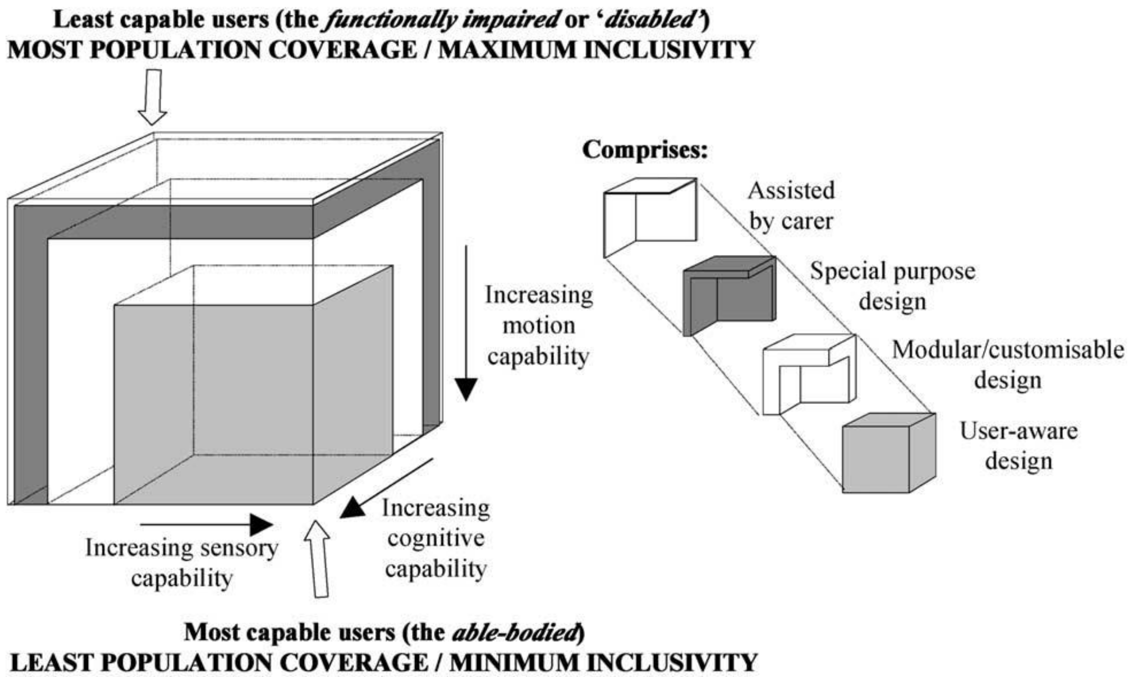


Figure 2.7.2 – The Inclusive Design Cube (IDC) [38]

Each axis represents a capability: motion (user motor/system input), sensory (user perception/system output) and cognitive (user understanding/system clarity). The less the application supports each capability, less is the population reached by it. Able-bodied users (least population coverage) are in a vertex, covering a small volume, and impaired users are in

the opposite vertex which covers the entire cube's volume (most population coverage). 5-level Design takes care of each IDC's capability, and defines the following approach:

- **Level 1 - User needs:** in this level system requirements are specified;
- **Level 2 – User perception:** system output is defined, that is, which are the output modalities that exclude minimum users from target audience;
- **Level 3 – User cognition:** system clarity is defined, given possible cognitive impairments;
- **Level 4 – User motor function:** user input/comfort is established, defining which input modalities will be used;
- **Level 5 – Usability:** a final system evaluation is made.

Designing for elderly and impaired users is therefore different from designing for users without disabilities. There are some questions seen above that have to be considered. Although 5-level Design is an interesting approach, which should even be considered for most applications, for this thesis the primary objective is not designing for all but for handicapped users, so probably the best approach is Interaction Design.

Interaction Design [43] is a process of how to build a system or interface so that the final user will be pleased with the product, that is, the final product reflects the users' needs regarding to usability and functionalities. The first step is to perceive what the users' needs are and how the interaction is made in existing systems or environments. This can be made by observing the users and asking them through interviews and questionnaires. In the second phase, the system is designed and specified using storyboards, sketches, etc. This is accompanied by building a prototype that could be low or high level and vertical or horizontal. Finally the prototype is tested with the users, and is changed according to tests results. This process ends when the prototype is free of errors and meets all the requirements.

In summary we have seen that multimodal applications for disabled people are focused on cognitive, socialization, entertainment and care support. We have also seen some examples of alternative existing solutions for disabled people, as gaze-detection and speech based systems. Finally some guidelines and methodologies were presented, mainly regarding to designing for disabled users.

2.8 Conclusion

In terms of computer interfaces a lot have been made. We evolved from keyboards to more natural interfaces like speech or gesture recognition. By doing this some hardware constraints disappeared. It is still hard or even impossible for a disabled person to interact with a computer using "traditional" interfaces like mouse and keyboard. Now disabled people can use speech and even gaze-based interfaces, in order to simply use a computer. But has we have seen, these interfaces by themselves have some problems that can be avoided just by using more

than one kind of input in an application. By creating a multimodal application, we are making the system more usable and more reachable. Important is to say that a true multimodal interaction should allow the total cooperation between modalities, so that the performance of the use is greater, and should not rely only on accepting one modality at a time.

In conclusion, multimodal interfaces have the objective of making human-machine interaction more natural. They are special indicated for elderly, impaired and handicapped people as they make possible to interact with technologies using alternative ways of interaction.

Chapter 3

Requirements Analysis

In this chapter we present two requirements analysis sessions and a user requirements list, specially targeted for the prototype design, based on the two sessions. The first one was a group interview, with the objective of making preliminary requirements analysis. This first session was very important, as all data collected helped us to define and better calibrate objectives for this thesis and contextualize it. In the second session, ten participants were individually interviewed and invited to do some prescribed tasks in order to perceive how usable current interfaces for a set of common services are.

Important is to say that, a subject number uniquely identifies a participant along all sessions in this thesis (e.g. Subject 2 that participated on preliminary requirements analysis session is the same Subject 2 that participated on requirements analysis session).

3.1 Preliminary Requirements Analysis Interview

Five subjects participated in this section, consisting of two quadriplegics and three paraplegics. The objective of this session was to determine global objectives and tendencies that will help to guide future studies. And so this was the first requirements analysis session that was made, and played an important role on defining objectives to this thesis. Being a preliminary session with the objectives referred above, only a small sample was necessary.

More details about the preliminary requirements analysis interview and its transcriptions can be found in Appendix B. In this section we present results and conclusions of the interview.

3.1.1 Results

All interviewees, consisting on three paraplegics and two quadriplegics, have some limitations using computers and cellphones. To increase keyboard usability, some of them use pens as a typing assistance. Particularly subject 5, due to his specific limitations, needs to resort to a gaze-based interface. This subject added that without this interface he would be unable to use a computer. However, due to specific adjustments needed to properly use the device, including camera adjustments, marker and eye glass placement, he still is not completely independent to use a computer whenever he needs to. Regarding cellphone usage, the majority of the test subjects have pointed small keys as one of their main difficulties while using these devices.

Regardless of these limitations, they all consider that using computers is extremely important, be it for work, entertainment or both purposes (see Figure 3.1.1). These individuals also stated that their computer usage pattern is intense, resourcing to them more than five hours a day (see Figure 3.1.2). Cellphone usage is, however, somewhat more limited, varying according to their personal and professional needs (see Figure 3.1.3).

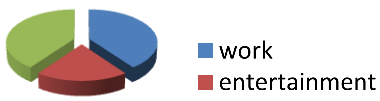


Figure 3.1.1 – Computer usage

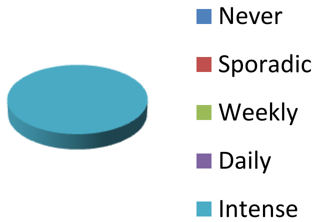


Figure 3.1.2 – Computer usage pattern

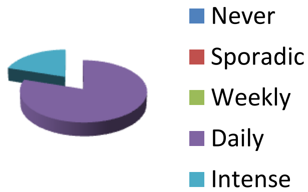


Figure 3.1.3 – Cellphone usage pattern

Considering that handicapped users have to stay at home more that they probably would like, one of the major advantages of computers is communication. The questionnaire's results show that these individuals believe that information and communication technologies (ICTs) would help them keep in touch mainly with family and friends, but also with co-workers and acquaintances (see Figure 3.1.4). They already use communication services, although mainly instant messaging (IM) and email (mainly Hotmail, Gmail and Outlook), as we can see in Figure 3.1.5. Regarding video and audio conferencing, their usage pattern is mostly defined as sporadic.

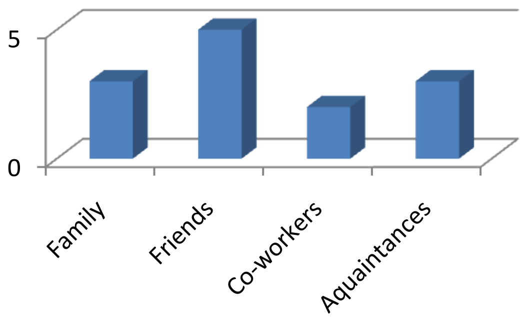


Figure 3.1.4 - Groups that respondents considered they will be in touch with, if communication technologies were used

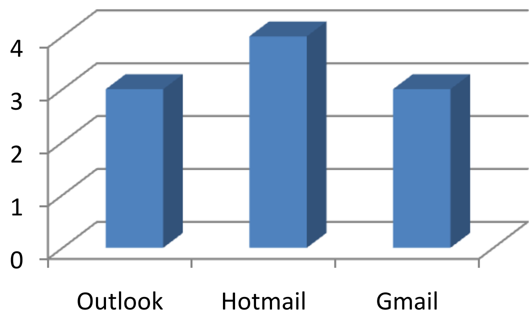


Figure 3.1.5 –Emails clients usage

All test subjects believe that usage of an electronic agenda is extremely important for their professional areas. Regarding currently used agenda software, subject 5 referred that he uses Outlook, while other subjects use either Gmail, integrated with Google Calendar, or nothing. Subject 5 also mentioned that one restriction he constantly notices in the Outlook agenda is the lack of synchronizations between devices such as his home and office computers.

Audio-visual media management is done off-line, with resource to physical storage media such as flash cards or hard drives, as can been seen in Figure 3.1.6. Most of the subjects, however, stated that it would be interesting to try something new, such as a media center or an on-line media storage solution.

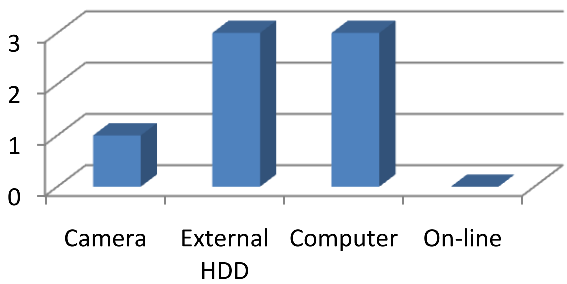


Figure 3.1.6 - Where audio-visual assets are stored

Overall, the test subjects believe that, in order to improve their interaction with digital devices, the availability of a speech input modality would greatly increase their experience. They also would like to have access to the proposed communication services in a ubiquitous way, that is, through a smartphone, a computer and, in the living room, through the television.

3.1.2 Conclusions

The results obtained with this preliminary questionnaire to a sample of the target population have shown that, although ICTs are already being used, in some cases quite intensively, there is still room for user experience improvement.

This opens up the possibility to explore multimodal interfaces in this particular context, especially through the combination of common keyboard and mouse modalities with speech and touch, which will be done throughout this work, as a way to improve mobility impaired users' communication using digital means and services, hopefully producing positive results and improvements in the usability of such services. Of course these interfaces must be capable enough to allow different users, with different kinds of impairments, to use a computer in a natural way. In this study we have noted that quadriplegic and paraplegic have different limitations, and as such, special care must be taken during our prototype development, as requirements are different for each subgroup of motor impaired people.

3.2 Requirements Analysis Session

In order to apprehend which difficulties and limitations motor-impaired people have, a user requirements analysis session was prepared. This session consisted on two parts:

1. **Usability evaluation of current interfaces for communication and entertainment services**: in this part, participants were invited to do some simple tasks with common ICT services, in which they had to use email, agenda, audio/video conference and media center applications.
2. **Usability evaluation of HCI modalities and associated hardware**: in order to realize which HCI modalities and associated hardware can be used, participants were invited to experiment new ways of interaction rather than by using traditional keyboard and mouse, such as, using speech and touch modalities.

In the beginning of each session, the session's goals were explained to each participant, and they were told to fill in a small questionnaire and a consent form (see Table 2 and Appendix B - section B.2.3). During each session, audio and video were recorded for further analysis.

1. On average, how would you describe your computer usage habits (according to scale A)?
2. On average, how would you describe your smartphone usage habits (according to scale A)?
3. On average, how would you describe your cellphone usage habits (according to scale A)?
4. How would you rank your level of easiness of use of a computer (according to scale B)?
5. How would you rank your level of easiness of use of a cellphone (according to scale B)?

Scale A:

1 - Never used

2 - Sporadic usage (less than once a week) 3 - Weekly usage (at least once a week) 4 - Daily usage (less than five hours a day) 5 - Intense usage (more than five hours a day) **Scale B:** 1 - Very Low 2 - Low 3 - Medium 4 - High 5 - Very High

Table 2 – Startup questionnaire

In this section we will describe the goals and the results of each part of the session. For comparison purposes a pilot test was made with a non-impaired person: this is our control test – used to define desired performance.

3.2.1 User Study Participants

As referred in the beginning of this chapter, participants were chosen by Associação Salvador. They were instructed to select participants randomly: with different ages, genders, impairments and computer experience. Below we present the subjects panel used for this session.

Participant	Gender	Age	Career	Impairment type
Control	Male	22	Student	None
Subject 6	Male	37	Unemployed	Quadriplegia
Subject 7	Male	26	Informatics Technician	Paraplegia
Subject 2	Male	43	Informatics Technician	Quadriplegia
Subject 1	Female	26	Life Sciences Technician	Paraplegia
Subject 10	Male	19	Student	Paraplegia
Subject 5	Male	28	General Manager	Quadriplegia
Subject 3	Male	47	Book Keeper	Paraplegia
Subject 9	Male	41	Informatics Engineer	Quadriplegia
Subject 8	Female	54	Technical Assistant	Paraplegic
Subject 11	Male	40	Enologist	Quadriplegia

Table 3 – Subjects panel for requirements analysis session

Participant	1.	2.	3.	4.	5.
Control	5	2	5	5	4
Subject 1	5	1	4	4	3
Subject 2	5	1	4	4	3
Subject 3	5	1	4	5	5
Subject 5	5	1	5	2	2
Subject 6	3	1	5	3	3
Subject 7	5	5	5	5	5
Subject 8	5	1	4	3	3
Subject 9	5	1	4	5	5
Subject 10	4	1	5	3	4
Subject 11	5	4	4	4	4

Table 4 – Startup questionnaire results for each participant (according to Table 2)

3.2.2 Usability evaluation of current interfaces for communication and entertainment services

The main goal of this part was to evaluate the easiness/difficulty of computer usage, in terms of using the following capabilities: communication and entertainment management. The first one is related with email, agenda and audio/video conference. The second one is related with audio-visual information management, in this case the use of a media center.

3.2.2.1 Methodology

The procedure consisted on asking participants to do some tasks (see Table 8). These tasks were the same to all participants and were prepared in advance. So, each participant was given the same information and was treated the same way. In order to prevent ordering effects, tasks did not have any order, and so they were made randomly, this is known as the counterbalance principle (see [43]). During or in the end of each tasks, we asked participants some questions (see Table 5).

1. Do you like the interface? Is it easy to use?
2. If not, what could be improved?
3. If you could interact with this applications using another modalities (e.g. speech, touch), do you think that interaction will be better?
4. Give examples of how you can use new modalities in this application.

Table 5 – Questions asked to participants, in the end or during each task

In some cases, participants took longer than expected to complete a task. And so, for those cases, tasks were aborted and participants were invited to move to the next task. Whenever difficulties on doing a task were detected, participants were helped.

All tests took place on controlled environments: meeting rooms or at participants' homes. For each task it was used the same laptop (see Table 6) for each subject, with same hardware, same software and same state (this means that at beginning of each session all programs were closed, all programs maintained same version – updates disabled – and all desktop, start menu and taskbar icons, remained the same). Every participant made an individual session, with the exception of Subjects 8 and 9 that made a joint session.

- Toshiba Tecra M4
- Intel Pentium M 2 Ghz
- 2 GB RAM
- Windows 7 Enterprise 32 bits
- 55 GB Toshiba hard drive
- For internet access were used (in preference order): WIFI, Ethernet and a CDMA card, depending on what was available on location
- Available browsers: Internet Explorer 8 and Mozilla Firefox 3.6

Table 6 – Characteristics of the laptop used in requirements analysis session

3.2.2.2 Tasks

Being the objective to evaluate in an abstract form, email, agenda, conference and media center services, participants could use any application available for each service. The objective was to ensure that participants would use applications that they normally use, in order not to affect the results. On other cases, they were invited to use the most similar application to the one that they often use, if their preferred application was not available. And, on some cases, they used applications that they have never used before, if they have never tried that service. There were provided test accounts, with username and password, for each application that require an account to login. Below, we present available software applications for each service group.

Email	Agenda	Conference	Media Center
• Microsoft Office Outlook 2007 • Windows Live Mail (v2009) • Gmail • Hotmail	• Microsoft Office Outlook 2007 • Windows Live Mail (v2009) • Gmail Calendar • Hotmail Calendar	• Skype (v4.2.0.155) • Windows Live Messenger (v2009)	Windows Media Center (Windows 7)

Table 7 – Available email, agenda, conference and media center applications

Below we present the description of each task. As a note, we would like to add that all tasks were dictated to all participants, and normally they had only to read accounts' credentials.

Email task	1. Open your email client (login with test account provided) 2. Open any email message in your email box 3. Send an email to apmultimodal@gmail.com and with CC to you (test account that you are using), with the following: a. Subject: *Email de teste!* b. Text: *Olá, este é um email de teste!* *Bem, responde-me.* *PS: Será que escrever o símbolo do euro, é complicado? Deixa cá ver: €.* 4. Open and reply to previous email (3), with following: a. Attach image *teste* that is on desktop b. Write: *Esqueci-me de te enviar isto*
Agenda task	1. Open your agenda 2. Check if there is something appointment for tomorrow 3. Create a new appointment for tomorrow, with the following: a. Go to the movies b. At *Colombo* c. From 16 to 18 d. If available, create a reminder for 1 hour before 4. Delete the previous appointment
Conference task	1. Start a new audio-conference with *ContactoTeste* 2. Stop the audio-conference. 3. Start a new video-conference with *ContactoTeste* 4. Stop the video-conference.
Media center task	1. Start media center 2. Check photos of album *Teste* 3. Start a new slideshow (of photos of album *Teste*) 4. Stop slideshow 5. Check and play a video on that same album (You can use ESC or backspace to back to previous menus or stop presentations)

Table 8 – Tasks description

3.2.2.3 Analysis Methods

Both qualitative and quantitative results will be presented. Qualitative results rely on observations and participants opinions during each task. For quantitative results, we have considered the following:

- **Time to complete a task** – time (in minutes and seconds) since participant was instructed to do a task, until task termination.
- **Number of helps** – number of times participant asked for help or was helped.

For qualitative results we have considered:

- **Result** – it could be:
  - **Successful completion** – participant successful terminated the task.
  - **Completed with errors** – participant completed the task but committed some errors
  - **Incomplete** – participant were told to terminate the task (e.g. if the task was taken too long than expected).
  - **NA** – participant did not do the task.
- **Observations** – our point of view of participants' actions, considering interaction with hardware and software
- **Participants' opinion** – some opinions given by participants about the task, in reply to questions referred on Table 5

3.2.2.4 Results

In this section we present the results grouped for each task. Observations and participant's opinions can be found on Appendix B (section B.2.2).

Email task

Participant	Email client used	Email client normally used	Result	Time to complete (minutes:seconds)	Number of helps
Subject 6	Gmail	Gmail	Incomplete	10:06	1
Subject 2	Windows Live Mail	Hotmail, Gmail	Successful completion	7:18	0
Subject 1	Gmail	Gmail	Successful completion	5:37	0
Subject 5	NA	NA	NA	NA	NA
Subject 9	Gmail	Gmail	Incomplete	5:05	0
Subject 11	Hotmail	Hotmail,	Successful	6:11	0

		Outlook	completion		
Subject 7	Hotmail	MSFT client, Hotmail	Successful completion	4:33	0
Subject 8	Hotmail	Hotmail	Incomplete	5:05	2
Subject 3	Hotmail	Hotmail	Incomplete	6:53	2
Subject 10	Hotmail	Hotmail	Successful completion	5:40	2
Control	Gmail	Gmail	Successful completion	4:13	0

Table 9 – Email task results

Agenda task

Participant	Agenda client used	Agenda client normally used	Result	Time to complete (minutes:seconds)	Number of helps
Subject 6	Gmail Calendar	Never used	Incomplete	3:36	3
Subject 2	Outlook	Rarely uses Outlook	Successful completion	3:00	1
Subject 1	Gmail Calendar	Never used	Completed with errors	3:48	0
Subject 5	NA	NA	NA	NA	NA
Subject 9	Outlook	Outlook	Successful completion	2:26	0
Subject 11	Outlook	Outlook	Successful completion	2:05	0
Subject 7	Hotmail	Hotmail	Successful completion	1:28	0
Subject 8	Gmail Calendar	Gmail Calendar	Completed with errors	2:50	2
Subject 3	NA	NA	NA	NA	NA
Subject 10	Windows Mobile agenda	His cellphone ‘s agenda	Completed with errors	3:33	2

Control	Gmail Calendar	Gmail Calendar	Successful completion	1:35	0

Table 10 - Agenda task results

Conference task

Participant	Conference client used	Conference client normally used	Result	Time to complete (minutes:seconds)	Number of helps
Subject 6	Skype	Skype	Completed with errors	1:51	1
Subject 2	NA	Skype, Windows Live Messenger	NA	NA	NA
Subject 1	Skype	Rarely Windows Live Messenger	Successful completion	2:55	1
Subject 5	NA	NA	NA	NA	NA
Subject 9	Skype	Cisco Telepresence, GmailTalk	Completed with errors	2:13	0
Subject 11	Skype	Skype	Successful completion	1:20	0
Subject 7	Skype	Office Communicator, Windows Live Messenger	Successful completion	1:52	0
Subject 8	Skype	Never used.	Completed with errors	2:13	3
Subject 3	NA	NA	NA	NA	NA
Subject 10	Skype	Windows Live Messenger	Completed with errors	1:53	1
Control	Windows Live Messenger	Windows Live Messenger, Skype	Successful completion	1:50	0

Table 11 - Conference task results

Media center task

Participant	Have already used any media center?	Result	Time to complete (minutes:seconds)	Number of helps
Subject 6	No	Completed with errors	2:47	3
Subject 2	No	Completed with errors	2:01	2
Subject 1	No	Completed with errors	1:56	2
Subject 5	NA	NA	NA	NA
Subject 9	No	Incomplete	1:05	0
Subject 11	No	Completed with errors	1:52	2
Subject 7	Yes	Successful completion	1:22	0
Subject 8	No	Incomplete	1:24	4
Subject 3	No	Successful completion	1:35	3
Subject 10	No	Completed with errors	1:49	0
Control	Yes	Successful completion	1:41	0

Table 12 - Media center task results

3.2.2.5 Results Analysis and Discussion

Email task

The Email task was not only for evaluating email interfaces and its problems, but also to see how motor-impaired people write on a computer.

As described above on Table 8, the email task consisted on using an email client. Participants were told to write a small text for an email, in which they had to use key combinations for uppercase letters and symbols.

Assuming that email is probably the most used feature for communication, all participants demonstrated at least some knowledge on using an email and all referred that the email interface is simple to use. However, there were some problems with email interfaces, as some interfaces could be too complex. For example, Subject 6 had some difficulties on finding an attach icon in Gmail and even on reading what was on screen (we have to consider that he is

quadriplegic and had to approach the screen, which was complicated due to physical barriers) . Another problem registered was with Subject 3 that failed to find CC option in Hotmail. He referred that that option was hidden.

In the case of the email interface, we can consider that new email interfaces for motor-impaired people, must be similar to existent ones and:

- possible simpler with just the essential features: subject, text, attach option and recipients;
- with a good readable interface, where all items and options are understandable and visible, considering that the interface's text/icons are big enough to be seen at some distance from the monitor.

Regarding hardware use, some problems were noticed from quadriplegic participants. As depicted below (Figure 3.2.1), time to complete email task, was completely affected by the writing speed. We noticed that normally quadriplegic users use only one finger at a time or at maximum two fingers to write. They all considered that their writing speed is slow and, using speech for email dictation (and not only), could be very useful. Another noticed problem regards key combinations. Subject 6 referred that normally he uses a bent wire for insert a simple arroba symbol (@). Subject 9, Subject 1 and Subject 2, all used their hands to insert arroba, although they all had problems with it. Subject 9 referred also that key combination limitations could depend of the keyboard format, that is, where keys like Alt or Ctrl are placed. Subject 11 (quadriplegic) mentioned that he normally uses the Sticky Keys functionality. All quadriplegic participants considered that a toolbar with large enough icons, like arroba or the euro symbol, selectable by touch or speech, would be very useful.

	Mean	**Standard deviation**	**Mean Differences**
Control	**04:13**	**-**	**02:03**
General	**06:16**	**01:41**	
Quadriplegic	**06:51**	**01:59**	**01:18**
Paraplegic	**05:32**	**01:00**	
Proficient Quadriplegic	**06:51**	**01:59**	**01:44**
Proficient Paraplegic	**05:06**	**00:47**	

Table 13 – Email task execution time results (tabular form)

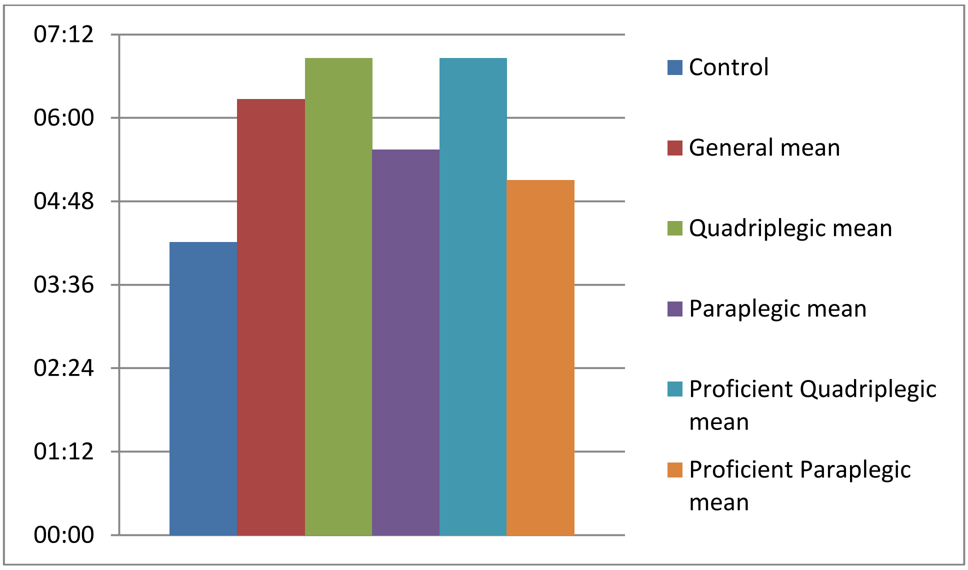


Figure 3.2.1 – Email task execution time results (graphical form)

As depicted in Figure 3.2.1, Control can be considered as the objective. Comparing Control with General mean, we can see that differences are quite large (about two minutes and three seconds). Considering problems that quadriplegic users had, which we have been discussed before, we have divided participants in two groups (quadriplegic and paraplegic) and we have made the mean for each one. Now, the differences are relatively smaller, one minute and eighteen seconds. Considering that Subject 8 and Subject 3 rarely use their email, and they took longer because they are not proficient email users, we have excluded them and we have obtained two new groups (proficient quadriplegic – all quadriplegic participants, and proficient paraplegic, all paraplegic excluding Subject 8 and Subject 3). Now the observed differences between these 2 proficient groups are larger, with 1 minute and forty four seconds separating quadriplegics from paraplegics.

Agenda task

The agenda task consisted on using an agenda, in which participants had to create a new appointment with some characteristics and then delete it.

Since the agenda can be managed more with cursor than with keyboard, none of the subjects had considerable problems in using the available hardware (interaction issues will be discussed in next section). Regarding to the interface itself, participants that normally use an agenda had no problems with it. Participants that have never used an agenda, felt some problems in the beginning, but after a while they managed to use the agenda without problems. One interesting issue raised, was with participants that tried Gmail Calendar. They found that it was difficult to find how to cancel an appointment, as Gmail Calendar is web-based and right-click functionality does not exist (although Hotmail Calendar is web-based too, but it has right-click functionality, much like a desktop-based application).

Most participants considered the interface easy to use. As an improvement, some participants considered that using speech commands for agenda management, as for example, in free hands systems, would be useful. Our development of agenda applications for motor-impaired people rely on using existent applications and possibly adding some speech synthesis (TTS) and recognition (ASR) functionality, as suggested by subjects.

	Mean	Standard deviation	Mean Differences
Control	**01:35**	**-**	**01:15**
General	**02:50**	**00:48**	
Quadriplegic	**02:59**	**00:44**	**00:22**
Paraplegic	**02:37**	**01:03**	
Proficient Quadriplegic	**02:15**	**00:14**	**00:15**
Proficient Paraplegic	**02:30**	**01:28**	

Table 14 – Agenda task execution time results (tabular form)

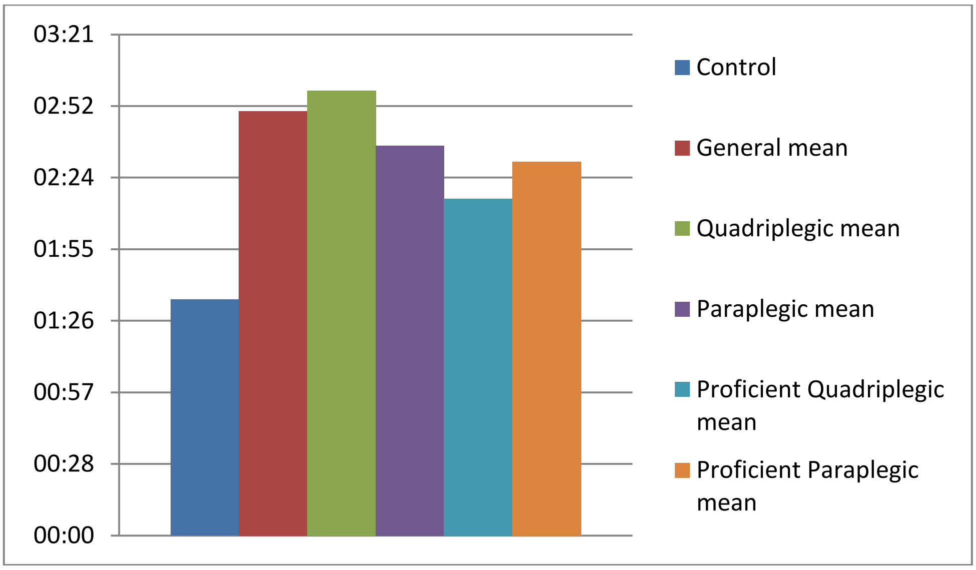


Figure 3.2.2 – Agenda task execution time results (graphical form)

As we can see, there are just small differences between paraplegics and quadriplegics. The observed minor differences can be due to lack of experience in using an agenda application. Therefore, we have considered Subject 9 and Subject 11 as proficient quadriplegics, as they referred that they use an agenda on daily basis. For the same reasons, we have considered Subject 10 and Subject 7 as proficient paraplegic. Now the differences between proficient quadriplegic and paraplegic is smaller, yet paraplegics took longer to complete the task. This can be explained because Subject 10 has used a device that he has never tried before, and so he took longer than expected. Also proficient quadriplegics considered here, are very “proficient”, as they deal with computers every day and their impairment is not so advanced as in other subject cases. For proficient quadriplegic there are no limitations regarding mouse usage,

although for quadriplegic we still consider that mouse offers a limitation (see Figure 3.2.9 in next section). Finally, there is a considerable difference (about one minute and fifty seconds) between Control and General mean, which can be explained by the fact that motor-impaired people take a little longer in using a mouse than non-impaired people.

Conference task

For this task, participants were told to first start an audio-only call and then a videocall.

We have noticed that major participants did not accomplish what was on the script. When asked to first start an audio-conference, most participants started a video-conference instead. For this case, it was not clear why they did it. One hypothesis could be that they stayed confused by the task. Another could be that the interface itself could be source of the problem. All participants used Skype for this task, but most situations Skype was a second choice, as Windows Live Messenger was not working properly in the set-up, for some reason. We have noticed that using Skype could be less error-prone for conference functionality, than using Live Messenger, as buttons for starting an audiocall and a videocall are separated and are understandable.

In terms of user interaction, we have noticed that there were some problems on using mouse and touchpad, in the case of quadriplegic subjects. For paraplegic participants, we haven´t noticed any problems. Quadriplegic participants referred that speech and even touch could improve conference applications usability. In terms of speech they referred that command and control would be nice (e.g. saying a contact's name or "start call"). Touch would be probably helpful for contacts' selection. Important is to say that, quadriplegic subjects considered that audio and video conference are preferable to instant messaging, since they have some writing problems (as we have seen above in the email task).

As a conclusion of this user study, in what concerns developing HCI for audio and video conference, applicable for motor-impaired people, we can derive the following recommendations:

1. The interface must be simple with audiocall and videocall buttons separated and understandable;
2. Developers must consider that audio and video conference are more easy and convenient to use than instant messaging;
3. Using speech for command and control could improve conference applications usability;
4. Simple touch could be helpful for interaction (with medium to large visible icons) and for contacts' selection (by dragging);

	Mean	Standard deviation	Mean Differences
Control	**01:50**	**-**	**00:12**
General	**02:02**	**00:29**	
Quadriplegic	**02:04**	**00:39**	**00:05**
Paraplegic	**01:59**	**00:11**	

Table 15 – Conference task execution time results (tabular form)

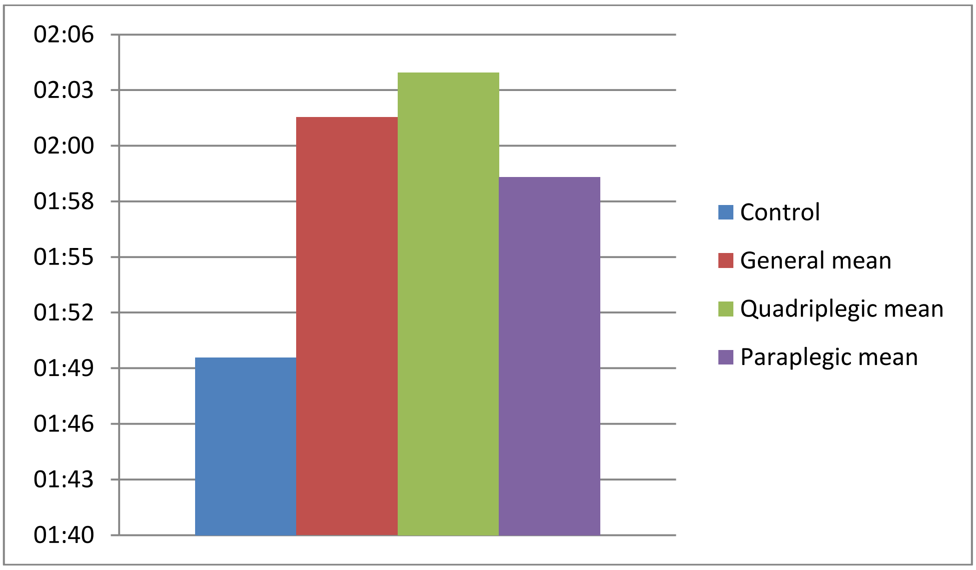


Figure 3.2.3 – Conference task execution time results (graphical form)

As in the agenda task, dealing with a conference application requires less or no keyboard interaction and so, results (Figure 3.2.3 – Conference task execution time results (graphical form), show that there are practically no differences between quadriplegic and paraplegic subjects and, even with Control. We have to refer again that most participants did not accomplished successfully this task, that is, they only did half of it, and so we have to consider that differences between motor-impaired and Control are actually larger (as in the agenda task).

Media center task

In the media center task, subjects were invited to try Windows Media Center (as most of them have never used a media center application before).

Regarding media center's interface, participants had some difficulties on terminating slideshows and video playbacks, as they could not find out how to do that. Most of them even closed the media center, because they clicked on the media center's exit button (X on right-top of the screen). In terms of the interaction itself, we have noticed that there were no problems to register.

Quadriplegics considered that using speech for media center control, would be a good idea. For example, Subject 11 said that speech could be used for controlling slideshow

commands, by uttering commands like “pause”, “next” or “previous”. Gestures could also improve usability, as for example, by pressing the right part of the screen, this could mean “next photo” and, by pressing the left part, “previous photo”.

	Mean	Standard deviation	Mean Differences
Control	**01:41**	**-**	**00:04**
General	**01:45**	**00:29**	
Quadriplegic	**01:56**	**00:36**	**00:23**
Paraplegic	**01:32**	**00:12**	

Table 16 – Media center task execution time results (tabular form)

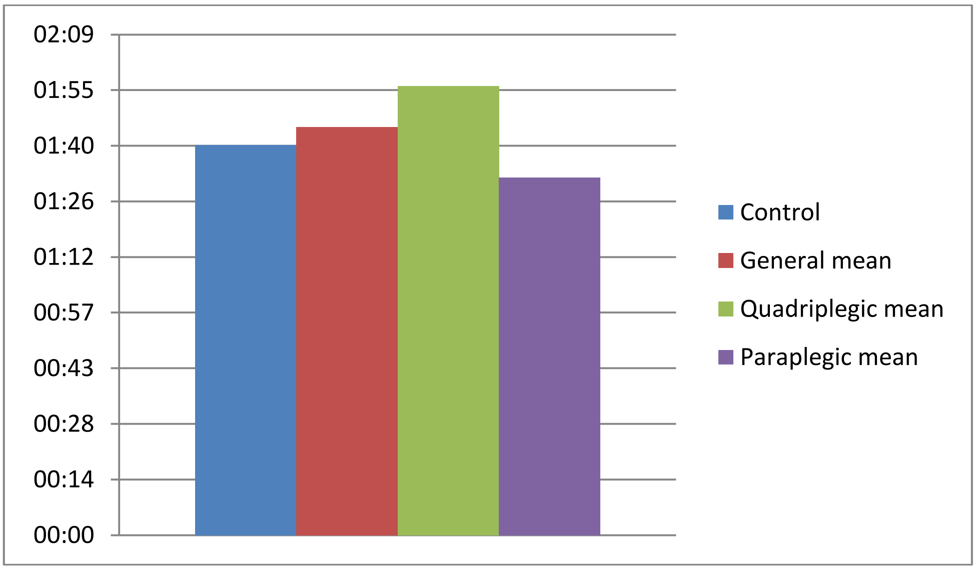


Figure 3.2.4 – Media center task execution time results (graphical form)

In order to access the media center, participants have used touchpads, mice and keyboards. The observed differences between quadriplegic and paraplegic are quite small (Figure 3.2.4 – Media center task execution time results (graphical form). Even so, we can conclude that quadriplegics take a little longer in using interfaces like touchpads, mice and keyboards (excluding writing - as we have seen before, differences between quadriplegics and paraplegics in writing using a keyboard are more significant). An interesting result is that paraplegics took even less than Control, which suggests that they have practically no problems on using traditional hardware interfaces and that, for this type of interfaces, they can considered as behaving like Control.

3.2.2.6 Conclusions

In this study, that considered the usability of common ICT services (e-mail, agenda, conference and media center), we have seen tasks that made participants use more pointing devices (agenda, conference and media center), and a task in which participants had to write a small text (email). By analyzing results, we can conclude that whenever quadriplegic users need

to do something that requires writing, they will be less productive as paraplegic and non-impaired people with currently available traditional HCI. For those cases, speech used in dictation was considered as an excellent alternative, as it could decrease writing time. Regarding touch and speech for command and control, participants considered that those modalities could improve usability, but differences will not be significant.

3.2.3 Usability evaluation of HCI modalities and associated hardware

In the previous section we have seen some difficulties that motor-impaired people have on using standard interface hardware such as keyboard and mouse. In this section, we will go deeper with this study and present how in fact quadriplegic and paraplegic users interact with current interface hardware, what are their limitations, and analyze if there are limitations or not regarding alternative HCI modalities (and its associated hardware), such as speech, touch and gesture. This section presents the second part of the requirements analysis session, which focus on alternative modalities for interaction, which imply the use of specific devices.

3.2.3.1 Methodology

In order to evaluate current interface hardware, we have simply observed how participants used a computer in their daily lives, that is, we have recurred to the first part of this study where participants made a couple of tasks as seen above.

To evaluate how alternative ways of interaction could be or not helpful, we have designed simple tasks for each device and modality, as described below (see Table 17).

Device (modality)	**Task**	**Observations/Objectives**
Tablet PC (Speech)	Try to say something to the computer (e.g. count from 1 to 6).	We have used an experimental program developed in MLDC that includes a command and control ASR engine (European Portuguese). Beyond perceiving if motor-impaired people could use speech without problems, this task also served for apprehend if they could put a headset by their own.
Tablet PC (Touch using a stylus device)	Try to draw something using Paint. Try to write something (using handwriting) on Office Word.	For this task we have used Microsoft Paint and Microsoft Office Word handwriting capabilities (English only). The main objective was to perceive if handwriting is more appropriate than using a keyboard for disabled people. Participants tried to write on the tablet's screen, being it on

		vertical or horizontal.
None (Multi-touch)	If this screen had Multi-touch capabilities, can you use them? Can you show us?	At the time, we did not have a multi-touch device, and so this was only a conceptual task. Participants were told do simulate multi-touch gestures.
Smartphone (3D gesture)	Try *Dice Game*. Shake the phone and watch dices being launched.	This task aimed to check if motor-impaired had problems on holding a smartphone and shook it (using its accelerometer).
Smartphone (Touch/gestures)	Try to start *Asphalt 4 Elite Racing* or *Resco Snake* or *Resco Bubbles* (Go to Start menu – drag down and select Games).	In this task participants tried to use the smartphone (using drag gestures and touch) and then they played a game that required spatial (3D) gesturing with the smartphone (using the accelerometer) and touch. This task had the objective of perceiving if motor-impaired people had limitations on producing 3D gestures with a smartphone.

Table 17 – Usability evaluation of HCI modalities and associated hardware tasks

As a Tablet PC we have used the laptop referred in Table 6. As a smartphone we have used the following set-up:

- Samsung Omnia 2 I8000
- Windows Mobile 6.5

Table 18 – Characteristics of the smartphone used

After all participants have terminated the tasks described above, they were told to answer a small questionnaire presented below.

1. Rank in terms of easiness/difficulty of interaction the following modalities:
    a. Touch (computer):
    b. Speech
    c. Touch (smartphone)
    d. 3D Gestures (smartphone)
    In the following scale: (1) impossible; (2) very difficult; (3) difficult; (4) medium; (5) easy; (6) very easy
2. From all modalities that you have tried, which one you liked most?
3. And less?
4. Considering that the prototype that we will develop will use modalities that you have tried earlier, in what way that prototype would improve your daily work?

Table 19 – Interfaces hardware and multimodal HCI usability evaluation questionnaire

3.2.3.2 Results

Below, we present for each participant the results of interface hardware and HCI modalities usability evaluation session.

Modality	Results
Speech	Participant managed to put on the headset with some problems. He could speak without any problem and considered that was easy. Subject considered voice interaction strange to use and as such he wouldn't like to use it on a daily basis.
Touch using a stylus	Subject used stylus without problems, but for drawing purposes only. He considered easy to use.
Multi-touch	Impossible, only one finger at a time. In that case, touching would be easy to do, including simple dragging.
3D Gestures	Participant managed to shake and perform spatial gesture with the smartphone without problems. He could also hold the smartphone with his hand.
Touch/gestures (touch screen)	Subject had some difficulties on performing dragging gestures on screen and the smartphone had to be fixed to the wheelchair. He used touch without problems, although he considered that icons could be bigger.
Keyboard	Subject could only write on the keyboard with only one finger at a time.
Mouse/touchpad	Participant normally uses touchpad, but with some difficulties.

Table 20 – Subject 6 evaluation results.

Modality	Results
Speech	Participant managed to put on the headset with no problems. He could speak without any problem and considered that it was easy to do and a good alternative.
Touch using a stylus	Participant used stylus device without any problems, and he also considered handwriting very easy and equivalent to keyboard writing, but screen must be in his lap (horizontal) (see Figure 3.2.7).
Multi-touch	Subject considered multi-touch impossible to do. He also considered that he could only using touch with one finger at a time and also considered that having two arms in the air would be complicated for him and even impossible for some quadriplegics.

3D Gestures	Participant managed to shake the smartphone without any problems. He could also hold smartphone with his hands, but with some difficulties.
Touch/gestures (touch screen)	Subject had no difficulties on holding the smartphone and in performing dragging gestures and touch, but for that, he had to use both hands. He considered that gestures were easy to do, although not very easy. Subject considered also that icons must be larger, since he used his knuckles.
Keyboard	Subject write with two knuckles at a time. Alternatively he can use two pens or pencils, which doubled writing speed approximately (see Figure 3.2.6).
Mouse/touchpad	Participant is left-handed only with mouse, because he uses the back of the finger (see Figure 3.2.5).

Table 21 – Subject 2 evaluation results.

Figure 3.2.5 – Subject 2 using mouse.

Figure 3.2.6 – Subject 2 using pencils to write (left) or his knuckles (right)

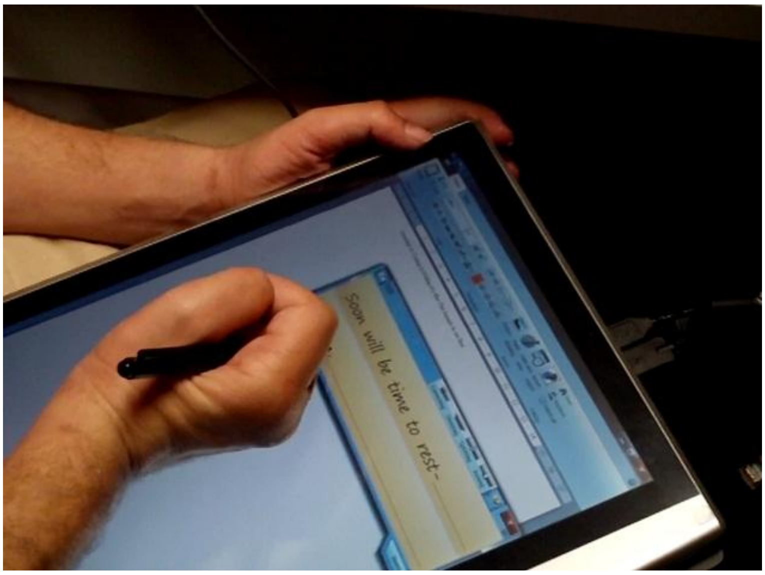


Figure 3.2.7 – Subject 2 using handwriting on Tablet PC

Modality	Results
Speech	Subject did not tried speech in this session, but he tried ASR applications in the past, and he considered that those applications did not work on open space environments.
Touch using a stylus	NA
Multi-touch	Subject considered multi-touch impossible to do. For simple touch, participant considered that if display screen was on his lap, simple touch would be possible, but still hard to do.
3D Gestures	It was impossible for him to hold a smartphone.
Touch/gestures (touch screen)	Subject could not hold smartphone and so, he tried with smartphone attached to the wheelchair. He did some dragging gestures and simple touch with some difficulties. Subject referred also that if icons were bigger interaction would be better (see Figure 3.2.8).
Keyboard	Participant said that he does not use a keyboard for writing, but instead he uses a gaze-detection system that interacts with an onboard keyboard.
Mouse/touchpad	NA

Table 22 – Subject 5 results

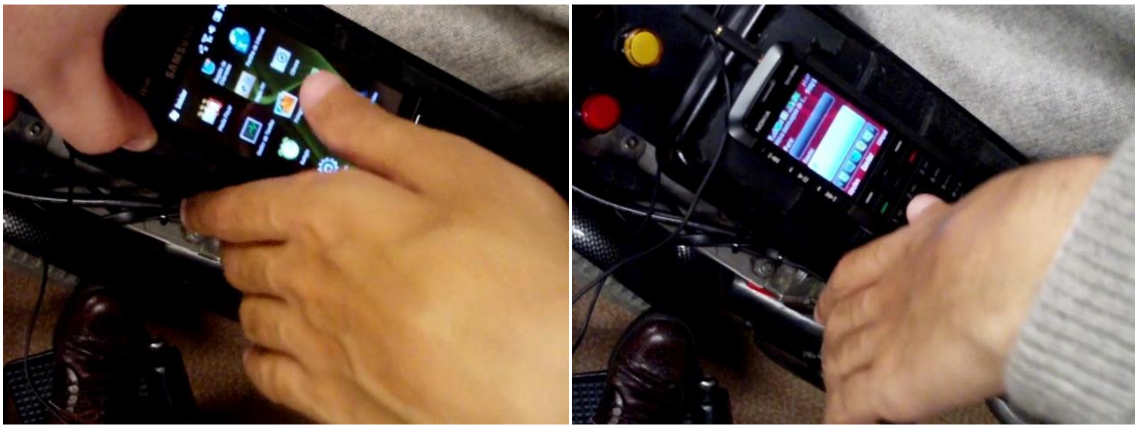


Figure 3.2.8 – On left we can see Subject 5 using smartphone and on right Subject 5 is using his cellphone

Modality	Results
Speech	Subject 1 managed to put on headset without problems, and considered even that putting on an auricular would be better. Participant used speech without any problems.
Touch using a stylus	Subject could not use stylus device on a vertical screen display. With the display placed on the horizontal she used stylus but with some difficulties (see Figure 3.2.11). She felt that handwriting on screen is not practical, as the pressure must be superior to the normal writing condition (on paper).
Multi-touch	Subject considered Multi-touch impossible to do. Regarding simple touch she considered that probably it would be possible, but if icons were large enough.
3D Gestures	It was impossible for her to hold smartphone.
Touch/gestures (touch screen)	Subject could not hold a smartphone. She did some dragging gestures but they were not successful, and subject considered impossible to do dragging. Subject could do some simple touch on screen, but she noted that icons were too small.
Keyboard	Participant wrote on the keyboard with one finger at a time (see Figure 3.2.10).
Mouse/touchpad	Subject used mouse with both hands (see Figure 3.2.9).

Table 23 – Subject 1 evaluation results.

Figure 3.2.9 – Subject 1 using mouse

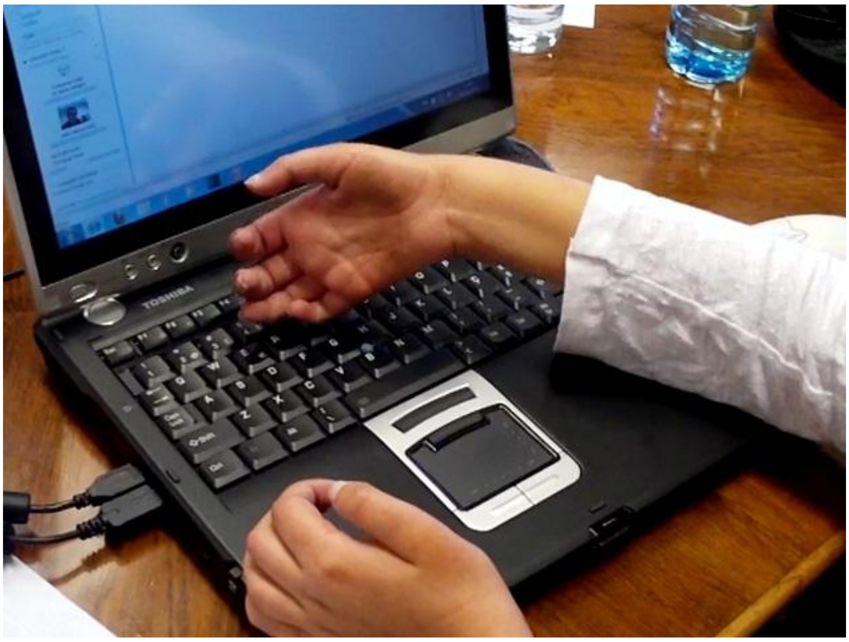


Figure 3.2.10 – Subject 1 using keyboard

Figure 3.2.11 – Subject 1 using Tablet PC with stylus device

Modality	Results
Speech	Participant put headset with no problems. He used speech with no problems too.
Touch using a stylus	Subject used stylus device with no problems, both with display screen on horizontal or on vertical, although he considered easier with display screen on horizontal. Subject 9 also considered that handwriting on a tablet PC would be faster than writing on keyboard.
Multi-touch	Subject considered Multi-touch hard to do. Simple touch would be easier, as he normally uses touchpad.
3D Gestures	Participant hold smartphone with some difficulties (see Figure 3.2.12), and he could also shake it. He referred that he can play tennis on Wii, which it's interesting.
Touch/gestures (touch screen)	Subject managed to use smartphone with some problems. He did dragging gestures and simple touch, but we noticed that icons were very small, which hampered usability.
Keyboard	Participant wrote on keyboard with one finger at a time. He noticed that key combinations were very difficult to do and even they are keyboard dependent.
Mouse/touchpad	Subject used touchpad and considered it easier than mouse.

Table 24 – Subject 9 evaluation results.

Figure 3.2.12 – Subject 9 playing *Resco Snake*

Modality	Results
Speech	Participant put headset with some difficulties and he used speech with no problems. He also considered that putting on an auricular would be easier.
Touch using a stylus	Subject could not use stylus device on a vertical display. With display on the horizontal he could use the stylus but yet with some difficulties. Participant referred also that the display's ideal position is inclined. Subject 11 tried handwriting but computer failed to recognize his words (only uppercase letters), and he considered that writing on the keyboard is still better and faster.
Multi-touch	NA
3D Gestures	Subject managed to issue 3D gesture with the smartphone (equipped with an accelerometer) without problems, and he considered even that those gestures could be considered as physiotherapy.
Touch/gestures (touch screen)	Subject hold the smartphone with difficulties. He could not do dragging gestures, but he did some simple touch on screen, although he considered that icons were too small.
Keyboard	Participant wrote using the keyboard with two fingers at a time (see Figure 3.2.13). For key combinations he used the Sticky Keys.
Mouse/touchpad	Subject used the mouse with both hands.

Table 25 – Subject 11 evaluation results.

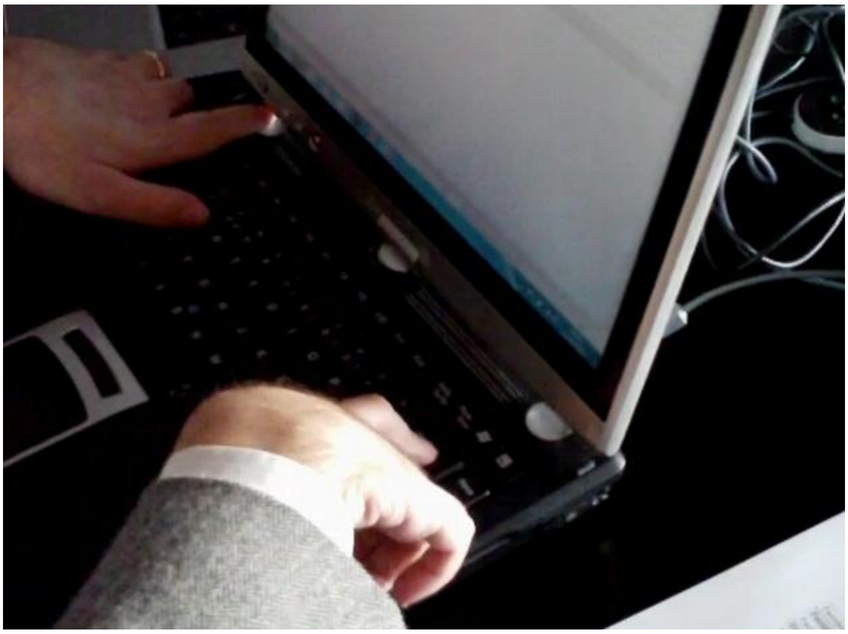


Figure 3.2.13 – Subject 11 using keyboard.

Modality	Results
Speech	Participants used speech without problems. Subject 10 used also Windows Speech Recognizer that he found very interesting, and he considered a good alternative for both command and control and dictation, but it has yet many faults. Subject 7 considered that speech is not a good alternative yet, but it can be used for alternative tasks as authentication (speaker identification by recognizing user's voice).
Touch using a stylus	Participants used stylus device with no problems. Subject 10 considered handwriting interesting but it is not yet a good alternative for some cases.
Multi-touch	Subjects considered that Multi-touch gestures easy to do.
3D Gestures	Subjects used 3D gestures without any problem.
Touch/gestures (touch screen)	Subjects did not have any difficulties on using smartphone's touch capability.
Keyboard	No problems were registered with keyboard use.
Mouse/touchpad	Subjects used mouse without problems.

Table 26 – Subject 7, Subject 8 and Subject 10 evaluation results.

Modality	Results
Speech	NA
Touch using a stylus	NA
Multi-touch	NA

3D Gestures	Subject used 3D gestures without any problem.
Touch/gestures (touch screen)	Subject did not have any difficulties on using smartphone's touch capabilities.
Keyboard	Subject only used one finger at a time, but that's because he is not experienced with computers.
Mouse/touchpad	Subject used mouse without problems.

Table 27 – Subject 3 evaluation results.

Now we present the answers for each participant when replying to the opinions questionnaire (see Table 19).

Participant		Question 1. a)	Question 1. b)	Question 1. c)	Question 1. d)
Subject 1		2	6	1	1
Subject 2		5	6	4	3
Subject 3		NA	NA	NA	NA
Subject 5		3	4	3	1
Subject 6		5	5	5	5
Subject 7		6	5	6	6
Subject 8		4	5	5	3
Subject 9		4	5	5	3
Subject 10		4	6	5	5
Subject 11		3	5	3	4
Mean		4 (Medium)	5,(2) (Easy)	4,(1) (Medium)	3,(4) (Hard)
Standard deviation		1,23	0,67	1,54	1,74
Paraplegic	**Mean**	4,67 (Easy)	5,(3) (Easy)	5,(3) (Easy)	4,67 (Easy)
	Standard deviation	1,15	0,58	0,578	1,53
Quadriplegic	**Mean**	3,67 (Medium)	5,17 (Easy)	3,5 (Hard to medium)	2,8(3) (Hard)
	Standard deviation	1,21	0,75	1,52	1,60

Table 28 – Questionnaire results for question one.

Participant	Question 2.	Question 3.
Subject 1	Speech	Touch on smartphone and

		handwriting on tablet PC
Subject 2	Speech	Smartphone accelerometer
Subject 3	NA	NA
Subject 5	NA	NA
Subject 6	Touch screen (digital stylus input)	Keyboard
Subject 7	Smartphone's touch screen, keyboard, mouse	Speech
Subject 8	Speech	NA
Subject 9	Smartphone's touch screen	Speech and handwriting on tablet PC
Subject 10	Speech	Handwriting on tablet PC
Subject 11	Speech, accelerometer interaction	Handwriting on tablet PC

Table 29 – questionnaire results for questions two and three

Participant	Question 4.
Subject 1	Subject 1 replied that voice recognition would greatly improve her daily interaction with computers, especially for work related tasks in which European Portuguese dictation support would help in text composition.
Subject 2	Due to his job requirements, Subject 2 is already using keyboard to typing. He believes, however, that voice recognition in command and control mode would help him interacting with complex environments.
Subject 3	NA
Subject 5	Subject 5 believes that voice recognition would greatly improve his text writing and computer interaction experience, especially when he's at home writing short texts or, larger texts at the office later in the day.
Subject 6	Subject 6 added that key combinations are hard for him to enter, believing that a virtual keyboard with special characters would simplify his interaction.
Subject 7	Subject 7 believes that if speech recognition was more developed tan it is today, it could help him in his daily tasks, especially for authentication (speaker identification) and dictation purposes.
Subject 8	Subject 8 believes that voice interaction, especially while in dictation mode, would help her a lot during her daily tasks.
Subject 9	Subject 9 believes that if speech recognition was more evolved that it is today, it could help him in his daily tasks.
Subject 10	Subject 10 believes that these alternative modalities won't significantly influence his daily activities.
Subject 11	Subject 11 noted that speech interaction, especially in dictation mode,

	would substantially increase his interaction with the computer. He noted that he eagerly waits for full dictation support in European Portuguese, estimating that he would use dictation in 90% of his daily computer interactions and voice command and control in 10% of his interactions.

Table 30 – questionnaire results for question four

3.2.3.3 Results Analysis and Discussion

As in the first part of this study, we can divide participants into two main groups: paraplegics and quadriplegics. Paraplegics used current interface hardware (mouse and keyboard) with no problems, and alternative interfaces were considered to be easy to interact. We can therefore conclude that restrictions are specially observed in the quadriplegics group.

Quadriplegics adapted themselves to interfaces and we have seen that they had different ways of using keyboard and mouse or touchpad. But there are yet some limitations and restrictions on interactions, and so, our recommendations derived from this user study for developing multimodal HCI specially target for **quadriplegics,** are:

- **Mobile interfaces**
  - Icons must be big enough, considering that quadriplegics use their knuckles or their palm finger of the thumb or the forefinger;
  - Interfaces must have good readability – not only icons must be big enough, but also texts or other data, because we have to consider that mobile phones will need to be attached to the wheelchair;
  - Multi-touch should be avoided, as all quadriplegics considered it hard or even impossible to perform;
  - There must be alternatives for gestures (drag or 3D gesture). For example, for menu browsing, the interface could accept dragging gestures or even use of 3D gesture (by means of an accelerometer in the smartphone), but, it is advisable to have too buttons to do that; as many quadriplegics failed to use both 3D gesture and dragging gesture;
  - In order to assure mobility, mobile interfaces should have a feature set as close as possible from the desktop one, meaning that mobile interfaces should offer user all functionalities that are offered on similar interfaces in the desktop platform.
- **Computer interfaces**
  - As in mobile interfaces, touchable icons must be large enough;
  - Multi-touch should be avoided, but single touch can be considered;
  - Interfaces must have good readability;
  - Touch should be minimized, if the display screen is on a vertical setting;
  - Display screen should be reachable;

- Keyboards and mice/touchpads are still good alternatives;
- Key combinations must be avoided, as most quadriplegics felt difficulties on using them. Instead, everything that requires key combination should be placed on a sidebar, with large enough icons, selectable by touch, speech or by cursor.

- For mobile and desktop interfaces, speech should be present, especially in dictation mode, but also in command and control mode. Speech for dictation mode was considered to be very important and prone to improve quadriplegics writing performance a lot. Speech could be accepted from various kinds of peripherals like headset (that we have seen to have almost no restrictions), auricular (participants considered easier to put on than headset) or microphone.

Below (Figure 3.2.14, Figure 3.2.15 and Figure 3.2.16), we can see results from Table 28 in graphical form. As we can observe speech was considered the easiest modality to use, followed by touch on smartphone. The hardest modality was 3D gestures on smartphone, that is, using its accelerometer.

Dividing our analysis in two groups, we can see that paraplegics have no problems on using interfaces and so, on average, they considered all interfaces easy to interact with. On the other hand, quadriplegics considered speech as the easiest modality, while other modalities were considered to have hard to medium difficulty. Touch on smartphone have dissonant results, probably because some quadriplegics considered that if smartphone was attached to their wheelchair, touch interaction would be easier than if they have to hold smartphone on their hands, which in some cases is impossible.

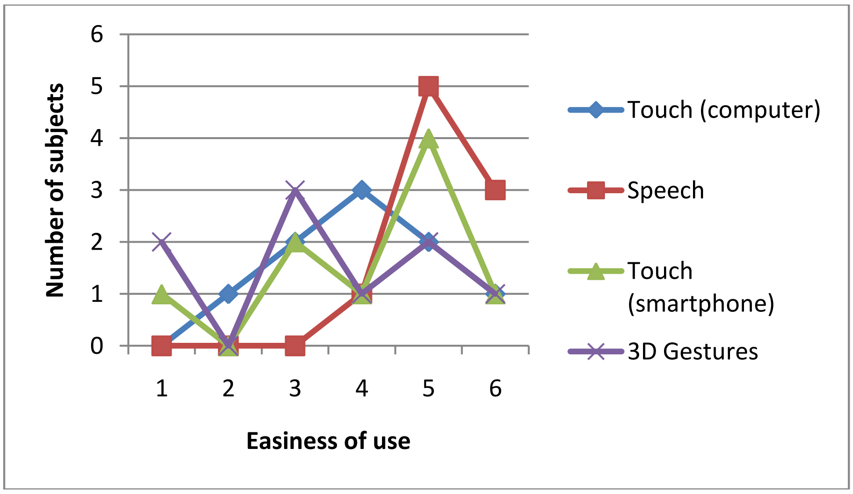


Figure 3.2.14 – Participants opinions of easiness of use of interfaces/modalities, from one (impossible) to six (very easy)

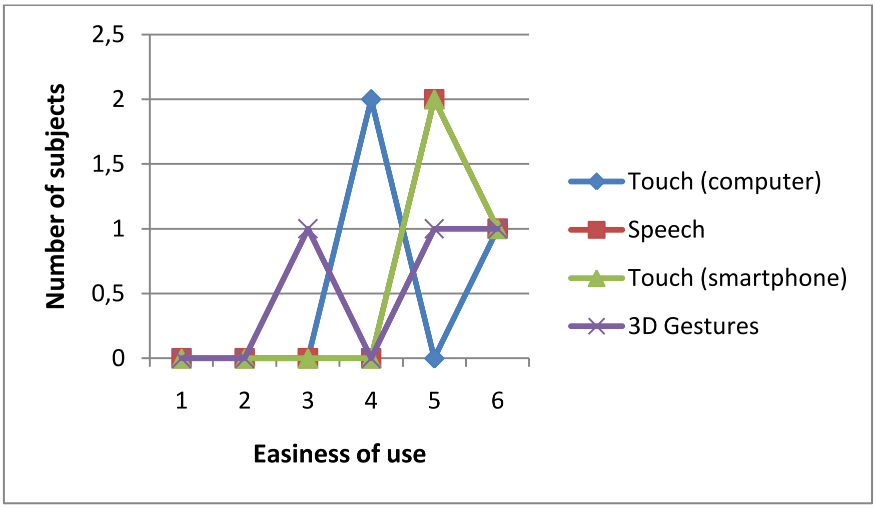


Figure 3.2.15 – Paraplegics opinions of easiness of use of interfaces/modalities, from one (impossible) to six (very easy)

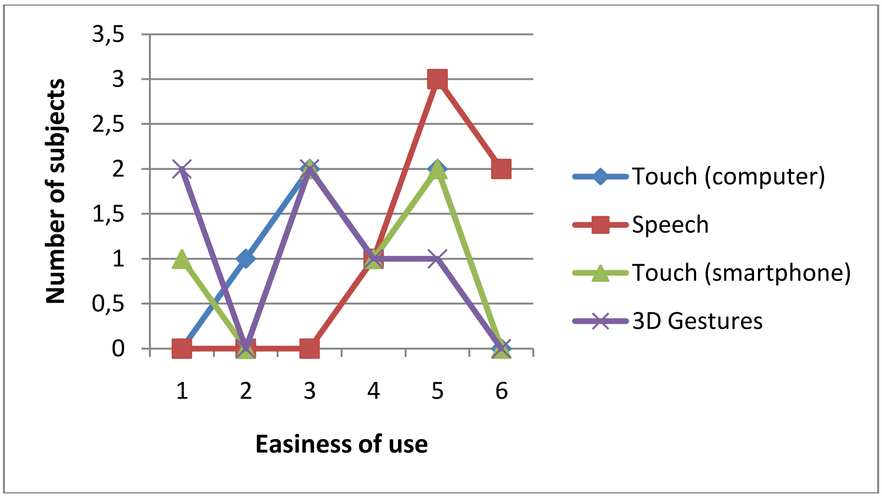


Figure 3.2.16 - Quadriplegics opinions of easiness of use of interfaces/modalities, from one (impossible) to six (very easy)

3.3 User requirements

After analyzing and compiling all subjects' restrictions and opinions in our usability evaluation study, we presented below, on Table 31, the user requirements for HCI for an integrated Personal Life Assistant, specially targeted for motor-impaired users, that will take care of accessing such services, as email, agenda, conference and media center.

ID	User requirements for HCI
	E-mail
1	The system should let user receive email messages from everyone
2	The system should let user send email messages to everyone (not only to registered

	contacts but also to any email address)
3	The email interface should have essential features (subject, text, attach option and recipients insertion option) and buttons/icons should be big enough in order to be touchable
	Agenda
4	The agenda interface should present clearly all options for appointment deletion, edition and creation
5	The agenda interface should avoid timeslot selection for appointment creation
6	The agenda should let user to receive notifications regarding appointments
	Conference
7	The conference interface should have separated and understandable buttons/icons for audiocall and videocall
8	The system should let user make audiocalls and videocalls
	Media Center
9	The media center interface should have understandable buttons for slideshow and video playback termination
10	The media center should allow speech interaction in command and control mode
11	The system should allow the user to manage his multimedia assets offline (media center)
	General HCI Requirements
12	All interfaces on all devices should have big enough icons/buttons in order to be touchable
13	All interfaces on all devices should have large enough texts/icons in order to be readable at some distance
14	All functionalities should be usable by current/traditional interface hardware (keyboard and mouse)
15	Data must be synchronized between devices (mobile and desktop)
16	The system should be available anytime, anywhere, through various devices (desktop and mobile at least)
17	Physical peripheral interaction should be avoided (e.g. usb drive removal, cd/dvd insertion/removal, computer power turned on)
18	There must be alternatives to key combination on all interfaces, such as: using an options sidebar selectable by touch, cursor or speech
19	Whenever user needs to insert large texts, there should be helpers as auto completion or alternatives like using dictation instead
20	Interaction should not be done exclusively by using multi-touch or speech interfaces
21	User should be able to interact with all interfaces using any of the available modalities, such as, speech, touch, gesture, keyboard/mouse
22	All interfaces should accept voice commands

23	Touch should be minimized if display screen is on a vertical setting
24	Gestures (touch) and 3D gestures should be minimized and there must be alternatives to them.
	Specific mobility requirements
25	Mobile device should be completely usable if it just stays fixed (e.g. on a wheelchair)
26	Mobile device should be easy to turn on (avoid small power on buttons)

Table 31 – User requirements

3.4 Summary

In this chapter we have presented our user evaluation studies, including the adopted methodology and obtained results. First a selected group of motion impaired users with paraplegia and quadriplegia, all members of Associação Salvador, was interview in the context of a preliminary requirements analysis and subjects considered that current computer and mobile interface devices and HCI are still restrictive for their daily work and lives. On the second part, we have presented a more structured requirements analysis session, in which we found what are the major problems, for motor-impaired persons, on using current interface hardware and HCI modalities, what how new ways of interaction can be used to improve usability and what are the limitations of current HCI to commonly used ICT services, such as email, conference, agenda and media center, for these users. Finally, and by analyzing all difficulties of motor-impaired people on using these interfaces, we have derived a user requirements list for multimodal Human-Computer Interface of a Personal Life Assistant, specially targeted for motor-impaired users, that will take care in an integrated way, of accessing such services, as e-mail, agenda, conference and media center.

Chapter 4

Prototype specification and development

As we have seen in the previous chapter, motor-impaired persons still have some limitations on using current interfaces. In the scope of this thesis, from the users requirements captured, whose results were presented in the previous chapter, we have specified, developed and tested a prototype with the objective of improve the usability of motor-impaired people regarding agenda, email and conference service interfaces. In this chapter we will define the prototype's architecture and functionalities.

4.1 General description

Taking into account restrictions, limitations, guidelines and user requirements (Table 31), extracted from the user study presented in the previous chapter, we specified that the prototype would have to:

- Be connected with remote servers or services for email, agenda and conference management;
- Have a desktop and a mobile component capable of dealing with touch and speech (in addition to traditional mouse and keyboard interfaces);
- Offer an integrated solution to connect mobile and desktop devices with services or servers.

Non-functional requirements are also presented:

- For email and agenda management *Exchange Server 2010* should be used;
- For conference and *Outlook Voice Access* with Exchange Server, *Office Communications Server 2007 R2* should be used;

- Given the availability of technology, speech recognition should be done with server-side computing using *UCMA Speech 2.0* [62].

After compiling the previous requirements, a final architecture for the prototype is presented. As we can see from Figure 4.1.1, there are two main regions: (1) **Home** - which represents the devices that user will have at home; (2) **Backend** - where backend remote servers are located. At home we have a mobile device (smartphone), a desktop device and a server (PLA Server) that works as a mediator between mobile, desktop and backend.

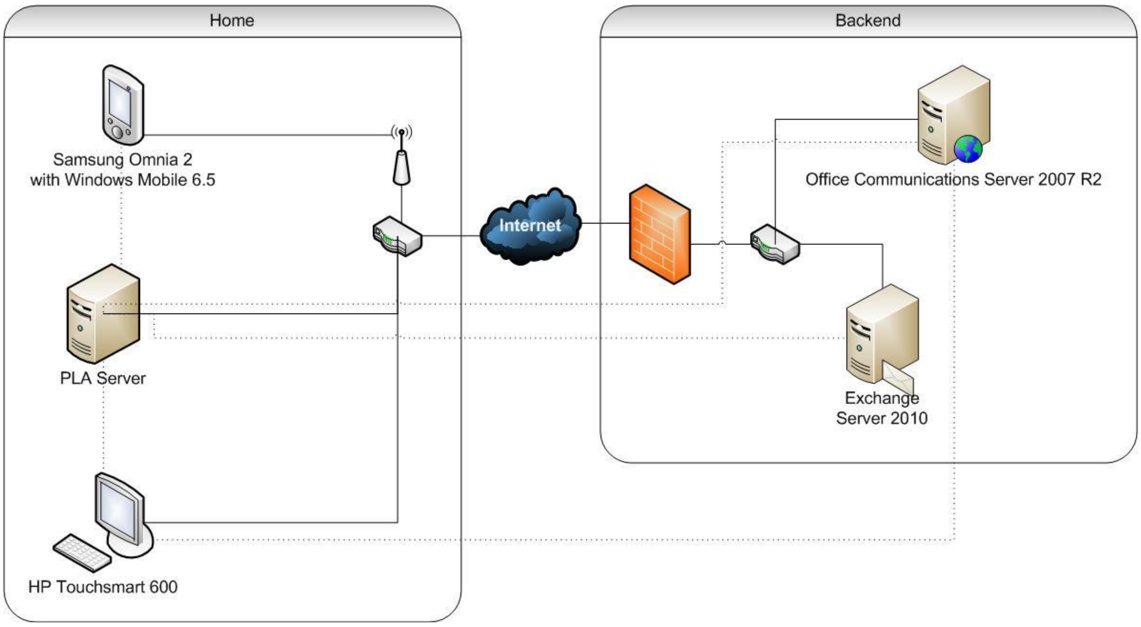


Figure 4.1.1 – Prototype's physical architecture

In the next section, details about each component will be presented.

4.2 Hardware and technologies description

4.2.1 Mobile Device

As mobile device, *Samsung I8000 Omnia II* was chosen. One reason that lead to its choice was that the device has a large touchable screen and its API (*Samsung Windows Mobile SDK 2.1* [57]) enables access to practically all smartphone's features (accelerometer API, orientation API, vibrate API, etc). Another reason relies with the fact that its operating system is Windows Mobile 6.5 Professional, which already supports by default touch functionality and 3D gestures implementation. For the speech implementation, *OpenCF.NET Smart Device Framework* [58] was used to play sounds (for TTS) and, a low level library (developed at MLDC), was used to record sounds (for ASR), as well as for dealing with hardware buttons. The mobile application uses Windows Forms and C# as programming language.

4.2.2 Desktop

The Desktop device consisted on an HP Touchsmart 600 PC, with the following specifications:

- **Operating system:** Microsoft Windows 7 Home Premium;
- **Processor:** Intel Core 2 Duo;
- **Monitor:** 23'; 1080p Full HD widescreen; with multi-touch technology;
- **Peripherals:** Wireless keyboard and mouse.

The main reason that lead to its choice was its large screen and of course its multi-touch capabilities. In order to use (single) touch, the device mapped automatically touch hits as mouse clicks, and so touch technology was available by default. For speech, a microphone and speakers are incorporated in the device.

Office Communicator 2007 R2 [59] was used to manage audio calls and video calls (in background) between the application and other contacts (making conference possible). *Communicator* was also used to provide server-side speech support (ASR and TTS). A continuous audio call running in background was used in order to link desktop with speech server (more details will be presented in the *Implementation details* section). *Automation API* (*Microsoft Office Communicator 2007 SDK* [60]), was used to start/stop *Office Communicator* conference calls.

In order to do audio streaming on the desktop, using UCMA on server-side, there are two options: (1) using *Office Communicator* as a mediator - this was the solution adopted, but it has some limitations, as *Communicator* must be running in background and it does not offer low level access to data; (2) using UCCA (*Microsoft Unified Communications Client API SDK* [61]) - this would be probably the best and the most dynamic solution, but UCCA at the time was still a low level API and developing a solution using it would be time expensive.

The desktop application was developed in C# and using the WPF framework.

4.2.3 PLA Server

As referred above, *PLA Server* is responsible for mediating communications between home devices and remote servers, being it a server to the devices.

PLA Server runs *Microsoft Windows Server 2008*, with *Internet Information Services 7*, providing *ASP.NET* environment for web services hosting. With this set-up, web services were used to provide data exchange capabilities between *PLA Server* and devices. For speech support, *UCMA Speech 2.0* [62] was used to provide ASR and TTS, both to smartphone and desktop, using the TTS/ASR pt-PT (European Portuguese) engines. For ASR, two pt-PT engines were used: one for dictation (an experimental engine) and another for command and control (voice commands recognition), which is in production in some Microsoft products like Exchange. To provide audio data transfer between the desktop and the PLA Server, *UCMA*

Core 2.0 [63] was used, which enables the establishment of audio calls between this server and *Office Communicator* on the desktop. Audio data between smartphone and server, was dealt recurring to *wav* files, which were transferred through web services.

Speech processing is done server-side, not only because it was a requirement (it´s the only available Microsoft pt-PT speech technology in production, at the time of writing of this thesis), but also because it has some advantages, such as: (1) performance (server has more memory and processing resources), (2) software reuse and device compatibility (the same code works for all devices and operating systems) and (3) configuration-free (clients do not need to install or configure TTS/ASR engines).

PLA Server also was responsible for communication with the Exchange Server. In order to do that, we have used the *Exchange Web Service Managed API 1.0* [64], which provides a complete API for agenda and email management (and more) on a remote Exchange server.

4.2.4 Backend

On backend we installed *Microsoft Office Communications Server 2007 R2*, providing support for audio and video calls and *Microsoft Exchange Server 2008,* to provide email and agenda facilities.

4.3 Prototype interfaces and functionalities

Taking into account Table 31, which describes the user requirements based on the user study and requirements analysis stage, use cases describing each service component (agenda, email and conference), were designed and will be presented in this section. We will also present a general description of how these components are working on mobile and desktop versions of the prototype, describing what the user can do, for each component.

4.3.1 General description

On all user interfaces, being them on mobile or desktop, three modalities are always available (with some exceptions): **speech**, **touch** and **hardware** (including keyboard, mouse and smartphone's buttons). Therefore, whenever a screen appears, the user can execute interactive actions, which include icon or button selection, text box focusing or text insertion, choosing the modality he/she wants. Giving this, we can classify the prototype interface as concurrent (see Table 1).

Below (see Table 32) we present a table describing all possible actions with the prototype, for each modality.

Action	Modality action		
	Speech	**Touch**	**Hardware**
Select a button or icon	Say its name or its meaning (e.g. buttons with an arrow normally mean something like “next”)	Touch the button or icon	Use mouse or touchpad to select it
Available on	Mobile (using Push To Talk - PTT) and desktop	Mobile and desktop	Desktop
Navigate to a text box	Say the name of the label associated with the text box, or the text presented on that text box (whenever a default text is shown)	Touch the label associated with the text box or the text box itself	Use mouse or touchpad to select it
Available on	Mobile (using PTT) and desktop	Mobile and desktop	Desktop
Insert text on a selected text box	If dictation mode is enabled, user only has to dictate the text. If not, user needs first to enable dictation (this button will appear whenever a text box is selected and dictation is possible) and then use it.	On desktop user can use Windows’ virtual keyboard. On mobile user can also use a virtual keyboard, by selecting the proper icon presented in the footer.	User can use the traditional keyboard
Available on	Desktop	Mobile and desktop	Desktop

Table 32 – Prototype’s user manual for common actions

More details about the prototype’s interface can be found on Appendix D.

4.3.2 Agenda

In terms of agenda management, user should be capable to do simple tasks such as: check for appointments, edit or cancel an appointment and of course create a new one (see Figure 4.3.1).

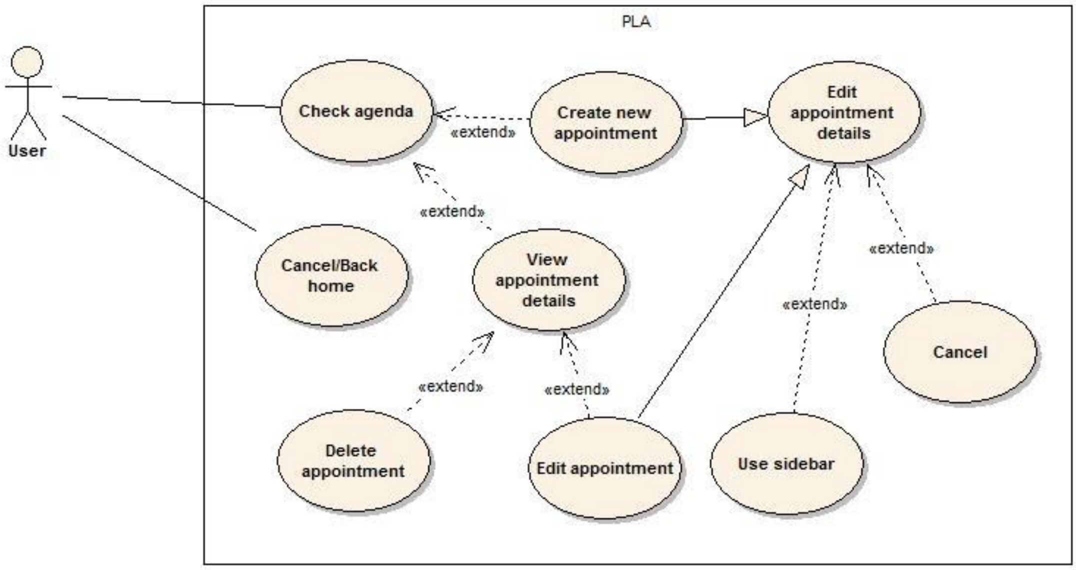


Figure 4.3.1 – Agenda use case

The Agenda interface use should be natural and simple, being its interface in a calendar form, for the month view and, on a list format for appointments of each day (as subjects demonstrated some difficulties on select hourly slots – see Chapter 3). As we can see below (see Figure 4.3.2), the main agenda interface has large buttons representing each day. In order to see more details, the user must say the number of the day or select (touch or mouse) the day, after which a list of all appointments for that day will be presented (see Figure 4.3.3).

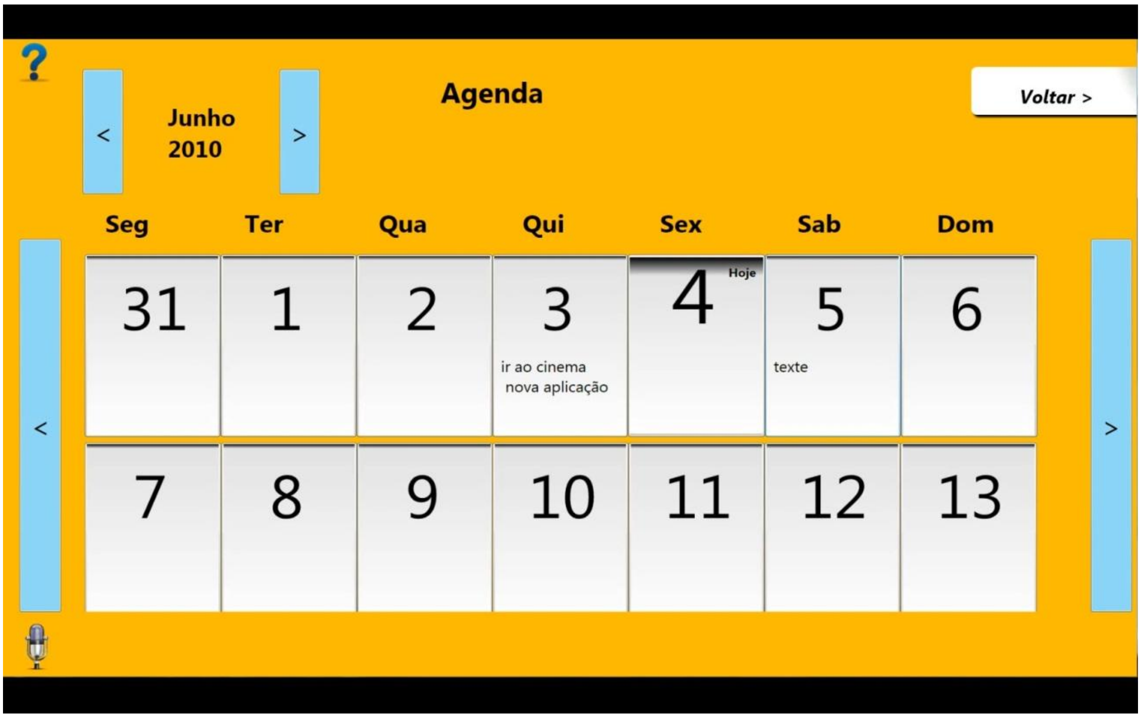


Figure 4.3.2 – Agenda main screen on desktop

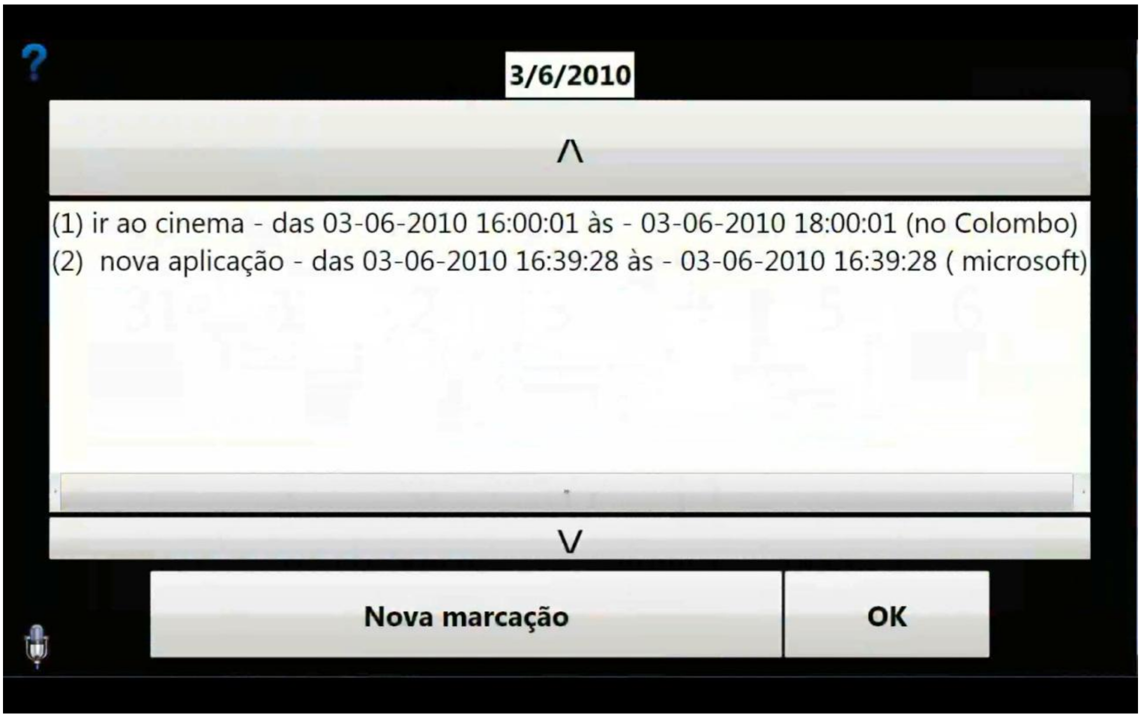


Figure 4.3.3 – Selected day screen (agenda interface) on desktop

The Agenda interface is similar on mobile, having the same features as on desktop. More details about agenda and other interfaces can be found on Appendix D.

4.3.3 Email

Regarding email, the user should be able to do the basics of email management, which is: read email messages, reply and forward email messages, attach documents and send any email message to any contact. Below (see Figure 4.3.4), we presented the email use case and email main screen (inbox) on mobile (see Figure 4.3.5).

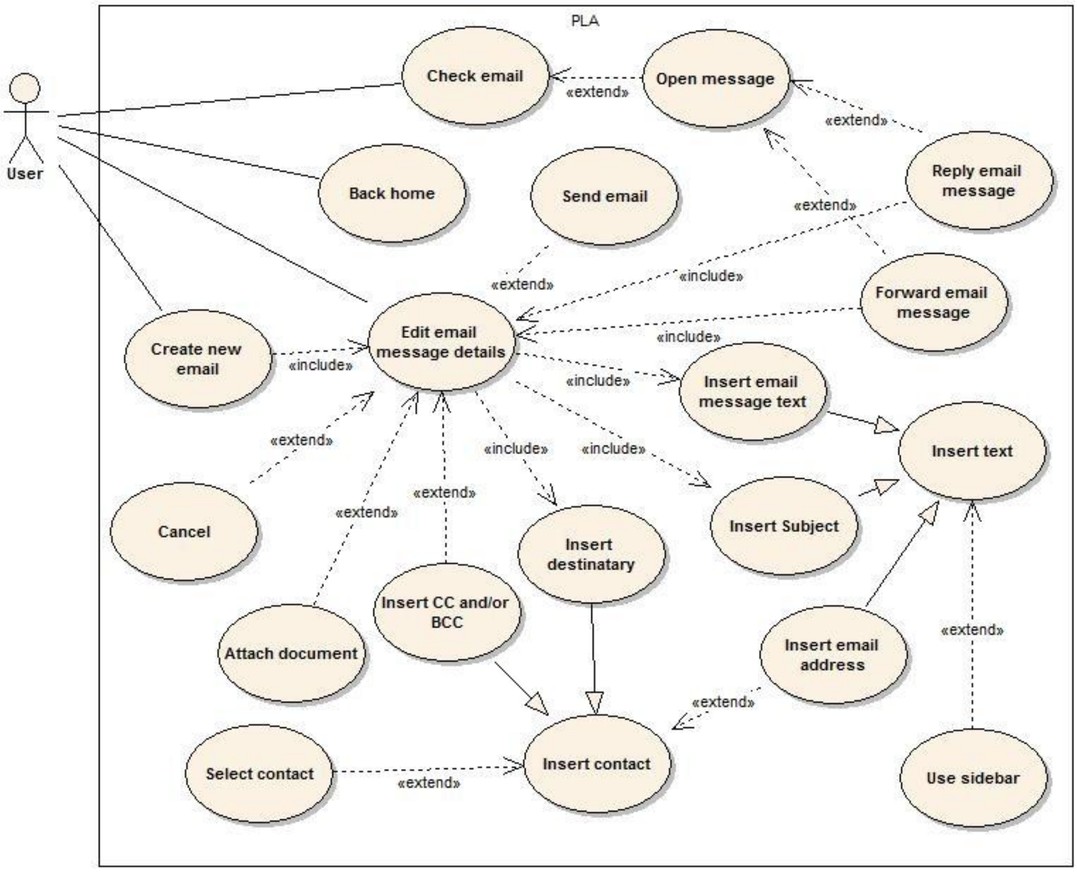


Figure 4.3.4 – Email use case

Figure 4.3.5 – Email interface (main page) on mobile

On mobile, 3D gestures were used in order to perform items selection over lists. In the list presented in Figure 4.3.5, user can tilt the smartphone backwards or forward, in order to select the next or previous item, respectively.

4.3.4 Conference

In terms of the conference service, its way of operation is quite simple: it should allow the user to start an audio or a video call with any contact and to stop a started call, as we can see below (see Figure 4.3.6).

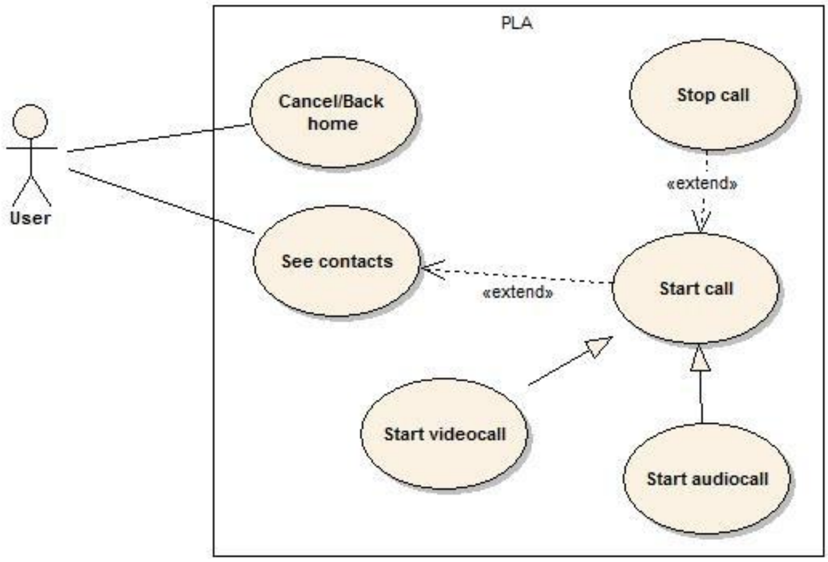


Figure 4.3.6 – Conference use case

Unlike agenda and email, the conference interface on mobile works as an extender (remote control) of the conference main interface, which is on desktop. Below (see Figure 4.3.7), we can see the conference interface on desktop, in which user can select (only by touch or mouse), a contact and initiate an audio or video call with it. Using mobile (see Figure 4.3.8), user can also select a contact (whose action also selects the same, on desktop) and initiate or stop a call (issuing orders to the desktop to start or stop a call).

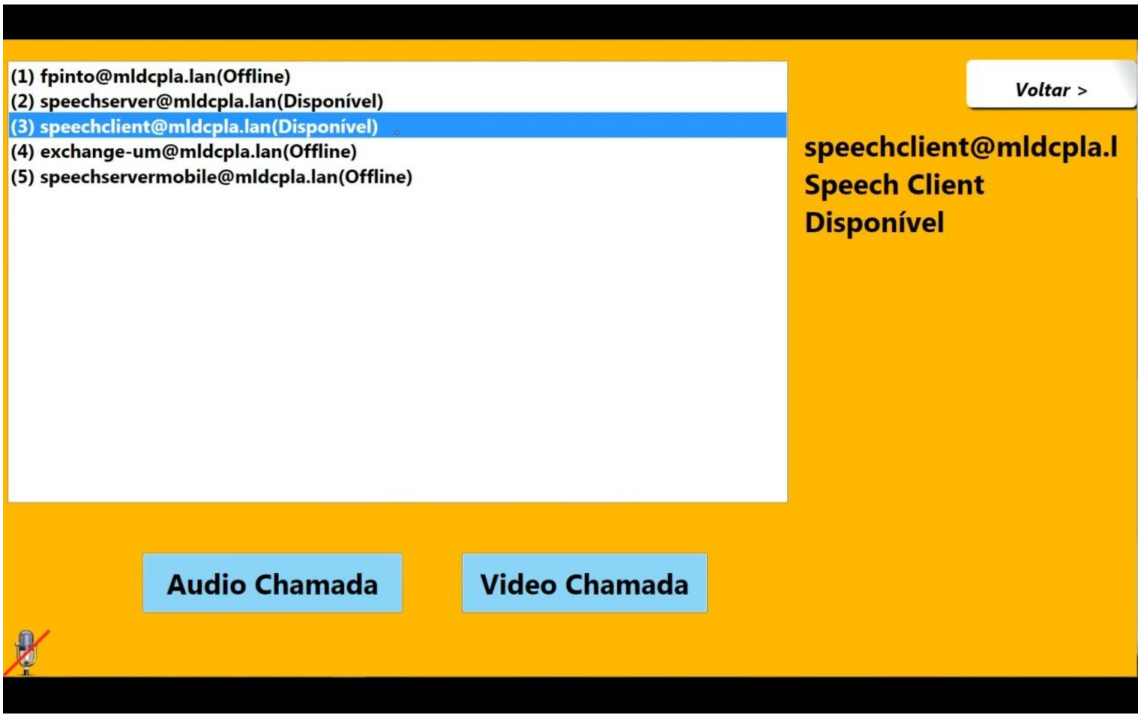


Figure 4.3.7 – Conference interface on desktop

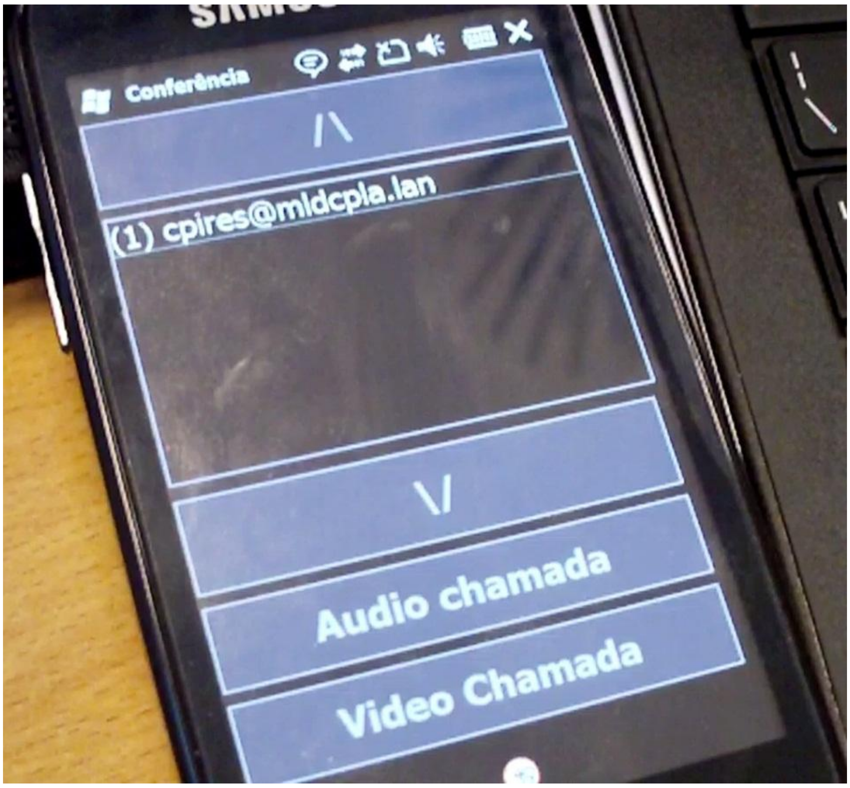


Figure 4.3.8 – Conference interface on mobile

4.3.5 Media Center

Although a media center component was planned to be incorporated in the prototype, and taking into account the amount of time available in the project, this task was not completed.

The planned work was to create a plug-in for the Windows Media Center, that is, create a background application that would run whenever Media Center started. This plug-in would be responsible to handle the speech HCI modality.

In fact there is already an application done in the context of another thesis (see [7]), that is capable of managing ASR and TTS in pt-PT, implementing a speech interface to the Media Center. However, speech engines on that application run locally and so they could not be adapted in a straightforward way, to our architecture (where speech engines run server-side). Even so, considering that media center would run only on the desktop, the integration of this component in our overall architecture would be possible. Unfortunately this application was compiled for Windows Vista and not Windows 7, and all configurations needed for porting it, would take too much time and therefore we have decided that the Media Center integration is out of scope of this thesis.

4.4 Implementation details

In the previous section we have presented how the prototype works in terms of its main features, that is, what the user can do. In this section we will describe, with some detail, how the prototype works in the background. We will see also its architecture in more detail.

Below (see Figure 4.4.1), we will present the prototype's deployment architecture. As we can see, there are three main components: mobile (smartphone), desktop and *PLA Server*. Details about each one, will be presented on the following sub-sections.

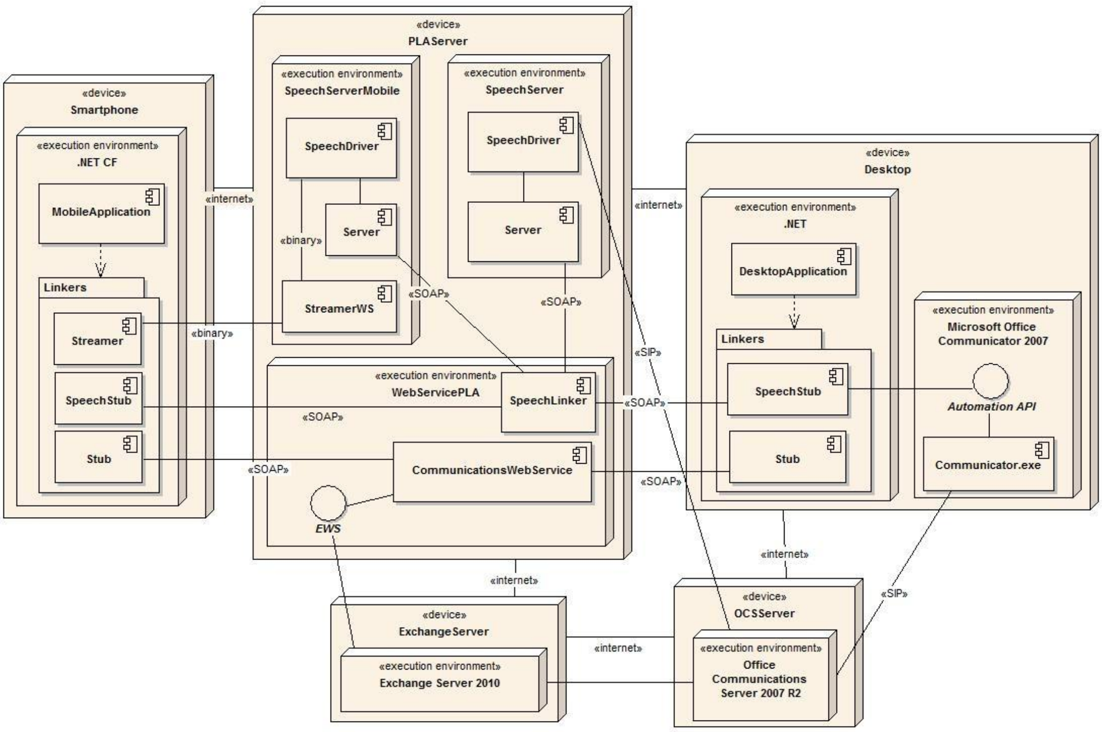


Figure 4.4.1 – PLA Prototype's deployment architecture

4.4.1 Mobile

The prototype's mobile version (smartphone) consists on a piece of software, encompassing graphical and logical parts (*MobileApplication* and *Linkers*). This application is linked with the PLA Server through three main links:

- **Speech link (*Streamer - StreamerWS*)**
  Mobile application records voice samples to *wav* files, and sends those files' raw data to *SpeechServerMobile* through a special web service (*StreamerWS*). For audio playback, the process is the reverse, that is, mobile application playbacks *wav* files received from the web service. In order to record and send a sample to the server, the solution adopted was PTT (Push-To-Talk), meaning that recording starts, when the user presses the smartphone's middle button and stops, when the user presses again the same button. After stopping the sample is automatically send to the server. PTT was adopted because there are currently no available solutions to detect the end of phrase on Windows Mobile, and the development of a solution would be very costly.
- **Speech events link (*SpeechStub - SpeechLinker*)**
  This link uses the main web service (*SpeechLinker*'s web service), and its purpose is to send and receive speech events. For example, whenever the application wants to synthesize (TTS) some text, it calls a function on that web service with the text to be synthesized as argument. After some time (milliseconds), the application receives the *wav* file containing the text synthesized (from the *SpeechServerMobile*) through **Speech link**, which is then played. For ASR the process is done reversely, that is, the recognized text is received from *WebservicePLA* whenever the ASR engine (on *SpeechServerMobile*) recognizes text from sent samples through **Speech link**.

  For notification events, all clients (including desktop) use a pooling thread that generates an event when some information is available. Being this link so useful to send/receive events, it is also used to control Conference state, sending and receiving events from when Conference interface on desktop changes (see *Prototype interfaces and functionalities* sub-section).
- **Application link (*Stub - CommunicationsWebService*)**
  This is the main link, responsible of sending and receiving email and agenda data. All email messages and appointments are managed using this link, that is, by using *CommunicationsWebService* API (a web service) it is possible to interact with Exchange server.

4.4.2 Desktop

Similarly to mobile version, desktop consists on an application with graphical and logical components, and it has too, three links:

- **Speech link (*SpeechStub - SpeechDriver*)**
  This link represents the logical audio streaming between desktop and *SpeechServer.* Audio is then sent and received using *Office Communicator* [59] on desktop, and *UCMA Core 2.0* [63] on *SpeechServer*. *SpeechStub* is responsible for managing that audio call and uses prototype's SDK, which in turn uses *Automation API* [60]. This API interacts directly with *Communicator*, which in turn is connected with *OCS*. *SpeechServer*'s *SpeechDriver* is responsible for connecting to *OCS* and accepting any audio call.
- **Speech events link (*SpeechStub - SpeechLinker*)**
  Like on the mobile version, this link is used to send text to be synthesized (TTS) and to receive recognized text (ASR), to and from the server, respectively. *SpeechServer* which is responsible of TTS and ASR functionalities, interacts with desktop using *SpeechLinker*.
- **Application link (*Stub - CommunicationsWebService*)**
  This link provides an API to access Exchange items (email and agenda), as seen previously on mobile.

4.4.3 Speechservers

Both *SpeechServer* (which handles speech on desktop) and *SpeechServerMobile* (which handles speech on mobile), are very similar. Both servers define grammars for each application state, that is, for each window on desktop or mobile, there are a list of keywords on server, mapping possible voice actions on that window. Whenever a client (desktop or mobile) changes its state, it notifies its corresponding speech server of that change (using *WebservicePLA*). After that, speech server loads the corresponding grammar and starts ASR with that grammar. As we have seen above, whenever some word/phrase is recognized, it notifies the corresponding client sending the recognized text to *SpeechLinker* (*WebservicePLA*).

As we have seen previously, *SpeechServer* and *SpeechServerMobile* differ on the way audio is received/sent. *SpeechServer* handles audio using *UCMA Core 2.0* ([62] and [63]), which enables it to accept audio calls from clients and redirect audio data to ASR and TTS engines. On other hand, *SpeechServerMobile* relies on a special web service (*StreamerWS*) that handles binary audio files (*wav* files). After receiving a new audio file, *SpeechServerMobile* redirects that file to ASR engine, which opens it and tries to recognize some text. The synthesis process is similar: TTS engine saves its output to a *wav* file, which is then sent to client through *StreamerWS*.

4.4.4 WebservicePLA

WebservicePLA is probably the most important component of the *PLA Server*. It provides basic communication with all other components.

We have used web services because it is an excellent way to connect various devices, being the integration independent from operating systems, devices and programming languages. Another advantage is that only one API was done, independently the number of different devices that will use it, and so both mobile and desktop use this service to access common resources as for example Exchange server. *WebservicePLA* is then divided in two parts:

- ***CommunicationsWebService***

  This web service exposes an API that lets direct interaction with agenda and email on Exchange server. The API belongs to prototype's SDK and uses the *EWS Managed API 1.0* [64], being adapted to the context of this prototype which is using Exchange's email and agenda facilities.

- ***SpeechLinker***

  We can consider *SpeechLinker* as the "heart" of the prototype, as all event notifications are managed by this component. All components (mobile, desktop, *SpeechServer* and *SpeechServerMobile*) listen and send information from and to *SpeechLinker*.

  Below (see Figure 4.4.2) we have the *SpeechLinker*'s logical architecture, and as we can see its architecture is based on *W3C Multimodal Interaction Framework* [65], having:

  - a **fusion engine** for modalities combination;
  - an **interaction manager** responsible of generating responses based on commands from the fusion engine;
  - a **generation component** responsible of generating proper responses to the application, after receiving data from the interaction manager;
  - a **session controller** that registers application's current state (screen where the user is at, is dictation mode enabled or disabled, etc.);
  - a **Web Service** that serves as *SpeechLinker*'s input and output mode;
  - a **grammar interpretation component** that is used by interaction manager in order to map recognized speech commands to corresponding events.

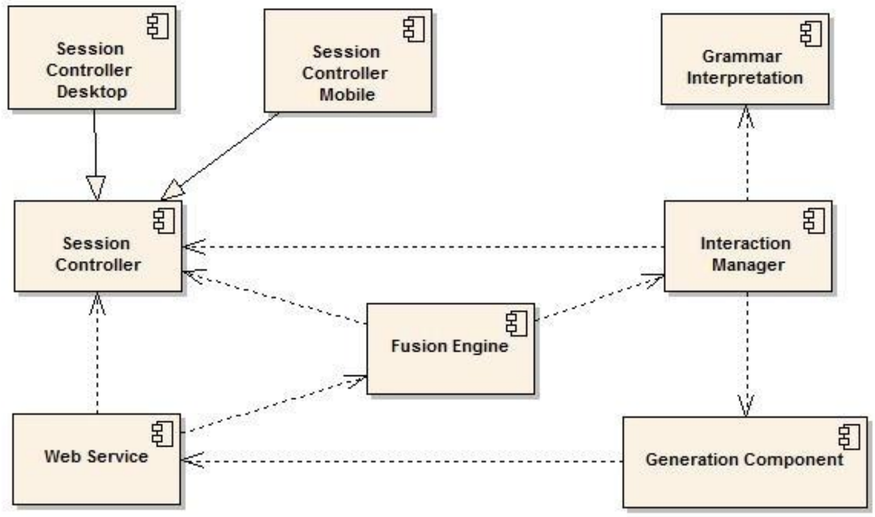


Figure 4.4.2 – *SpeechLinker*'s logical architecture

PLA Server has a multimodal architecture capable of dealing with various input modalities and output devices. Taking into account time available for this thesis development and despite the design and availability of a multimodality architecture, as well as an *EMMA* (*Extensible MultiModal Annotation* [66]) engine, only speech is managed on the server and therefore, as we will see below (see Figure 4.4.3), multimodal actions are not being done (e.g. modalities fusion) by our prototype.

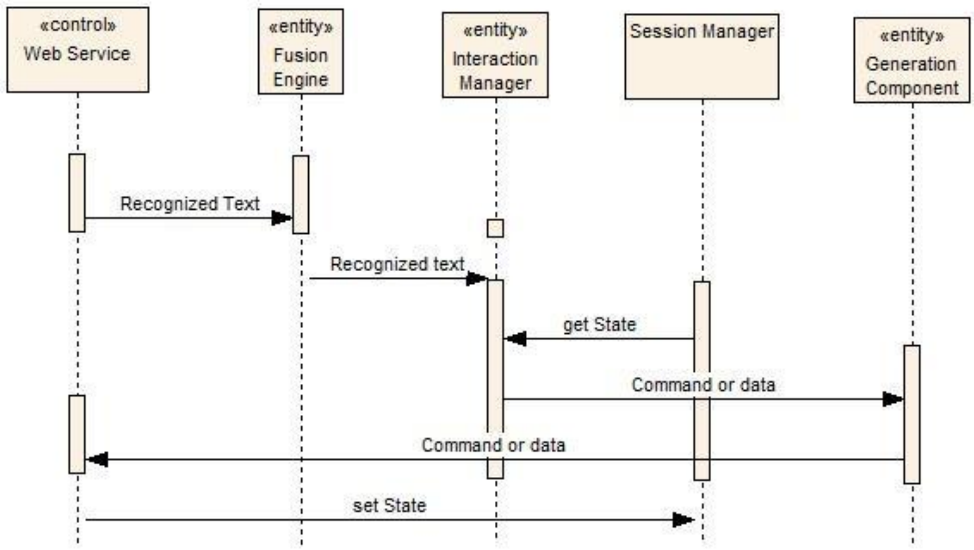


Figure 4.4.3 – *SpeechLinker*'s logical architecture

In Figure 4.4.3, we have an interaction diagram, showing the process of activating a speech generated event. As seen previously, speech servers send recognized texts to *SpeechLinker*'s web service. This component redirects the recognized text to the fusion engine,

which in turn redirects it to the interaction manager. The Interaction Manager takes into account the current state, by checking the session manager (in order to know which commands have to be translated), and then generates a single word which is the name of the method to be called on client. In other cases (when there is no match), recognized text is directly sent to generation component (e.g. dictated text). Then this word or text is sent to generation component which redirects it to web service to be consulted by client. Whenever occurs some change on client, as for example user changes to another screen, web service is notified ("set State") of it by the client itself, which in turn updates the session manager.

Sending the name of the method available on the client, makes it possible to dynamically invoke that method, that is, the client uses reflection in order to call a method that has the name of the received string. If there is no method, then is treated like a text, for dictation mode or variable commands as numbers.

4.4.5 Prototype SDK

The prototype SDK consists of a group of libraries that abstract the access to some resources. As referred in this sub-section, prototype SDK encompasses the following features:

- **EmailLib** - Provides access to Exchange email account, enabling email messages reading, creation, forwarding, replying and deletion, including attachments management. Provides also reading access to Exchange contacts. This library uses *EWS Managed API 1.0* [64] and is used on *CommunicationsWebService* (*WebservicePLA*).
- **AgendaLib** - Provides access to Exchange agenda account, letting appointments creation, edition, deletion and reading. This library uses *EWS Managed API 1.0* [64] and is used on *CommunicationsWebService* (*WebservicePLA*).
- **ConferenceLib** - Provides access to *Office Communicator 2007* that must be running on background. This library enables simple actions on *Communicator* as: login/logout, start/stop audio and video calls, contacts listing and contacts' status change notification. This library uses *Microsoft Office Communicator 2007 SDK* [60] and is used on *desktop*.
- **PLA Server Engine –** This component also known as *SpeechLinker*, provides functionalities for speech events handling and also multimodal management (through its architecture and *EMMA* [66] support).

4.5 Conclusions and future work

As we have seen, the PLA prototype consists not only of a software architecture running on a single device, but of software modules and on a group of devices including a desktop computer, a smartphone and a server architecture (which more than one server). Various technologies were used in order to make the prototype supporting speech and touch on both

platforms (mobile and desktop), as new HCI modalities in addition to standard ones such as keyboard and mouse.

From our analysis of the specification and development, there is still room for prototype improvements, essentially from the technical point of view. Below we present some proposals for future work, regarding our prototype:

- **Audio streaming and dictation on mobile** – currently, speech on mobile only works by using "push-to-talk" (PTT), and this feature somewhat restricts speech on smartphone, not only making dictation difficult, as well as imposing actions on to user to start and stop listening mode;
- **Use UCCA (*Unified Client Communications API SDK* [61]*)* to establish calls on desktop** - *Automation API* [60] is used to establish a background audio call to enable speech on desktop, as well as to establish audio or video calls with contacts (conference). For this, *Office Communicator* [59] must be running on the background, and all functionalities rely on using *Communicator*'s interfaces. This solution is very limiting and using UCCA, would allow low level access to data, making the prototype more independent and dynamic;
- **Multimodal integration –** In order to implement a synergistic (see Table 1) interface, not only speech need to be handled on server, but other HCI modalities too. Adapting *PLA Server* to support more modalities and modality fusion would be a great improvement on the prototype. This will include *EMMA* integration, meaning that all speech events (and events from other modalities as well), would be translated to the *EMMA* language and then fusion engine would interpret them, possibly using a grammar (see Chapter 2 and [41]);
- **Contacts management –** User should be able to manage his/her contacts;
- **Email accounts integration –** The email interface could handle more email accounts rather than only one. This integration could be done on the *PLA Server*, offering it to all clients (mobile and desktop);
- **Instant messaging service –** As we saw in Chapter 3, instant messaging is considered to be very important especially to quadriplegic users, as another way of real-time communication. It would be important for the *Personal Life Assistant* to have an *IM* interface, with speech and touch support, available on both desktop and mobility.

4.6 Summary

In this chapter we have presented some details about the developed prototype, including HCI features, architectural issues and software and hardware development environment, with the aid of use case diagrams. This prototype, also called *Personal Life Assistant*, has the objective of testing if the users requirements presented in the end of Chapter 3, have been correctly mapped or not (as we will see in next chapter, Prototype evaluation). Finally, some topics were presented regarding future prototype improvements, mainly from the technical point of view.

Chapter 5

Prototype evaluation

In this chapter we present the results and conclusions of a final user study aimed at evaluating the prototype PLA application. Five participants, recruited from the initial set, were asked to perform a set of tasks both on the mobile and desktop versions of the prototype. Auxiliary information about this user study can be found in Appendix C.

5.1 User study participants

For this evaluation study we were able to recruit five of the participants that took part in the previous requirements analysis study, three paraplegics and two quadriplegics (see Table 33). For calibrating the study tasks and for comparing results, a non-impaired user, called Control subject, performed the structured tasks (see *Tasks* section).

Participant	Gender	Age	Career	Impairment type
Control	Female	25	Assistant Manager	None
Subject 7	Male	26	Informatics Technician	Paraplegia
Subject 10	Male	19	Student	Paraplegia
Subject 8	Female	54	Technical Assistant	Paraplegia
Subject 5	Male	28	General Manager	Quadriplegia
Subject 9	Male	41	Informatics Engineer	Quadriplegia

Table 33 - Subjects panel for prototype evaluation session

5.2 Methodology

Following what was done on the requirements analysis session (see chapter 3), the prototype evaluation session consisted on asking participants to do some simple tasks, which are described in section 5.3. Tasks were performed without a pre-defined order, to minimize possible task sequence bias in the interpretation of the results across participants. Upon completion of each task, participants answered the set of questions provided in Table 34.

1. Do you like the interface?
2. What could be improved?
3. Do you felt difficulties? Which ones?
4. Do you like more this interface or the one you usually use? Why?
5. Would you use this interface in your daily life?
6. Why did you use more often modality X? (If applicable)

Table 34 - Questions asked in the end of each task

The evaluation study took place on a controlled environment and all participants worked with same hardware and software (see Chapter 3 for details). The study was performed in individual sessions with each participant. At the end of their session, each participant was asked to respond to a small questionnaire (see Table 35).

1. Please rate in terms of easiness/difficulty the following modalities (according to scale A):
   a. Touch (desktop)
   b. Speech (desktop)
   c. Speech (smartphone)
   d. Touch (smartphone)
   e. 3D gestures (smartphone)
2. Please rate in terms of satisfaction the following modalities (according to scale B):
   a. Touch (desktop)
   b. Speech (desktop)
   c. Speech (smartphone)
   d. Touch (smartphone)
   e. 3D gestures (smartphone)
3. Do you think that this prototype could improve your daily life?
4. Do you found prototype’s interface easy to use and intuitive?
5. Which prototype’s version do you liked more? Why?
   a. Desktop version
   b. Mobile version

6. What would you like to see in the assistant beyond email, agenda and conference? **Scale A:** 1. Impossible 2. Very difficult 3. Difficult 4. Medium 5. Easy 6. Very easy **Scale B:** 1. I did not like it 2. I liked it a little 3. I liked it 4. I liked a lot 5. I loved it

Table 35 - Prototype evaluation questionnaire

5.3 Tasks

In order to evaluate the prototype, tasks regarding email, agenda and conference were designed, for both mobile and desktop. Below, the description of each task is presented.

Email task (desktop)	1. Open your email box 2. Open any email message in your inbox 3. Create a new email with the following: a. Subject: *Email de teste* b. Text: *Olá, este é um email de teste!* *Bem responde-me.* *PS: Será que escrever o símbolo do euro é complicado? Deixa cá ver: €* c. Attach an image d. Recipients: i. To: apmultimodal@gmail.com and *Fernando Pinto* (existent contact) ii. Cc: *Eu Próprio* (existent contact)
Agenda task (desktop)	1. Open your agenda 2. Create a new event for tomorrow, with the following: a. Subject: *Ir ao cinema* b. Description: *Ira o cinema ver um filme*

	c. Location: *Colombo* d. Start hour: *16h00* e. Duration: *2 (h)* 3. Delete the previous event
Conference task (desktop and mobile)	1. Start conference on mobile and then on desktop 2. Start an audiocall with contact *3* using mobile. 3. Stop the current audiocall using desktop. 4. Start a videocall with contact *3* using desktop. 5. Stop the current videocall using mobile.
Mobile task	1. Open your email box 2. Select a message (you can use the accelerometer) 3. Send an email message to someone in your contact list and with subject and content at your choice 4. Open your agenda 5. Create a new appointment for any day with details at your choice

Table 36 - Prototype evaluation tasks description

As we can see, the first ones (email and agenda), are structured tasks, from which quantitative results such as execution times can be compared. Mobile task relies on giving the participants the possibility of freely trying the prototype mobile version and after that collect their opinions. Conference task can be considered as a hybrid task, because its purpose is to test if there are advantages or not on using mobile devices as extenders, that is, remote controllers.

5.4 Analysis methods

Qualitative results were extracted from all tasks. Quantitative results were extracted only from structured tasks (email and agenda).
As qualitative results we considered the following:

- **Result**, that could be:
  - **Successful completion** - participant successful completed the task;
  - **Incomplete** - participant did not performed all tasks successfully;
- **Observations** - our point of view of participants' performance on doing the task;
- **Participants' opinion** - opinions given by participants about the task in reply to questions referred on Table 34.

For quantitative results we considered (available only for email and agenda task on desktop):

- **Time to complete a task** - time (in minutes) since the participant was instructed to do a task until task termination
- **Number of helps** - number of times the participant asked for help or was helped
- **Modality count** - number of times a modality was used to accomplish a single action (select a text box or a button). A modality was counted only when all three modalities were available:
    - **Speech** – the participant could use command and control to select a text box or a button or dictation to write a text
    - **Touch** – the participant could use touch to select a text box or a button or virtual keyboard to write a text
    - **Hardware** – the participant could use traditional hardware input devices as mouse/touchpad or keyboard

  Note that all three modalities are mutually exclusive and counted as well. In case of failure of one modality, only the first chosen modality was counted.

5.5 Results

In this section, we present all the tasks results (grouped by each one) and the questionnaire results as well.

5.5.1 Email task

Participant	Time to complete	Number of Helps	Modality count			Result
			Speech	Touch	Hardware	
Control	04:07	2	**14**	1	0	Successful completion
Subject 7	04:10	2	6	**14**	1	Successful completion
Subject 10	04:55	3	4	**9**	2	Successful completion
Subject 8	05:24	2	0	**13**	3	Successful completion
Subject 5	02:37	1	**11**	0	0	**Incomplete**
Subject 9	05:35	0	2	**18**	0	Successful completion

Table 37 - Email task results

Participant	Observations	Participant's opinion
Subject 7	Participant had difficulties on finding "insert new email" option Some dictation problems were registered, and subject 7 had to switch to keyboard	Participant said that the interface is intuitive and he liked it Subject 7 considered that he did not have any difficulties and speech is more appropriate for

		home usage
Subject 10	Participant did not have any problems on using interface, excepting on using dictation mode (bad recognition)	Participant considered the interface simple and he liked it Participant referred that he did not like dictation mode as it had too many failures
Subject 8	Participant had some difficulties on inserting a new email and deselecting a selected contact from the "To" list Participant had some difficulties on selecting the "Open" button from the browsing window (attachment), because it was too small	Participant liked the interface and she said that she would not change anything She used touch modality more often, because she considered to be easier
Subject 5	Participant did not accomplished the entire task: he did not attached a document (no voice support for browsing window), and he did not wrote the entire message text Participant only was capable of using speech and had some difficulties with dictation	Participant considered the interface easy to understand and intuitive Although, he uses Outlook with gaze-based interface device, from which he is accustomed and that's why he would not change for this interface Participant considered also that interface has a small feature set, but there are functionalities that he would like to have on Outlook
Subject 9	Participant tried to use dictation but due to successive failures, he used virtual keyboard instead, which was used to write all the texts Participant had some difficulties selecting "Open" button from browse window (insert attachment) Participant used his knuckles in order to use touch	Participant referred that the symbols sidebar should be integrated with virtual keyboard and not aside Participant said that writing large texts is better using a keyboard than speech, just because the process of formulating a sentence He used more often the touch modality, because he considered it to be less tiring than using a mouse, which he had to use both

		hands Participant considered speech for text box selection, very effective and useful Subject 9 referred that we would use this interface on daily basis

Table 38 - Email task observations and participant's opinion

5.5.2 Agenda task

Participant	Time to complete	Number of Helps	Modality count			Result
			Speech	Touch	Hardware	
Control	03:16	3	**16**	1	0	Successful completion
Subject 7	02:16	1	3	**16**	2	Successful completion
Subject 10	04:02	4	**9**	7	1	Successful completion
Subject 8	04:09	2	0	**11**	3	Successful completion
Subject 5	03:05	2	**15**	0	0	Successful completion
Subject 9	03:30	1	5	**13**	0	Successful completion

Table 39 - Agenda task results

Participant	Observations	Participant's opinion
Subject 7	Participant accomplished the task with no problems	Participant considered the interface easy to understand, intuitive and referred also that he would not change anything Subject 7 said also that he would like to use the interface on a daily basis and did not felt any difficulty
Subject 10	Participant did not have any problems using the interface Participant used dictation as the primary mode of text insertion	Participant considered the interface easy to understand, He referred also that he liked the interface and he would use it on a daily basis
Subject 8	Participant used interface without major problems	Participant considered the interface intuitive and helpful: "For those who know nothing like me, this (interface) is easier" "I would use this interface

		everyday" Participant referred that she used touch more often because of her hoarse voice
Subject 5	Participant used interface with no problems, excepting the appointment's start hour, which was not optimized for speech input and had to be edited manually	Participant said that he would not use this interface because he already had one (gaze-based interface) and he considered speech slower than gaze
Subject 9	Participant accomplished the task with no problems Participant used dictation for every text inputs Participant used his knuckles in order to use touch	Participant liked interface and had no difficulties on using it Participant referred that he used dictation just for trying, and also in that situation (write small texts) it is better and simpler to use dictation than keyboard

Table 40 - Agenda task observations and participant's opinion

5.5.3 Conference task

Participant	Observations	Participant's opinion	Result
Subject 7	Participant accomplished the task without problems	Participant found interesting the fact that smartphone operated like a remote control, although he did not found that functionality very useful Subject 7 considered interface easy to interact with	Successful completion
Subject 10	No problems were registered	Participant referred that he did not have any problems "I would certainly use this interface" Subject 10 found interesting that smartphone was used as a remote control	Successful completion
Subject 8	Participant did not have any problems using the interface	Participant liked the interface and would use it everyday	Successful completion

Subject 5	Participant accomplished the task with no problems Participant only used touch on mobile	Participant considered that smartphone operating like a remote control could bring many advantages and he would use the interface As an improvement, subject 5 referred that videocall image could be bigger	Successful completion
Subject 9	Participant did not have any problems using the interface Participant used his knuckles in order to use touch	Participant referred that the of the smartphone operating as a remote control, could be very useful, when he is far from the computer Participant 9 referred also that he liked the interface and would not change anything Participant considered that must be alternatives to conference, as at least on business environments, we have to consider the privacy issue Participant said that the only difficulty he felt was using touch on mobile (necessary pressure was too excessive)	Successful completion

Table 41 - Conference task results

5.5.4 Mobile task

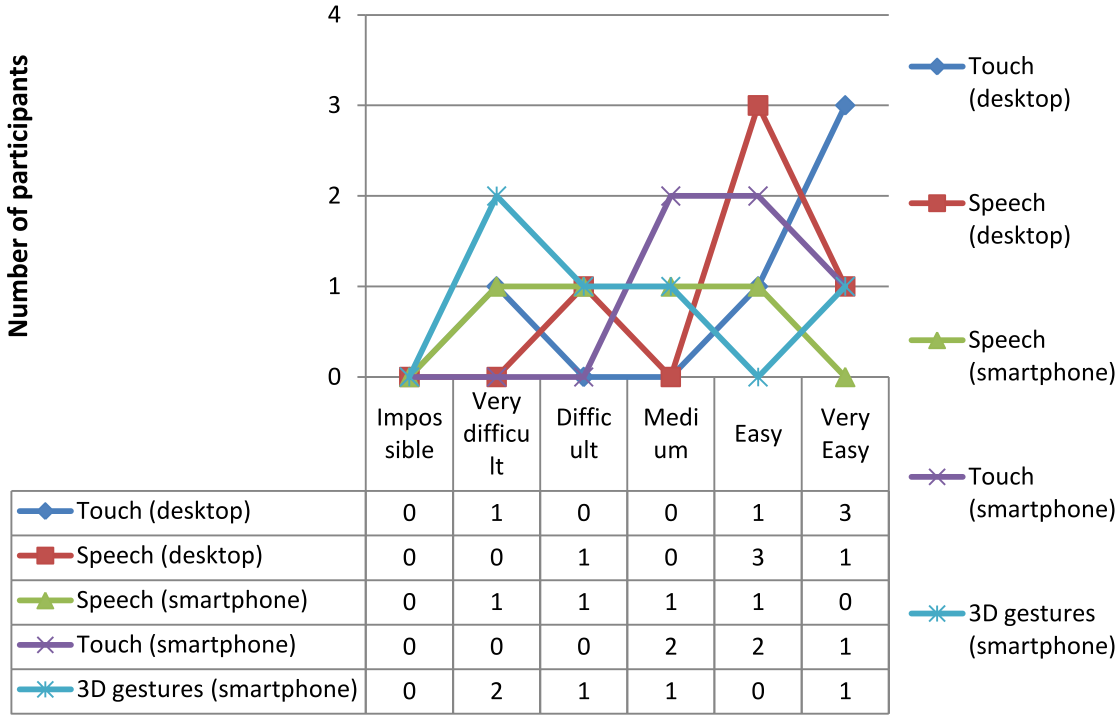
Participant	Observations	Participant's opinion
Subject 7	Participant had some difficulties on finding some options (submenus) Participant managed to use virtual keyboard with no problems, although the option to activate/de-activate was not clear We noticed that the accelerometer was used unintentionally when participant was handling the smartphone and it bothered him	Participant referred that mobile version requires more learning than desktop one Participant considered contacts' selection (on email) interesting Subject 7 said also that he would use the application on a daily basis, although ASR should be continuous and not "push-to-talk" (PTT)

	Participant used PTT activated by button	
Subject 10	Participant managed to use speech with no problems, although after some time of use push-to-talk states were desynchronized Participant had no problems on using virtual keyboard neither accelerometer for items selection Subject 10 had some difficulties on selecting recipients (email) Participant used PTT activated by button	Participant found application very interesting and he would use it on daily basis He said also that push-to-talk solution is useful as it can be used to switch off/on speech on noisy environments or whenever he wants to speak with somebody
Subject 8	Participant did not use speech due to technical difficulties Subject 8 used accelerometer with no major problems	Participant found mobile version interesting
Subject 5	Participant managed to write with virtual keyboard and found it functional Participant had some problems on selecting smaller buttons (footer buttons) and navigating through application menus Both speech and touch were used Smartphone was attached to a table Participant used PTT activated by proximity sensor	Participant considered that speech would be more useful if ASR was continuous (and not PTT) Participant managed to use PTT with no major problems Participant referred also that it was very important for him to have his PC and phone synched (which this application offered)
Subject 9	Participant had difficulties on using touch, because he did not managed to apply the right pressure on smartphone's screen Due to touch limitations, participant had issues on using virtual keyboard to write Participant used PTT activated by proximity sensor	Subject 9 considered accelero-meter functionalities (list items selection or week navigation) bother, as these 3D gestures were activated when subject had adjust the smartphone Subject considered PTT (with proximity sensor) very problematic, as like 3D gesture it was activated unintentionally

Table 42 - Mobile task observations and participant's opinion

5.5.5 Questionnaire results

This section presents the responses to the questionnaire given by users after completion of all tasks (Table 35).

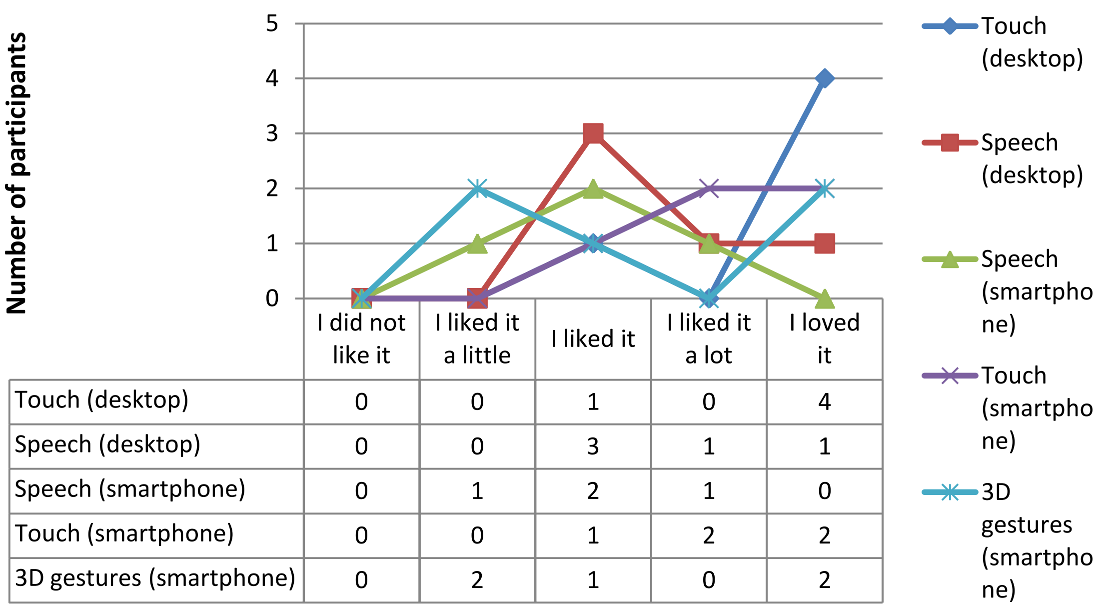


	Impossible	Very difficult	Difficult	Medium	Easy	Very Easy
Touch (desktop)	0	1	0	0	1	3
Speech (desktop)	0	0	1	0	3	1
Speech (smartphone)	0	1	1	1	1	0
Touch (smartphone)	0	0	0	2	2	1
3D gestures (smartphone)	0	2	1	1	0	1

Figure 5.5.1 - Questionnaire (see Table 35) results for question one

	Quadriplegic 1	**Quadriplegic 2**
Touch (desktop)	Very difficult	Very Easy
Speech (desktop)	Easy	Easy
Speech (smartphone)	Difficult	Medium
Touch (smartphone)	Medium	Medium
3D gestures (smartphone)	Very difficult	Difficult

Table 43 - Quadriplegic questionnaire (see Table 35) results for question one



	I did not like it	I liked it a little	I liked it	I liked it a lot	I loved it
Touch (desktop)	0	0	1	0	4
Speech (desktop)	0	0	3	1	1
Speech (smartphone)	0	1	2	1	0
Touch (smartphone)	0	0	1	2	2
3D gestures (smartphone)	0	2	1	0	2

Figure 5.5.2 - Questionnaire (see Table 35) results for question two

	Quadriplegic 1	**Quadriplegic 2**
Touch (desktop)	I liked	I loved it
Speech (desktop)	I liked	I loved it
Speech (smartphone)	I liked	I liked a lot
Touch (smartphone)	I liked	I liked a lot
3D gestures (smartphone)	I liked	I liked a little

Table 44 - Quadriplegic questionnaire (see Table 35) results for question two

Participant	**Question 3.**	**Question 4.**
Subject 7	Interface was considered intuitive, easy to use and useful. The multimodality was found useful, since one can choose which modality to use on each occasion.	The participant response was yes.
Subject 10	The participant considered that with this assistant he could do multitasking, e.g. he could dress-up, while sending an email using speech.	Considered interface easy to use.
Subject 8	"Yes it can"	Considered this interface easier to use than other applications.
Subject 5	The participant considered that this assistant can have an important role regarding management of his daily tasks and synchronization between mobile and PC.	The participant response was yes.

Subject 9	Considered with this prototype it is easier to read and write email messages.	The participant response was yes.

Table 45 - Questionnaire (see Table 35) results for questions three and four

Participant	**Question 5.**	**Question 6.**
Subject 7	Preferred the **desktop** version over the mobile one, because it had more work area and it was easier to use.	Participant said that the prototype was good.
Subject 10	Preferred the **desktop** version because it had bigger buttons and it was easier to use.	Participant suggested that the assistant should be more "assistant", that is, it should offer: news, weather forecasts, movies of the week, radio, etc.
Subject 8	"**Desktop** version, because it is bigger".	Participant said that the prototype was well.
Subject 5	Preferred the **mobile** version, just because he already has an adapted PC and the desktop version would not improve his interaction. Also, he referred that he was not aware of any mobile applications capable of doing what he want them to do.	Participant mentioned that access to music and Office applications would be great.
Subject 9	"**Desktop** version, because it is more *colorful* and touch screen has more sensitivity"	Participant replied that email management (with other email accounts) in one email box would be a good improvement.

Table 46 - Questionnaire (see Table 35) results for questions five and six

5.6 Results Analysis and Discussion

In this section we present the analysis of the quantitative results, discussing each task separately. We also discuss the qualitative results and provide conclusions derived from this study.

5.6.1 Email task

The email task is a structured task, and therefore not only qualitative results were obtained, but also some quantitative results were registered as well, as we have seen on Table 37.

As we have seen on Chapter 3, one of the major limitations of quadriplegics is their writing capability, that is, due to their mobility impairments, writing speed and insertion of some characters are affected. Thus, the email task was designed primarily to evaluate users' writing capabilities with the multimodal prototype.

We can see below (Table 47 and Figure 5.6.1), there are practically no differences between paraplegic and quadriplegic. Quadriplegics that successfully accomplished the email task (only subject 9), took longer only forty-five seconds more than paraplegics (all paraplegics managed to terminate successfully email task). Even the differences between all disabled and control subject are quite low, only twenty-four seconds. This could be an indicator, that in fact using alternative modalities could improve mobility impaired users usability. In fact, almost all participants used alternative modalities (see Table 37).

	Mean	Standard deviation	Mean Differences
Control	**04:37**	**-**	**00:04**
General	**04:32**	**01:12**	
Successful General	**05:01**	**0:37**	**00:24 (with control)**
Quadriplegic	**04:06**	**02:05**	**00:43**
Paraplegic	**04:49**	**00:37**	
Successful Quadriplegic	**05:35**	**-**	**00:45**
Successful Paraplegic	**04:49**	**00:37**	

Table 47 - Email task execution time results (tabular form)

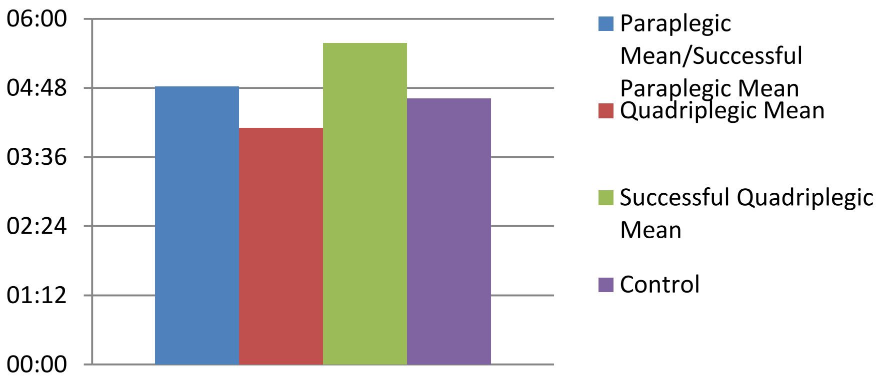


Figure 5.6.1 – Email task execution time results (graphical form)

As we can see on Table 37, touch was the most used modality (except Subject 5 that only used speech due to his advanced condition). Another interesting result is that quadriplegics did not use any traditional hardware interface and paraplegics used more often other modalities such as speech or touch rather than keyboard or mouse. Speech was not used more often, because dictation mode had a medium-low recognition rate.

As the subject 9 mentioned, most quadriplegics have to use both hands to handle a mouse, and so it is faster to use touch or a simple touchpad than a mouse. Subject 9 also managed to replace keyboard with the virtual one. But on some cases, like subject 5, where the impairment level is high, speech was the only alternative as we can see in Table 37.

In terms of easiness of use, we can observe that all participants took practically the same time to accomplish the task (except for subject 5 whose task was aborted due to some limitations). Considering that subject 8 is non-proficient with email applications and that she took practically the same time as others in this task, seems to indicate that the way the interface is designed could improve email usability for the general public due to its simplicity.

5.6.2 Agenda task

Similarly to email task, the agenda task is also structured, and so quantitative results are analyzed here. Regarding the task itself, the agenda task was designed to evaluate not only writing but also cursor handling capabilities. Using a cursor can have some limitations, as we have seen in chapter 3, since some quadriplegics have to use both hands to handle a mouse.

From Table 48 and Figure 5.6.2 the results are even better, when compared with the email task, since the differences between quadriplegics and paraplegics are minimal (only eleven seconds), and quadriplegics even took less than paraplegics.

	Mean	**Standard deviation**	**Mean Differences**
Control	**03:16**	**-**	**00:08**
General	**03:24**	**00:46**	
Quadriplegic	**03:17**	**00:17**	**00:11**
Paraplegic	**03:29**	**01:03**	

Table 48 - Agenda task execution time results (tabular form)

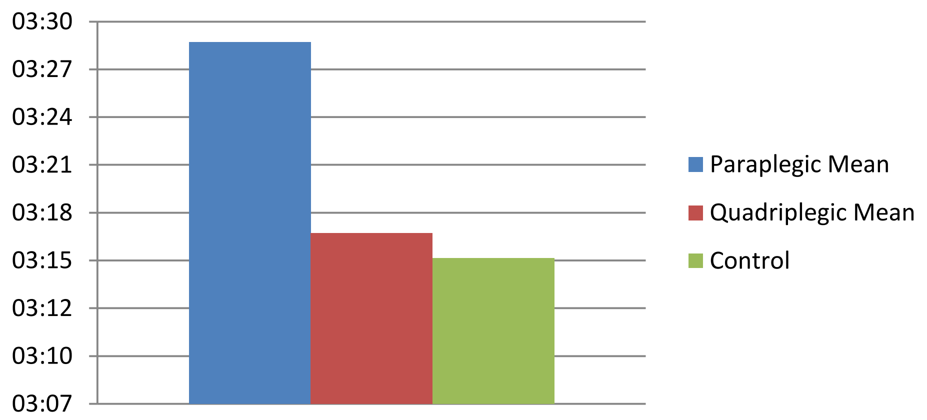


Figure 5.6.2 – Agenda task execution time results (graphical form)

The agenda task results are not surprising, since on the previous requirements analysis study (see Table 14) differences were also small. Although in this case we have to consider that: (1) subject 5 (which has by far the participant with most advanced disability) did not performed any communication and entertainment services tasks in the requirements analysis study, and performed all tasks in this one, and his execution time was similarly to the rest of the panel; (2) subject 8 which is non-proficient, did not take longer than the rest of the panel; (3) independently of the limitation, all participants (including control) took practically the same time to accomplish this task (excepting subject 7 who was the fastest).

Again, participants chose to use speech or touch to accomplish this task, and they practically replaced keyboard and mouse (see Table 39). With this approach, they managed to overcome their limitations using alternative modalities, and we can then conclude that a multimodal agenda interface offers less or inexistent limitations to mobility-impaired users.

5.6.3 Conference task

Conference task can be considered as a hybrid task, since it required the simultaneous use of the prototype's mobile and desktop versions. The objective was simple: participants had to manage audio and video calls using the desktop interface, which was the main interface, and the mobile interface, which worked as a remote controller. Considering that mobility-impaired people rely on their mobile phones to be in contact, and as we have seen on Chapter 3, some quadriplegics had their phone attached to their wheelchair, and mobile devices could have an important role. In cases of advanced limitations, as we have seen with subject 5, speech is the main modality that is used (at least on desktop). During this task no speech support was available (due to technical limitations), and so we hypothesized: if speech is not available on desktop, can a remote controller substitute that fault?

Participants had no problems on using conference interface, and they all considered it very intuitive and easy to use. Regarding the remote controlling functionality, paraplegics found it interesting but not very useful. Quadriplegics on the other hand, considered that the fact of the smartphone can operate as a remote control, could bring many advantages and it could be very useful, namely when they were far from desktop. During evaluation session, we observed that participant 5 (see Table 41) used only smartphone to interact with desktop, in order to initiate/terminate an audio or video call.

We can then conclude that whenever "wireless" modalities, like speech, are not available, mobile touch devices could overcome them. Being smartphones usable by quadriplegics (if guidelines in the end of Chapter 3 were followed), they can operate as remote controllers and then making another modality available.

5.6.4 Mobile task

In order to evaluate mobile version of the prototype, including email and agenda interfaces, an informal task was designed. Participants were told to manage their email and agenda using the smartphone, and again they were free to choose which modality to use (see mobile task on Table 36).

As we can see from Table 42, prototype's mobile version had still some limitations, mainly for quadriplegics. Subject 9 had some difficulties on using smartphone's touch screen, as the necessary pressure to activate the components was too excessive. Subject 5 (also quadriplegic) did not have any problem on using touch on mobile.

Regarding the application itself, all participants had initial difficulties on using mobile interface, but after some time they learned how to use it. They all considered the interface interesting, although participants considered PTT solution not very effective as well the way how 3D gestures were used, since they can be activated unintentionally. As future improvements, it was suggested that:

- ASR should be done continuously, with the option to mute/un-mute recognition (as in desktop);
- whenever 3D gestures are present, there must be an option to switch them off/on;
- dictation mode should be available;
- smartphone should have a capacitive display instead of a resistive one (as the necessary pressure in order to use touch on capacitive displays is less than on resistive displays).

As we have seen, there is still room for improvement, but this mobile application is closer to the objective of giving more mobility to quadriplegics and paraplegics, not only overcoming interaction limitations (as button and icon size) but also making the assistant available everywhere (as all emails and appointments are synchronized with desktop).

5.6.5 Questionnaire analysis

Participants tried different modalities both on mobile and desktop, and were asked about the easiness/difficulty of use of each modality as well as how they rated each one according user satisfaction.

As we can see from Table 49 and Figure 5.5.1, participants considered touch and speech on desktop the easiest ones, followed by touch on smartphone. As we have seen, smartphone's touch screen requires some pressure that some quadriplegics cannot apply. Finally, speech on smartphone was considered to be medium (see Table 49), probably because of the solution adopted (PTT). 3D gestures were considered to be difficult by quadriplegic, as they bother more than were useful.

Regarding classification according user satisfaction, results are quite similar (see Table 50 and Figure 5.5.2). Participants considered touch (both on mobile and desktop) the most preferable modality, followed by speech desktop, 3D gestures and finally speech on mobile. Once again, speech on mobile and 3D gestures received the lower rates. An interesting result is that, paraplegics classified speech on mobile with the lowest rate, as probably they feel that with PTT, speech is not so useful and functional as it should be, while quadriplegics probably found speech on mobile a good feature, at least for the future.

In terms of modalities, we can conclude that touch and speech are the preferable alternative modalities, but on one hand touch screens must be sensible enough and on other hand automatic speech recognition should be done continuously. Accelerometer (3D gestures) does not play an important role on in this kind of interfaces yet, and it can even bring some problems as we have seen with subject 9.

Modality	Mean	Paraplegic Mean	Quadriplegic Mean
Touch (desktop)	5 (**Easy**)	5,(6) (**Very Easy**)	4 (**Medium**)
Speech (desktop)	4,8 (**Easy**)	4,(6) (**Easy**)	5 (**Easy**)
Speech (smartphone)	3,5 (**Medium**)	3,5 (**Medium**)	3,5 (**Medium**)
Touch (smartphone)	4,8 (**Easy**)	5,(3) (**Easy**)	4 (**Medium**)
3D gestures (smartphone)	3,4 (**Difficult**)	4 (**Medium**)	2,5 (**Difficult**)

Table 49 - Question one (see Table 35) overall results

Modality	Mean	Paraplegic Mean	Quadriplegic Mean
Touch (desktop)	4,6 (**I loved it**)	5 (**I loved it**)	4 (**I liked a lot**)
Speech (desktop)	3,6 (**I liked it**)	3,(3) (**I liked it**)	4 (**I liked a lot**)
Speech (smartphone)	3 (**I liked a little**)	2,5 (**I liked it**)	3,5 (**I liked a lot**)
Touch (smartphone)	4,2 (**I liked a lot**)	4,(6) (**I loved it**)	3,5 (**I liked a lot**)
3D gestures (smartphone)	3,4 (**I liked it**)	4 (**I liked a lot**)	2,5 (**I liked it**)

Table 50 - Question two (see Table 35) overall results

All participants considered the prototype's interface intuitive and easy to work with, and they said also that the prototype would improve their daily life (see Table 45), not only by making them more "mobile" but also because of the simplification of the interface itself. Most

participants preferred prototype's desktop version, because they considered its interface more appellative and easy to understand. Although subject 5 (who preferred mobile version instead), pointed out the importance of an assistant present in a mobile device, and how it could improve his life (see Table 46).

As future improvements, some interesting suggestions were made (see Table 46): the assistant could provide daily information (weather forecasts, news, etc); the assistant email could managed various email accounts; and of course all improvements mentioned above.

5.6.6 Conclusions

The results presented in this chapter indicate that in fact the existence of alternative modalities improves disabled users' interaction. We have seen in all tasks that more than one modality was used by all participants, except subject 5 who had severe mobility limitations and chose to use only speech modality. Other participants used and combined all available modalities in order to optimize their experience.

On email and agenda tasks, we observed an interesting point, which was the fact that the hardware interfaces (keyboard and mouse) were almost replaced by touch and speech. This indicates on one hand that for quadriplegic users, speech and touch are less tiring or simply the only way to interact with interfaces. On the other hand, paraplegic found more effective and simpler these new modalities than the traditional ones.

As we have seen on email and agenda task, the time needed by each participant to complete the tasks, were not so different from participant to participant. Giving this, we can say that quadriplegics had the same performance as paraplegics, and so improving their interaction. Also, non-proficient users had the same performance as proficient ones, which mean that the interfaces simplification improved usability as well.

Finally we have seen that with some improvements, mainly to the speech interface, the prototype's mobile version, in which more modalities are available, can play an important role in mobility-impaired people's daily life, as it offers the possibility of giving them more independence.

5.7 Summary

In this chapter we presented the results and conclusions of the prototype evaluation session. We have concluded that in fact the multimodal prototype could improve the mobility-impaired users' interaction with communication services, as study's participants managed to overcome their limitations by using alternative modalities like touch and speech and even replace the traditional ones such as keyboard and mouse.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Mobility-impaired individuals are forced to stay at home more than they would like to. In order to overcome isolation, Information and Communication Technologies (ICT) are already being used by these individuals, to communicate not only with family but also with friends and acquaintances. Therefore, ICTs take an important role in mobility-impaired people's lives, fighting isolation. Services like email, conference and agenda, take an important role in communication and social life management, being essential features to be included on a personal life assistant.

In Chapter 2, we have reviewed examples and applications of systems using alternative modalities, such as speech or touch. We also mentioned restrictions and limitations that those alternative modalities exhibit, when operating alone. As the literature reveals, multimodal applications, by accepting more than one HCI modality, make the systems more accessible as users can choose which modality to use. Considering these observations, a multimodal prototype can potentially improve the usability of communication and entertainment applications by mobility-impaired users, offering alternative ways of interaction to overcome specific physical limitations.

In Chapter 3, we presented two user studies. The first user study consisted on a preliminary controlled interview with a group of mobility-impaired individuals, with the objective of gaining insights into their current use of ICTs in general, and communication and entertainment applications, in particular. This first study enabled us to derive some guidelines for future work. The second study aimed at understanding how current interface hardware limit and affect mobility-impaired users' interaction and how alternative HCI modalities (such as

speech or touch) can improve it. Also, we observed difficulties that participants had on using ICT in terms of email, agenda, media center and audio/video conference applications. The first two goals of this thesis were achieved by these two user studies. From these studies, we concluded that, there are still some limitations that affect interaction with current technologies. The traditional interfaces, such as keyboard and mouse, raise some barriers especially to quadriplegics. We have seen that within the whole group of mobility-impaired study participants, there are significant differences between individuals, for example, some can use keyboard others cannot. Paraplegics managed to use interfaces without many difficulties, much like non-impaired people, while quadriplegics on other hand showed more limitations. Even within the quadriplegic sub-group, some have more limitations than others and therefore, restrictions on using current interfaces go from simply low writing speed, to inability to use some HCI modalities or devices.

Based on results and design guidelines from the first two user studies we derived a set of user requirements for the development of a multimodal and multi-platform system for assisting mobility-impaired users in accessing communication services, such as email, agenda and audio/video conferencing, through a single application. In Chapter 4, we provided a detailed description of this prototype, including interface description and functionalities, as well as technologies used.

In order to accomplish the thesis main goal, which was to test our hypothesis and draw conclusions about the use of multimodal interfaces for communication and entertainment services, a prototype was developed and evaluated through a user study with mobility-impaired individuals, as described in Chapter 5. In this study, participants were invited to do common tasks regarding email, agenda and conference usage. We observed that participants, including paraplegics and quadriplegics, used alternative modalities more often than traditional ones, improving their interaction with devices and the application itself. We concluded that the multimodal application, which was developed taking into account the specific requirements of mobility-impaired users, can in fact improve these users' ability to interact with a set of communication services that were said to be important in their daily lives.

One way to overcome physical impairments consists on adopting other HCI modalities, such that if a user has difficulties on using his hands, he could opt to use speech instead. But by making a modality available, interaction simply does not improve. As we have seen some modalities may not work properly on some environments or conditions, or simply their adoption does not make a user autonomous, as they could rely on using external devices that had to be configured by someone else. Multimodal systems improve interaction, as making various modalities available, a user can choose the best for him or for any situation. As the last study revealed, a multimodal application offering speech and touch, in addition to traditional modalities, could in fact improve mobility-impaired user's access to ICT, in terms of email, agenda and conference services. Multimodal applications are especially important for

quadriplegics with severe limitations, as they can overcome their limitations regarding the use of traditional modalities and their independence from the need of using still complex devices of their assistive technologies. A solution available not only on a computer, but also on a mobile device, contributes to mobility-impaired citizens' independence and mobility, making the Personal Life Assistant available everywhere.

6.2 Contributions

The work carried out in the first two user studies, to which the author of [68] also contributed, was submitted and accepted for presentation at the at the Social Mobile Web 2010 Workshop, co-located with the 12th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2010) - see Appendix E.

6.3 Future work

Being one of the objectives of this thesis to study with real users the advantages of multimodality, and since only three different modalities were tested in more detail, it would be interesting to study in more deep other HCI modalities, such as: biological sensors, vision-based 3D gestures; gaze-based interfaces, etc.. It would be also very interesting to perform a more in-depth study with a larger mobility-impaired group, including people with more limitations, that could reduce even more the usable modalities set, as for example, study users with: cerebral palsy, voice impairments, visual impairments or even amputees. In order to perceive if the developed Personal Life Assistant, really improves motor-impaired people's lives, a long-term evaluation study could also be performed, in which users would try the prototype in their homes for an extended period of time.

In terms of the prototype, there is a large amount of work to be done on top of that was developed. As suggested by participants, the prototype feature set could be wider, including for example, access to daily reports, news, weather forecasts, movies of the week, radio (which is possible via the media Center), etc. Additionally, and considering that the prototype is a Personal Life Assistant, it could be available on more devices and access more services. Other possible technical improvements could include: multimodal integration, audio streaming and dictation on mobile and IM.

References

[1] A. Jaimes and N. Sebe, "Multimodal Human Computer Interaction: A Survey," 2005.

[2] R. Rosenfeld, D. Olsen, and A. Rudnicky, "Universal Human-Machine Speech Interface: A White Paper," Março, 2000.

[3] Y. Matsusaka, T. Tojo, S. Kubota, K. Furukawa, D. Tamiya, K. Hayata, Y. Nakano, and T. Kobayashi, Multi-person Conversation via Multi-modal Interface — A Robot who Communicate with Multi-user —.

[4] C. Müller and R. Wasinger, Adapting Multimodal Dialog for the Elderly, 2000.

[5] S. Zhu, J. Feng, and A. Sears, "Don't Listen! I Am Dictating My Password!," ASSETS, 2009, pp. 229-230.

[6] X.T. Gao, S.K. Ong, M.L. Yuan, and A.Y. Nee, "Assist Disabled to Control Electronic Devices and Access Computer Functions by Voice Commands," ACM, 2007, pp. 37-42.

[7] M.M. Henriques, C. Teixeira, and M.S. Dias, "Windows vista media center controlled by speech," Media, 2007.

[8] S. Oviatt, "Designing Robust Multimodal Systems for Universal Access," WUAUC, 2001, pp. 71-74.

[9] M. Turunen, J. Hakulinen, J. Hella, J. Rajaniemi, A. Melto, J. Rantala, T. Heimonen, T. Laivo, H. Soronen, M. Hansen, P. Valkama, T. Miettinen, and R. Raisamo, "Multimodal Interaction with Speech, Gestures and Haptic Feedback in a Media Center Application," Ifip International Federation For Information Processing, 2008, pp. 836-837.

[10] M. Turunen, J. Hakulinen, A. Melto, J. Hella, J. Rajaniemi, J. Rantala, T. Heimonen, T. Laivo, H. Soronen, M. Hansen, P. Valkama, T. Miettinen, and R. Raisamo, "Speech-based and Multimodal Media Center for Different User Groups," Interspeech, 2009, pp. 1439-1442.

[11] E. Tse, S. Greenberg, and C. Shen, "Exploring Interaction with Multi User Speech and Whole Handed Gestures on a Digital Table," UIST, 2006.

[12] F. Ferri, A. D'Andrea, P. Grifoni, and A. D'Ulizia, "A Multimodal Pervasive Framework for Ambient Assisted Living," PETRA, 2009.

[13] J. Hugunin and V. Zue, On the design of effective speech-based interfaces for desktop applications.

[14] M. Nielsen, T.B. Moeslund, and E. Granum, "A procedure for developing intuitive and ergonomic gesture interfaces for HCI."

[15] "Project Natal." Retrieved January 10, 2010, from: http://www.xbox.com/en-US/live/projectnatal/

[16] "Sony EyeToy." Retrieved January 10, 2010, from: http://www.us.playstation.com/PS2/Games/EyeToy_Play/ogs/

[17] "Audience Entertainment." Retrieved January 12, 2010, from: http://www.audienceentertainment.com/

[18] "Nintendo Wii." Retrieved January 20, 2010, from: www.nintendo.com/wii

[19] "Apple iPhone." Retrieved January 20, 2010, from: www.apple.com/iphone

[20] "Microsoft Surface." Retrieved January 16, 2010, from: http://www.microsoft.com/surface/

[21] L. Baillie, R. Schatz, R. Simon, H. Anegg, F. Wegscheider, G. Niklfeld, and A. Gassner, Designing Mona: User Interactions with Multimodal Mobile Applications, 2001.

[22] C. Elting and G. Michelitsch, A Multimodal Presentation Planner for a Home Entertainment Environment.

[23] D. Vergados, A. Mariolis, A. Alevizos, and M. Caragiozidis, "Intelligent Services for Assisting Independent Living of Elderly People at Home," PETRA, 2008.

[24] D. Lalanne, L. Nigay, P. Palanque, P. Robinson, and J. Vanderdonckt, "Fusion Engines for Multimodal Input: A Survey," ICMI-MLMI, 2009, pp. 153-160.

[25] R. Iglesias, N.G. Segura, and M. Iturburu, "The Elderly Interacting with a Digital Agenda through an RFID Pen and a Touch Screen," MSIADU, 2009, pp. 63-70.

[26] "My first time on a computer." Retrieved December 5, 2009, from: http://news.bbc.co.uk/2/hi/technology/8353465.stm

[27] "System targets older web users." Retrieved December 5, 2009, from: http://news.bbc.co.uk/2/hi/technology/8353468.stm

[28] "MIDAS: Multimodal Interfaces for Disabled and Ageing Society," Information Technology for European Advancement, 2008.

[29] "MyTobii." Retrieved December 22, 2009, from: http://www.tobii.com/assistive_technology/products/product_overview.aspx

[30] "Magic Key." Retrieved December 21, 2009, from: http://www.magickey.ipg.pt

[31] "EyeWriter." Retrieved January 12, 2010, from: http://www.eyewriter.org/

[32] P.A. Condado and F.G. Lobo, "EasyVoice: Breaking barriers for people with voice disabilities," ICCHP, 2008.

[33] T. Guerreiro, P. Lagoá, H. Nicolau, D. Gonçalves, and J.A. Jorge, "From Tapping to Touching: Making Touch Screens Accessible to Blind Users," IEEE Computer Society, 2008, pp. 48-50.

[34] "JAWS." Retrieved January 15, 2010, from: http://www.freedomscientific.com/products/fs/jaws-product-page.asp

[35] L.M. Reeves, J. Lai, and J.A. Larson, "Guidelines for Multimodal: User Interface Design," Communications of the ACM, vol. 47, 2004, pp. 57-59.

[36] J. ABASCAL and C. NICOLLE, "Moving towards inclusive design guidelines for socially and ethically aware HCI," Interacting with Computers, vol. 17, 2005, pp. 484-505.

[37] Z. Callejas and R. López-Cózar, "Designing Smart Home Interfaces for the Elderly," Sigaccess Newsletter, 2009, pp. 10-16.

[38] S. Keates, P.J. Clarckson, and P. Robinson, "Developing a practical inclusive interface design approach," Interacting with Computers, 2002, pp. 271-299.

[39] S.L. Oviatt, P. Cohen, L. Wu, J. Vergo, L. Duncan, B. Suhm, J. Bers, T. Holzman, T. Winograd, J. Landay, J. Larson, and D. Ferro, Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions, 2000.

[40] R.A. Bolt, "Put-that-there: Voice and gesture at the graphics interface",*Computer Graphics*, vol. 3, 1980, pp. 262-270.

[41] Y. Sun, F. Chen, and V. Chung, "QuickFusion: Multimodal Fusion Without Time Thresholds," MMUI, vol. 57, 2006.

[42] "SmartNav." Retrieved February 18, 2010, from: http://www.naturalpoint.com/smartnav

[43] J. Preece, Y. Rogers, and H. Sharp, Interaction Design: Beyond Buman-Computer Interaction, John Wiley & Sons, Inc., 2002.

[44] M. Jonhston, "Multimodal interfaces," AT&T, 2009.

[45] A. Edwards, A.D. Edwards, and E.D. Mynatt, CHI 2003 Tutorial: Designing for Users with Special Needs, 2003.

[46] M. Turunen, J. Hakulinen, A. Kainulainen, A. Melto, and T. Hurtig, "Design of a Rich Multimodal Interface for Mobile Spoken Route Guidance."

[47] B. Toth and G. Nemeth, "Challenges of creating multimodal interfaces on mobile devices," Elmar 2007, 2007, pp. 171-174.

[48] D.O. Tóth, Bálint (Budapest University Of Technology And Economics and D.O. Németh, Géza (Budapest University Of Technology And Economics, "XML Based Multimodal Interfaces on Mobile Devices in an Ambient Assisted Living Scenario."

[49] S. Harada, J.O. Wobbrock, J. Malkin, J.a. Bilmes, and J.a. Landay, "Longitudinal study of people learning to use continuous voice-based cursor control," Proceedings of the 27th international conference on Human factors in computing systems - CHI '09, 2009, p. 347.

[50] M. Johnstion, S. Bangalores, A. Stent, G. Vasireddy, and P. Ehlen, "Multimodal Language Processing for Mobile Information Access," AT&T Labs-Research, pp. 2237-2240.

[51] S. Dusan and J. Flanagan, "An Adaptive Dialogue System Using Multimodal Language Acquisition," CAIP, Rutgers University, 2001.

[52] Y.S. Choi, C.D. Anderson, J.D. Glass, and C.C. Kemp, "Laser Pointers and a Touch Screen : Intuitive Interfaces for Autonomous Mobile Manipulation for the Motor Impaired," ASSETS, 2008, pp. 225-232.

[53] B. Dumas, D. Lalanne, and R. Ingold, "HephaisTK : A Toolkit for Rapid Prototyping of Multimodal Interfaces," ICMI-MLMI, 2009, pp. 231-232.

[54] B. Dumas, R. Ingold, and D. Lalanne, "Benchmarking Fusion Engines of Multimodal Interactive Systems," ICMI-MLMI, 2009, pp. 169-176.

[55] P. Mistry and P. Maes, "SixthSense: A Wearable Gestural Interface," International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH ASIA 2009, 2009.

[56] L. Nigay and J. Coutaz, "A design space for multimodal systems: concurrent processing and data fusion.," Proceedings of the INTERCHI '93 Conference on Human Factors in Computing Systems, pp. 172-178.

[57] “Samsung Windows Mobile SDK 2.1”. Retrieved June 14, 2010, from: http://innovator.samsungmobile.com/down/cnts/toolSDK.detail.view.do?platformId=2&cntsId=5960

[58] “OpenNETCF Smart Device Framework”. Retrieved June 14, 2010, from: http://www.opennetcf.com/CompactFramework/Products/SmartDeviceFramework/tabid/65/Default.aspx

[59] “Microsoft Office Communicator 2007 R2”. Retrieved June 14, 2010, from: http://office.microsoft.com/en-us/communicator-help/microsoft-office-communicator-2007-product-overview-HA010203715.aspx

[60] “Microsoft Office Communicator 2007 SDK”. Retrieved June 14, 2010, from: http://www.microsoft.com/downloads/details.aspx?FamilyID=ed1cce45-cc22-46e1-bd50-660fe6d2c98c&displaylang=en

[61] “Microsoft Unified Communications Client API SDK”. Retrieved June 14, 2010, from: http://msdn.microsoft.com/en-us/library/bb878684(office.12).aspx

[62] “Unified Communications Managed API 2.0 Speech SDK”. Retrieved June 14, 2010, from: http://msdn.microsoft.com/en-us/library/dd266409(v=office.13).aspx

[63] “Unified Communications Managed API 2.0 Core SDK”. Retrieved June 14, 2010, from: http://msdn.microsoft.com/en-us/library/dd253291(v=office.13).aspx

[64] “Exchange Web Services Managed API 1.0 SDK”. Retrieved June 14, 2010, from: http://msdn.microsoft.com/en-us/library/dd633710(v=EXCHG.80).aspx

[65] “W3C Multimodal Interaction Framework”. Retrieved March 22, 2010, from: http://www.w3.org/TR/mmi-framework/

[66] “EMMA: Extensible MultiModal Annotation markup language”. Retrieved March 22, 2010, from: http://www.w3.org/TR/emma/

[67] R. T. Azuma et al., "A survey of augmented reality", Presence: Teleoperators and Virtual Environments, vol. 6, no. 4, pp. 355-385, August 1997.
[68] Fernando Miguel Pinto, "Multimodal Access to Social Media Services", Unpublished manuscript, Faculdade de Engenharia da Universidade do Porto.

Appendix A

Associação Salvador

Associação Salvador[3] (Salvador's Association) was founded in November 23th 2003 by Salvador Mendes de Almeida (founder and general manager), who became quadriplegic as a result of a traffic accident. The main objective of this association is promoting the interests and rights of mobility-impaired people. Their areas of action focus on integration (quality of life, events, job support), tourism, physical accessibility and road prevention. They also promote research and development and make international cooperation.

Associação Salvador's logo

[3] http://www.associacaosalvador.com retrieved February 10, 2010.

Appendix B

Requirements Analysis Sessions

Being this thesis a Human Computer Interaction study, the first stage consists on apprehend what the user's needs are and what are the difficulties using existing systems. With this aim, some user sessions were carried with a sample of mobility impairments users, selected by ***Associação Salvador*** (view Appendix A).

B.1 Preliminary Requirements Analysis Interview

Below it is the transcription of the preliminary requirements analysis interview.

B.1.1 Subjects Panel

Below we present the subjects panel regarding to this interview:

Participant	Gender	Age	Career	Impairment type
Subject 1	Female	25	Social Psychology intern	Paraplegic
Subject 2	Male	43	Administrative Worker (Informatics)	Quadriplegic
Subject 3	Male	47	Administrative Worker	Paraplegic
Subject 4	Female	26	Unemployed (Social Psychologist)	Paraplegic
Subject 5	Male	28	General Manager	Quadriplegic

B.1.2 About the Interview

This initial questionnaire was conducted in an interactive environment, over the *Microsoft Live Meeting* platform, in January 19th 2010 between 15:30 and 17:20, and took place in FEUP and Associação Salvador. It is important to say that this was the first contact with Associação Salvador.

B.1.3 Interview Transcription

1. On average, how would you describe your computer usage habits:
    a. Never used
    b. Sporadic usage (less than once a week)
    c. Weekly usage (at least once a week)
    d. Daily usage (less than five hours a day)
    e. Intense usage (more than five hours a day)

Participant	Response
Subject 2	5
Subject 3	5
Subject 5	5
Subject 4	5
Subject 1	5

2. On average, how would you describe your smartphone usage habits:
    a. Never used
    b. Sporadic usage (less than once a week)
    c. Weekly usage (at least once a week)
    d. Daily usage (less than five hours a day)
    e. Intense usage (more than five hours a day)

Participant	Response
Subject 2	1
Subject 3	1
Subject 5	1
Subject 4	1
Subject 1	1

Notes: Smartphone's keypads are too small and hard to use.

3. On average, how would you describe your cellphone usage habits:
    a. Never used
    b. Sporadic usage (less than once a week)
    c. Weekly usage (at least once a week)
    d. Daily usage (less than five hours a day)
    e. Intense usage (more than five hours a day)

Participant	Response
Subject 2	4
Subject 3	4
Subject 5	5
Subject 4	4
Subject 1	4

4. How would you rank your level of easiness of use of a computer:
    a. Very Low
    b. Low
    c. Medium
    d. High
    e. Very High

Participant	Response
Subject 2	4
Subject 3	5
Subject 5	2
Subject 4	4
Subject 1	4

Notes: Subject 2 uses pens in order to use a keyboard. Subject 5 uses a gaze-detection system. All subjects demonstrated interest in improving their interaction.

5. How would you rank your level of easiness of use of a cellphone:
    a. Very Low
    b. Low
    c. Medium
    d. High
    e. Very High

Participant	Response

Subject 2	3
Subject 3	5
Subject 5	2
Subject 4	2
Subject 1	3

6. Do you use the computer for entertainment, work, or both?

Participant	**Response**
Subject 2	He uses the computer essentially for work, but at home he sometimes uses it for entertainment purposes
Subject 3	Answered that more than 90% of his time spent in front of a computer is for work purposes, but he still uses it a bit for entertainment, at home.
Subject 5	He uses it mostly for work related activities.
Subject 4	Since she's unemployed, most of her usage is personal.
Subject 1	Since she works at home, she mostly uses the computer for work.

7. Do you use the cellphone for entertainment, work, or both?

Participant	**Response**
Subject 2	He uses the cellphone essentially for work.
Subject 3	More than 90% of his time spent with a cellphone is for work purposes.
Subject 5	He uses it mostly for work related activities.
Subject 4	Since she's unemployed, most of her usage is personal.
Subject 1	Since she works at home, she mostly uses the cellphone for work.

8. What do you usually do on your computer/cellphone?

Participant	**Response**
Subject 2	He browses the web, uses management applications, SAP in his work place, e-mail, some chatting at night. He also likes to watch movies and photos. Subject 2 also mentioned that he does not do audio or video conferencing.
Subject 3	Uses these devices to read daily sports, economy on-line newspapers and to e-mail. He also mentioned that he does not know, nor have the time to play games.

Subject 5	He uses his computer for e-mail, to elaborate spreadsheets in Excel and presentations in PowerPoint, as well as on-line newspaper reading and researching. He also uses his computer for a bit of casual gaming, and once or twice a month for video-conferencing with friends and for work related situations.
Subject 4	Uses her computer for e-mail, Instant Messaging, audio and video-conferencing, as well as for photo viewing, document writing in Word and presentation elaboration in PowerPoint. She also mentioned that she doesn't play games on her computer.
Subject 1	She uses her computer for web browsing, Instant Messaging, e-mail, social network access, document elaboration in Word and presentation creation in PowerPoint. She doesn't use her computer for audio or videoconferencing, as well as for gaming.

9. What are your main difficulties while using a computer?

Participant	**Response**
Subject 2	Key combinations such as ctrl+alt+delete, usb thumbdrive removal and cd/dvd insertion and removal are his main difficulties.
Subject 3	Mentioned that his main difficulty is having to change his glasses to ones designed specifically for computer usage.
Subject 5	He has difficulties writing for long periods of time (more than 5 hours) on the keyboard, suggesting the adoption of features such as auto-completion or most frequently typed words. Some of his other issues regard high movement effort that he has to sustain in order to do simple tasks, as well as dependency on others to perform simple tasks such as putting on eye-glasses designed for gaze control, as well as turning on the computer. He suggested adoption of a Multi-touch interface as a way to simplify turning on the computer.
Subject 4	Her main difficulties are with keyboard and mouse usage, suggesting the adoption of auto-completion writing or voice recognition technologies as keyboard aids and alternative ways of hands-free mouse control.
Subject 1	Her difficulties are focused on keyboard writing, due to some issues with her hands. She also suggested adopting auto-completion writing or voice recognition as additional ways of interfacing with the computer.

10. What are your main difficulties while using a cellphone?

Participant	Response
Subject 2	Answered that his difficulties are related with small keys on the cellphone, forcing him to use his knuckles to type.
Subject 3	Did not have any difficulties with cellphone usage.
Subject 5	Also has issues with small keys on the cellphone, forcing him to use his thumb as a way to write. He mentioned that, due to the position of the power button, it was also very hard for him to turn the cellphone on or off, suggesting adoption of a Multi-touch way to control this feature.
Subject 4	Mentioned that she has difficulties regard using small keys too.
Subject 1	Did not reply to this question.

11. Which interaction modalities have you used to interact with computers and cellphones?

Participant	Response
Subject 2	Has used keyboard, mouse and adaptive pens, noting that he is very accustomed to keyboard and mouse interaction.
Subject 3	Has only used keyboard and mouse.
Subject 5	Has used the keyboard and mouse combination, as well as Multi-touch screens in Windows 7, noting that gestures such as pinch+zoom weren't very easy for him to used, as he can't move his fingers very easily. He has also used voice recognition software such as Philips' Freespeech and IBM's Viavoice, noting that these older systems didn't work very well in open spaces due to the presence of other voices in the environment.
Subject 4	Has also only used keyboard and mouse.
Subject 1	Has also only used keyboard and mouse.

12. If you could use just one modality, which would you choose and why?

Participant	Response
Subject 2	He would use voice recognition and synthesis, as long as its usage was efficient. When asked if he would use gesture commands, he answered that it wouldn't be adequate for him to use as it requires too much movement.
Subject 3	Answered that, out of curiosity, he would like to try voice

	interaction.
Subject 5	Due to his limitations, answered that voice and gaze interaction would be the more adequate interfaces.
Subject 4	Would also like to try voice and maybe gaze interaction.
Subject 1	Would choose voice, due to her finger dexterity limitations.

13. Have you ever used hardware or software targeted towards mobility impaired users?

Participant	Response
Subject 2	Has only used adaptive pens.
Subject 3	He never used.
Subject 5	Uses gaze control glasses and his wheelchair's joystick, as an alternative way to control his mouse's pointer, noting that he uses these devices since he can produce more accurate movements with his neck, than with his hands.
Subject 4	She never used.
Subject 1	She never used.

a. What was good about your interaction with them?

Participant	Response
Subject 2	Mentioned that the pens he uses can be found almost anywhere, thus having a low price point, and that it's also very easy to find alternatives to them.
Subject 3	N/A
Subject 5	Replied that without those devices, he couldn't use any computer.
Subject 4	N/A
Subject 1	N/A

b. What was bad about your interaction with them?

Participant	Response
Subject 2	He did not reply.
Subject 3	N/A
Subject 5	Replied that the particular eye-glasses gaze interface he uses isn't very easy to deploy on other computers, due to requiring specific wiring and software, thus forcing him to always bring his own computer with him, whenever needed.

Subject 4	N/A
Subject 1	N/A

14. Have you ever used or are you using any videocall or audiocall system?

Participant	Response
Subject 2	He sporadically uses Skype.
Subject 3	Rarely uses these kinds of systems.
Subject 5	He uses 3 times per month Skype for videocall.
Subject 4	Once a week she uses MSN Messenger.
Subject 1	She never used videocall but rarely she uses audiocall.

a. What are the difficulties that you have with these systems?

Participant	Response
Subject 2	Did not reply.
Subject 3	Did not reply.
Subject 5	He said that dealing with the hardware (webcam) requires another person, in order to adjust it. But it is better to speak than to write.
	Did not reply.
Subject 4	Talking with other people (using these systems) is confusing.
Subject 1	

15. Have you ever used or are you using email clients?

Participant	Response
Subject 2	Outlook, Hotmail and Gmail.
Subject 3	Outlook and Hotmail. He referred also that he sends about 6 mails per day.
Subject 5	He replied that he uses Outlook, not only the email functionality but also the calendar for work managing. It is used essentially for work issues (80%).
Subject 4	Hotmail and Gmail.
Subject 1	Hotmail and Gmail.

a. You maintain contact with…

Participant	Response

Subject 2	He maintains contact with tens of people.
Subject 3	Did not reply.
Subject 5	He maintains contact with hundreds of people.
Subject 4	Did not reply.
Subject 1	She maintains contact with some people.

16. Have you ever used or are you using any management agenda software?

Participant	Response
Subject 2	He rarely uses Outlook.
Subject 3	Does not use anything.
Subject 5	As referred in 15 he uses Outlook.
Subject 4	She prefers to use a paper agenda.
Subject 1	Hotmail and Gmail.

a. What are the difficulties?

Participant	Response
Subject 2	Did not reply.
Subject 3	Did not reply.
Subject 5	He referred that the main difficulty was the lack of integration between computer and mobile phones (non-existent synchronization).
Subject 4	Did not reply.
Subject 1	Did not reply.

17. Regarding to email, videocall and audiocall capabilities, it is important to be in contact with whom?

Participant	Response
Subject 2	Family, friends and coworkers.
Subject 3	With anyone that is important in a certain moment.
Subject 5	He replied that it is important to be in contact with friends and disabled people that are isolated at home, for information exchange and sharing.
Subject 4	Friends.
Subject 1	Family, friends and other people.

18. Have you ever used or are you using any media center?

Participant	Response
Subject 2	No.
Subject 3	No.
Subject 5	No.
Subject 4	No.
Subject 1	No.

a. Do you find interest in using one?

Subject 2	Yes.
Subject 3	Eventually.
Subject 5	Yes.
Subject 4	Yes.
Subject 1	Yes.

b. Where do you store your multimedia assets (photos, videos, etc)?

Subject 2	Digital camera and external disc.
Subject 3	Personal Computer.
Subject 5	Personal Computer and external disc.
Subject 4	Personal Computer.
Subject 1	External disc.

19. Where you would like that this Personal Life Assistant was present (everywhere, only at home)? And how (desktop, mobile phone, TV, etc)?

Subject 2	Everywhere if possible (mobile phone, tablet PC and TV).
Subject 3	Notebook.
Subject 5	He said that everywhere would be nice and speech could be a good choice. He said also that he can control some home systems (open/close doors and blinds) using the joystick of his wheelchair.
Subject 4	She said that everywhere would be better.
Subject 1	Everywhere if possible (mobile phone, tablet PC or laptop and TV).

B.2 Requirements Analysis Session

Below we present some details of the requirements analysis session,

B.2.1 About the Session

Ten personal sessions were conducted between the days of 17 and 19 of March 2010, in Lisbon, Oeiras and Sintra, as well as in Matosinhos (Oporto) on March 22nd 2010. These sessions focused on testing user interaction in the fields of agenda, email, audio and video conference and media center (more details on Chapter 3).

B.2.2 Observations and participant's opinions

Email task

Participant	Observations	Participant's opinion
Subject 6	• Subject had difficulties in dealing with key combinations – inserting @ (he uses an bent wire) • Subject could not find attach icon • Subject had some difficulties on reading screen, as he had to get ahead • Due to excessive writing time, participant was told to not to write all text	• Participant considered that if an arroba icon could be inserted using his voice or even by touching a screen with a sidebar of icons (that require key combination), would be very helpful • Subject considered Gmail interface easy to use
Subject 2	• Participant had no problems in using the interface	• Subject considered interface very easy to use and did not presented any improvements, as he uses email everyday
Subject 1	• Participant had no problems in using the interface	• Subject considered interface easy to use • Participant considered that speech in dictation mode could help writing big texts • Participant considered that key combination is complicated, and

		could be improved by icons on screen, selected by mouse or even touch (in case of big icons)
Subject 5	Subject did not do the task.	NA
Subject 9	• Participant had no problems in using the interface • Subject considered key combination complicated and keyboard format dependent • Due to excessive writing time subject was told to not do part 4	• Participant considered that key combination is complicated, and could be improved by an icon tool on screen
Subject 11	• Subject demonstrated very experience in using email • Participant used *Sticky Keys* for key combinations	• Participant referred that if he could use speech for dictate large texts it could reduce writing time • Subject considered interface easy to use
Subject 7	• Subject demonstrated very experience in using email • Subject had no problems in using key combinations	• Participant considered interface very easy to use
Subject 8	• Subject demonstrated some difficulties in using interface • Due to excessive writing time subject was told to not do part 4	• Subject referred that interface is easy but she needed more practice
Subject 3	• Subject demonstrated some difficulties in using interface • Due to excessive writing time subject was told to not do part 4	• Subject referred that CC option was hidden • Subject considered interface easy to use and speech and touch could improve usability
Subject 10	• Participant managed to use email without any problem, excluding the fact that he did not know what CC was	• Participant considered that interface was easy to use and there was nothing that could be improved by alternative modalities

Agenda task

Participant	Observations	Participant's opinion
Subject 6	• Subject had some difficulties on using the interface, probably because it was the first time he use it • Participant tried to use right-click, which is not available on web interface (Gmail Calendar) • Subject did not know how to cancel an appointment	• Subject referred that an agenda would be useful for medical appointments
Subject 2	• Participant had some initial problems using the interface, but as he referred he rarely uses it	• Subject considered interface very easy to use and did not presented any improvements
Subject 1	• Participant had no problems in using the interface, although she failed in select correct time slot	• Subject considered interface easy to use • She referred that although existent limitations, she must try to use current interfaces • Participant said that using voice for dictation would be useful, and for command and control not really
Subject 5	Subject did not do the task.	NA
Subject 9	• Participant had no problems in using the interface	• Participant considered that using speech for agenda management, would be very useful
Subject 11	• Subject demonstrated very experience in using agenda	• Participant said that he was accustomed in using Outlook's agenda • Subject considered that eventually touch and speech will be helpful • Subject considered interface easy to use
Subject 7	• Subject demonstrated very experience in using agenda	• Participant considered interface very easy to use, as he uses it everyday
Subject 8	• Subject demonstrated some	• Subject referred that interface is

	difficulties in using interface • Participant did not know how to cancel the appointment and failed to select correct time slot	hard to use
Subject 3	• Subject did not do the task.	• NA
Subject 10	• Participant used agenda without problems, although did not place location in the location slot	• Participant considered that interface was easy to use and probably speech could help

Conference task

Participant	Observations	Participant's opinion
Subject 6	• Subject used application without problems, although there were some difficulties on finding the end call button • Subject were told to initiate an audio-only call, and he started a videocall instead	• Participant considered that the application is simple to use, but searching contact list could be complicated for him • He referred that if speech was available, probably it would be nice to say "Start call"
Subject 2	Subject did not do the task due to technical difficulties.	NA
Subject 1	• Participant used interface without problems, considering that it was her first time of use	• Subject considered interface easy to use and friendly • She referred that conference would be easier to use than instant messaging
Subject 5	Subject did not do the task.	NA
Subject 9	• Participant had no problems in using the interface	• Participant considered that using speech for calls management, as in free hands systems, would be very useful
Subject 11	• Subject demonstrated very experience in using Skype	• Participant considered that it would be nice saying contact's name for initiate a call • Subject considered that eventually touch and speech will be helpful

		• Subject considered interface easy to use
Subject 7	• Subject completed this task without problems	• Participant considered interface easy to use
Subject 8	• Subject demonstrated a lot of difficulties in using interface, but we have to notice that she never used a call application	• Subject referred that these kind of application is very interesting
Subject 3	• Subject did not do the task.	• NA
Subject 10	• Subject were told to initiate an audio-only call, and he started a videocall instead	• Participant considered that interface was easy to use and would not change anything

Media center task

Participant	Observations	Participant's opinion
Subject 6	• Subject used application with some problems - there were some difficulties on exiting slideshow • We have to consider that it was the first time participant tried a media center, and it was in English – participant did not have many English knowledge	• Participant considered media center very interesting • Subject referred that after some training he could use media center without any problems
Subject 2	• Subject manage to use media center without major problems • Participant had some problems on terminating slideshow, he closed media center instead	• Subject considered that after some practice he would easily use media center • Subject referred that using touch for media center control, could be complicated, has quadriplegic could have some arms movement problems, and so speech would be better
Subject 1	• Participant used interface without problems, considering that it was her first time of use • Participant had some problems	• Subject considered interface easy to use • She referred that speech would be interesting but for her would

	on terminating slideshow, he closed media center instead	not bring advantages: *I have hands, if i can use them i will*
Subject 5	Subject did not do the task.	NA
Subject 9	• In the beginning subject felt a little lost, but after some time he used interface without problems • Due to excessive time, point 5 was not done	• Participant referred that it would be interesting using touch to switch between albums and pictures (touching on left or right part of the screen, or even using dragging gestures) • Participant also said that voice would be a nice alternative, using command and control
Subject 11	• Participant used interface without problems, considering that it was her first time of use • Participant had some problems on terminating slideshow, he closed media center instead	• Subject considered interface easy to use • If voice was available, he considered that using it to control states (play video, pause video) and actions (say album name) would be interesting • Participant also said that controlling media center with touch would be nice, using single touch on monitor sides
Subject 7	• Subject completed this task without problems	• Participant considered interface easy to use
Subject 8	• Due to excessive time, point 5 was not done • Subject failed to use interface by herself	• Subject said that after training, she could use a media center without any problems
Subject 3	• Subject needed help to complete this task	• Subject said that he do not consider use any multimedia management tool
Subject 10	• Participant used interface without problems, considering that it was her first time of use • Participant had some problems on terminating video playback, he closed media center instead	• Participant considered that interface was easy to use • Subject said that he would like to use speech and gestures for media center control

B.2.3 Consent form (in Portuguese)

Formulário de consentimento (Original)

Antes de mais obrigado por participar neste estudo. A sua colaboração é essencial para a realização das nossas teses.

Este estudo insere-se no âmbito de 2 teses, cujo objectivo é estudar novas formas de interacção com os dispositivos, recorrendo a interfaces multimodais (que englobam várias modalidades como voz, toque, etc). A tese do Fernando Pinto (FEUP) é mais orientada ao acesso às redes sociais e a do Carlos Pires (FEUP) prende-se com utilização de audio/video-conferência, email, agenda e media center. No final pretende-se desenvolver um protótipo funcional, a fim de testar os resultados do estudo.

Ambas as teses são orientadas pela Profª Eduarda Mendes Rodrigues da FEUP e pelo Profº Miguel Sales Dias da Microsoft e ISCTE.

Durante esta sessão iremos pedir-lhe que realize algumas tarefas simples, com as quais deverá estar familiarizado, e que experimente também alguns dispositivos. No final da sessão iremos recolher as suas opiniões.

Esta sessão irá ser filmada e gravada. Todos os dados em suporte vídeo e audio recolhidos são confidenciais e acessíveis apenas às pessoas envolvidas neste estudo (acima referidas).

No entanto, e para fins meramente ilustrativos, autoriza que algumas imagens/videos aqui recolhidos, sejam publicados nas nossas teses, em conferências, revistas científicas, etc?

[ ] Sim [ ] Não

Vamos-lhe pedir que preencha os seguintes dados (os dados fornecidos são estritamente CONFIDENCIAIS e servem apenas para tratamento estatístico):

- Nome: ____________________
- Idade: ____________________
- Tipo de Limitação: ____________________
- Profissão: ____________________

[Preencher apenas se não participou no estudo preliminar]

- Nível de uso do PC (em média):
  1- Nunca usou,
  2- Uso esporádico (menos de 1 vez por semana)
  3- Uso semanal (pelo menos 1 vez por semana)
  4- Uso diário (menos de 5 horas por dia)
  5- Uso diário (mais de 5 horas por dia)
- Nível de aptidão (conhecimentos de informática):
  1- Baixa
  2- Suficiente

3- Média
4- Boa
5- Elevada

- Usa telemóvel? [ ] Sim [ ] Não
- Usa smartphone? [ ] Sim [ ] Não
- Nível de aptidão no uso de um telemóvel/smartphone:
  1- Baixa
  2- Suficiente
  3- Média
  4- Boa
  5- Elevada
- Nível de uso do telemóvel/smartphone (em média):
  1- Nunca usou
  2- Uso esporádico (menos de 1 vez por semana)
  3- Uso semanal (pelo menos 1 vez por semana)
  4- Uso diário (menos de 5 horas por dia)
  5- Uso diário (mais de 5 horas por dia)

[ ] Assinale a caixa se pretender receber actualizações deste estudo no futuro? (Se sim, indique-nos o seu email: ______________________________)

Declaro que li e compreendi os objectivos desta sessão, participando de livre vontade na mesma.
Assinatura: ____

Os Investigadores:
Fernando Pinto: ____
Carlos Pires: ____

Formulário de consentimento (Duplicado)

Antes de mais obrigado por participar neste estudo. A sua colaboração é essencial para a realização das nossas teses.

Este estudo insere-se no âmbito de 2 teses, cujo objectivo é estudar novas formas de interacção com os dispositivos, recorrendo a interfaces multimodais (que englobam várias modalidades como voz, toque, etc). A tese do Fernando Pinto (FEUP) é mais orientada ao acesso às redes sociais e a do Carlos Pires (FEUP) prende-se com utilização de audio/video-conferência, email, agenda e media center. No final pretende-se desenvolver um protótipo funcional, a fim de testar os resultados do estudo.

Ambas as teses são orientadas pela Profª Eduarda Mendes Rodrigues da FEUP e pelo Profº Miguel Sales Dias da Microsoft e ISCTE.

Durante esta sessão iremos pedir-lhe que realize algumas tarefas simples, com as quais deverá estar familiarizado, e que experimente também alguns dispositivos. No final da sessão iremos recolher as suas opiniões.

Esta sessão irá ser filmada e gravada. Todos os dados em suporte vídeo e audio recolhidos são confidenciais e acessíveis apenas às pessoas envolvidas neste estudo (acima referidas).

No entanto, e para fins meramente ilustrativos, autoriza que algumas imagens/videos aqui recolhidos, sejam publicados nas nossas teses, em conferências, revistas científicas, etc?

[ ] Sim [ ] Não

Vamos-lhe pedir que preencha os seguintes dados (os dados fornecidos são estritamente CONFIDENCIAIS e servem apenas para tratamento estatístico):

- Nome: ____________________
- Idade: ____________________
- Tipo de Limitação: ____________________
- Profissão: ____________________

Declaro que li e compreendi os objectivos desta sessão, participando de livre vontade na mesma.

Assinatura: ____________________

Os Investigadores:

Fernando Pinto: ____________________

Carlos Pires: ____________________

Appendix C

Prototype evaluation session

Below we present some details about the prototype evaluation session.

C.1 About the Session

Due to some dropouts, only five subjects participated in this session. Therefore, five personal sessions were conducted between the days of 1 and 4 of June 2010, in Microsoft, Porto Salvo, Oeiras. Like on requirements analysis session, all participants signed a consent form, stating that they participated in this session by their free will.

C.2 Consent Form (in Portuguese)

Formulário de consentimento

Antes de mais obrigado por participar neste estudo. A sua colaboração é essencial para a realização das nossas teses.

Este estudo insere-se no âmbito de 2 teses, cujo objectivo é estudar novas formas de interacção com os dispositivos, recorrendo a interfaces multimodais (que englobam várias modalidades como voz, toque, etc). A tese do Fernando Pinto (FEUP) é mais orientada ao acesso às redes sociais e a do Carlos Pires (FEUP) prende-se com utilização de audio/video-conferência, email,

agenda e media center. Esta é a fase final do estudo na qual se pretende avaliar um protótipo funcional, desenvolvido tendo em conta resultados de sessões anteriores.

Ambas as teses são orientadas pela Profª Eduarda Mendes Rodrigues da FEUP e pelo Profº Miguel Sales Dias da Microsoft e ISCTE.

Durante esta sessão iremos pedir-lhe que realize algumas tarefas simples, usando o nosso protótipo, que consiste num computador (touchsmart) e num smartphone (omnia). No final da sessão iremos recolher as suas opiniões.

Esta sessão irá ser filmada e gravada. Todos os dados em suporte vídeo e audio recolhidos são confidenciais e acessíveis apenas às pessoas envolvidas neste estudo (acima referidas).

No entanto, e para fins meramente ilustrativos, autoriza que algumas imagens/videos aqui recolhidos, sejam publicados nas nossas teses, em conferências, revistas científicas, etc?

[ ] Sim [ ] Não

Vamos-lhe pedir que preencha os seguintes dados (os dados fornecidos são estritamente CONFIDENCIAIS e servem apenas para tratamento estatístico):

- Nome: ____

- Idade: ____
- Tipo de Limitação: ____
- Profissão: ____

Declaro que li e compreendi os objectivos desta sessão, participando de livre vontade na mesma.

Assinatura: ____

Os Investigadores:

Fernando Pinto: ____

Carlos Pires: ____

Appendix D

Prototype user manual

In this section we will explain the prototype interface, regarding how a user can interact with it. As referred on Chapter 4, user can execute common actions using all three modalities (see Table 32).

D.1 Desktop

General considerations:

- Functionalities available on all windows

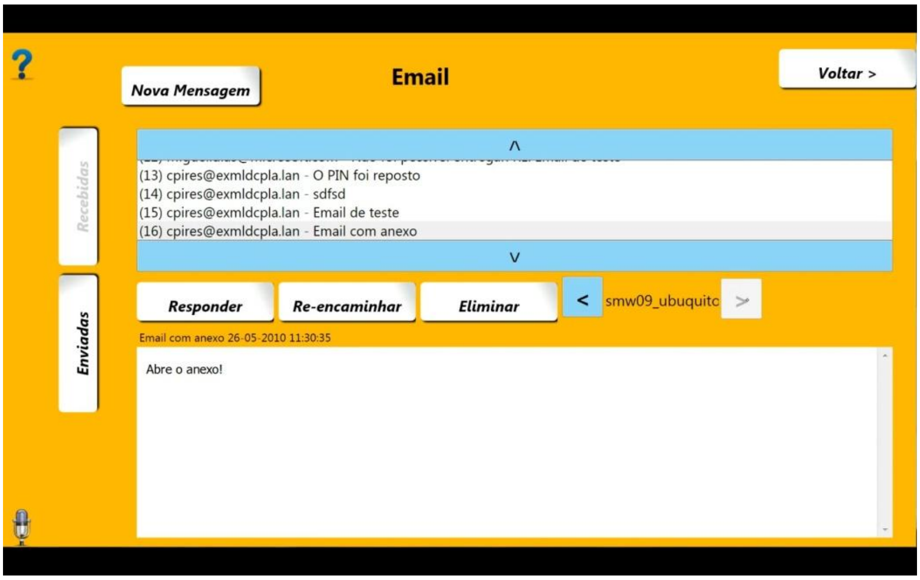
- **Help -** help is available on all windows and it is activated by saying "ajuda" (help) or by pressing its button. Help button is present on the upper left corner. After selected, TTS will describe what it is possible to do on current screen.
- **Microphone status** - Microphone status icon (displayed on the bottom left corner), indicates weather microphone is muted or not. If *PLA i*s muted, what user says is ignored by it. To mute it, user can say "não ouças" (stop listening) or touch the icon. To start listening user can say "ouve" (listen) or touch the icon.
- **TTS control** - in order to stop help or whatever TTS is saying, user can say "cala-te".

Main menu- this screen appears after *PLA* (desktop) starts

User can:

- select Email, Agenda or Conference;
- select "Assistente" to start an audio call to Exchange's OVA;
- select "Sair" to close *PLA*

NOTE: "Redes Sociais" is related with Fernando's work (see [68])

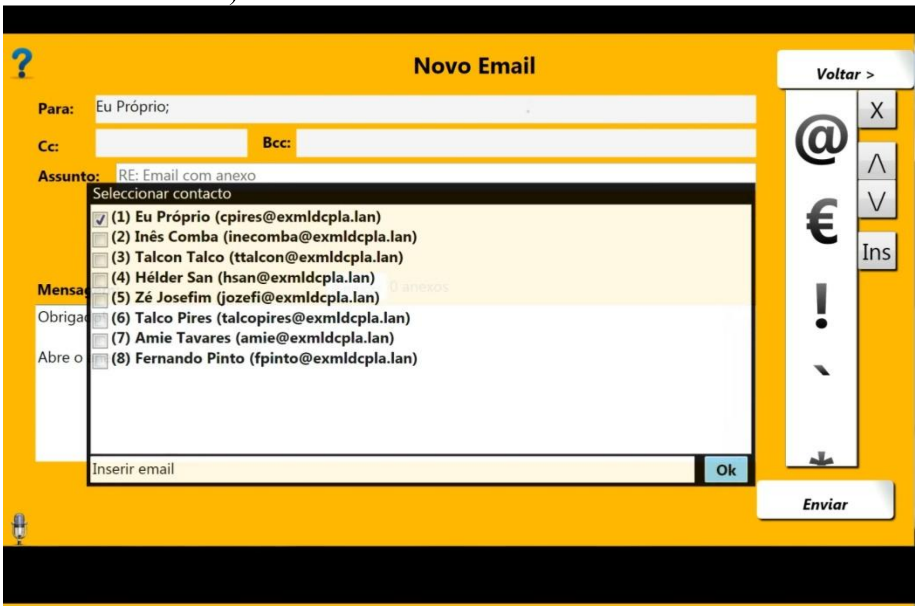


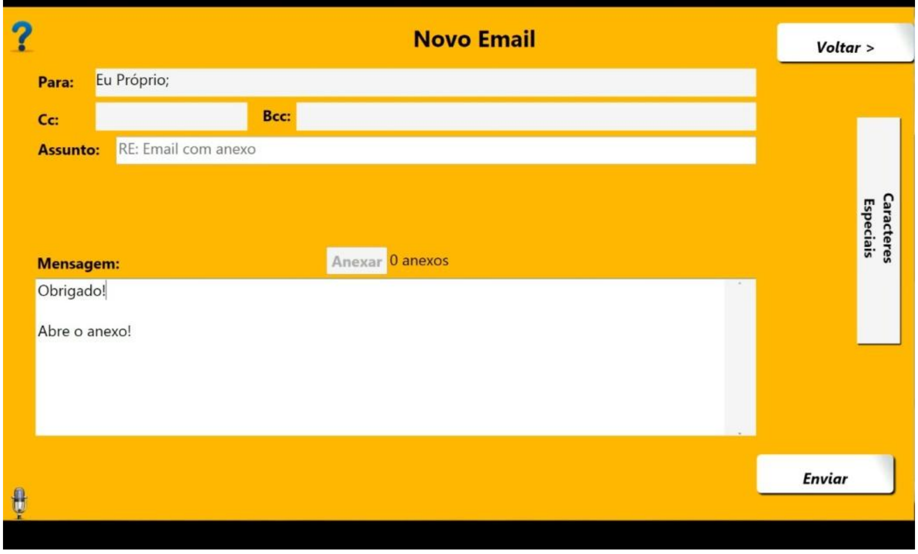
Email main screen- this screen appears after selecting "Email" on main menu

User can:

- return to main menu ("Voltar >");
- create a new email message ("Nova Mensagem");
- select an email message from the list, by saying its number - message details will appear on the window's bottom half;
- open the email message's attachments (if available), by first select the attachment to open (user can say "próximo anexo" or "anexo anterior") and then click on it, after user can save the file (by saying "guardar ficheiro") or open the file (by saying "abrir ficheiro") - attachments menu is beside "Eliminar" button;

- see received messages (selecting “Recebidas”) or see sent messages (selecting “Enviadas”).

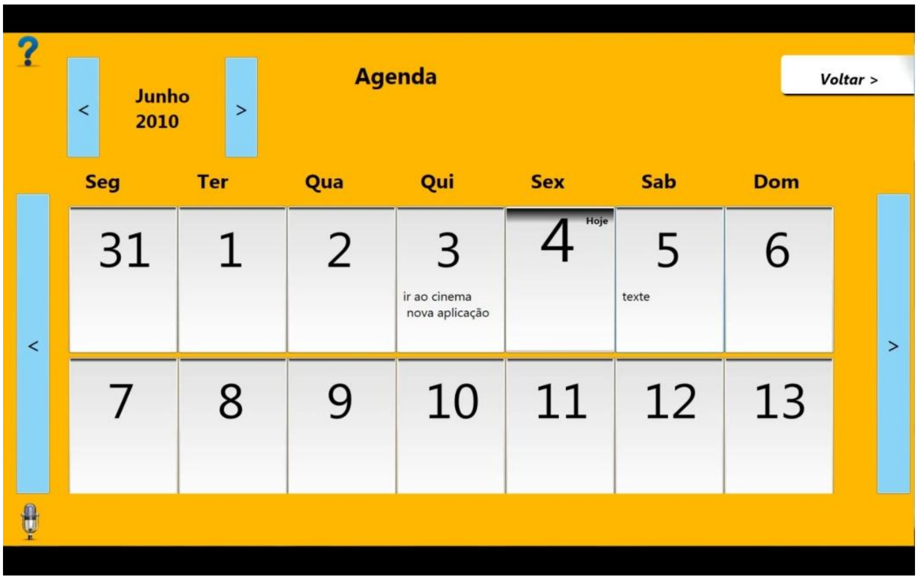




Email message details screen - this screen appears after selecting New email message, reply to an email message or forward an email message on email main screen

User can:

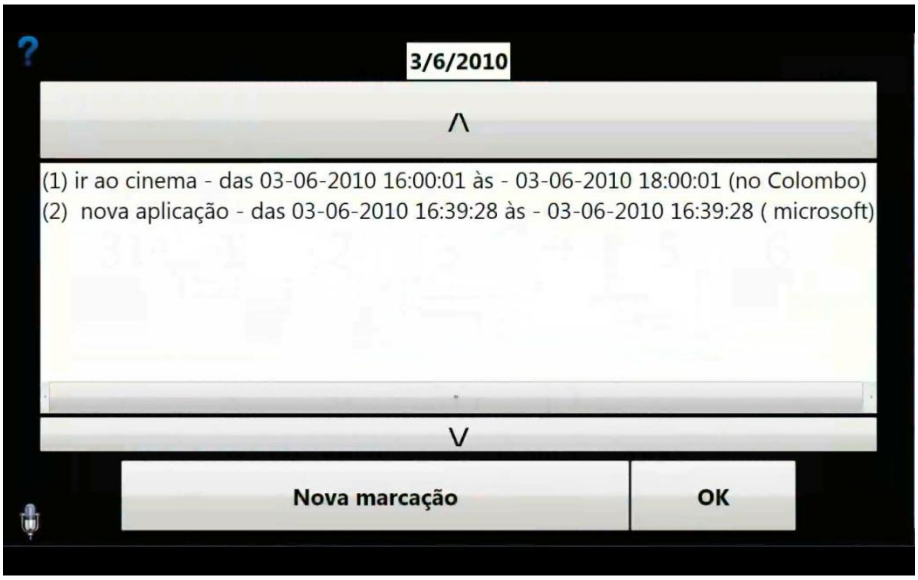
- return to email main screen (“Voltar >”);
- select recipients from contact’s list or insert a new email address (“Para”, “Cc” or “Bcc”) - see the first screenshot;
- edit message’s subject “Assunto” or message’s body “Mensagem”;
- insert or delete attachments (“Anexar”);
- use the sidebar (“Caracteres Especiais”) in order to insert symbols on a selected text box
- send the email (“Enviar”);



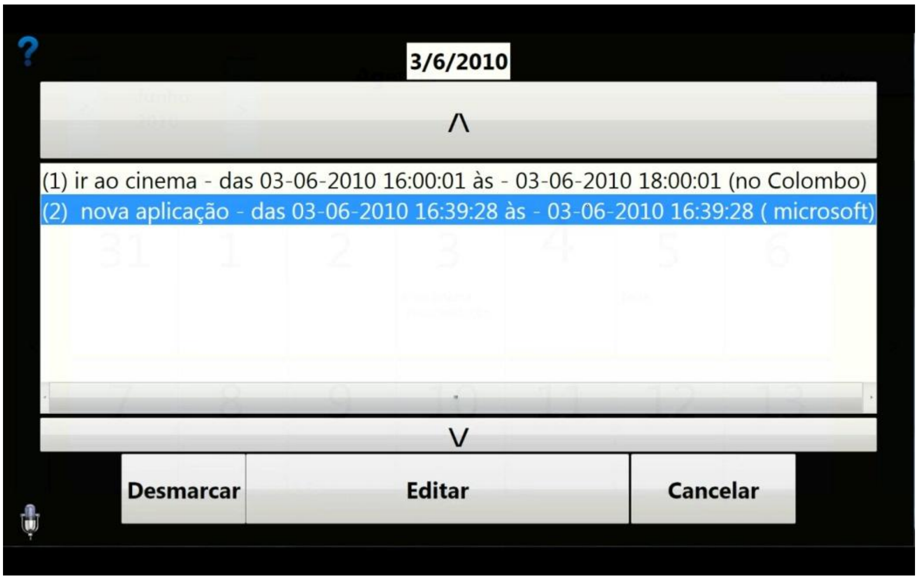
Agenda main screen - this screen appears after selecting "Agenda" on main menu

User can:

- return to main screen ("Voltar >");
- select a day by saying its number;
- navigate one week to the future or to the past, by saying "Próxima semana" (next week) or "Semana anterior" (previous week);
- navigate to a month by saying its name or to the next or previous month (by saying "Mês seguinte" or "Mês anterior" respectively);

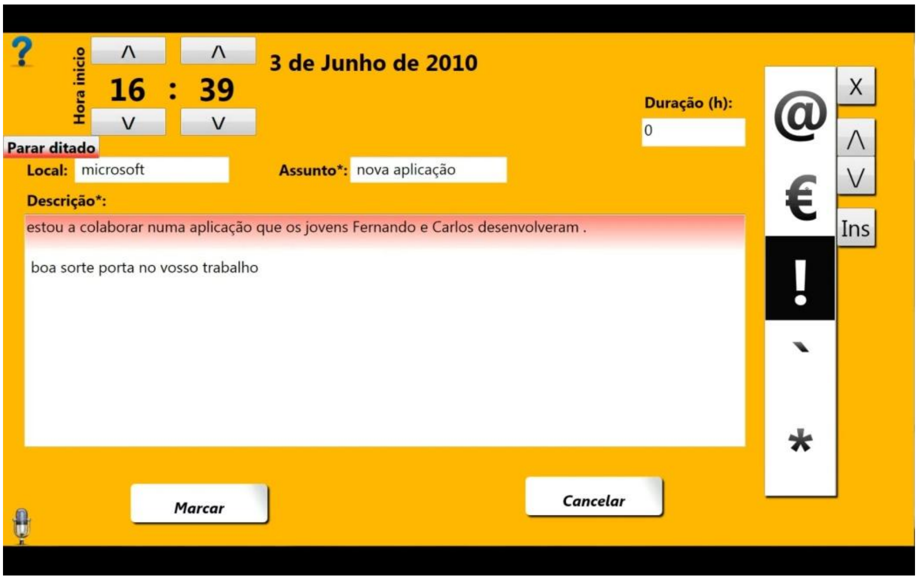


Selected day screen - this screen appears after a day is selected on agenda, showing selected day's appointments



User can:

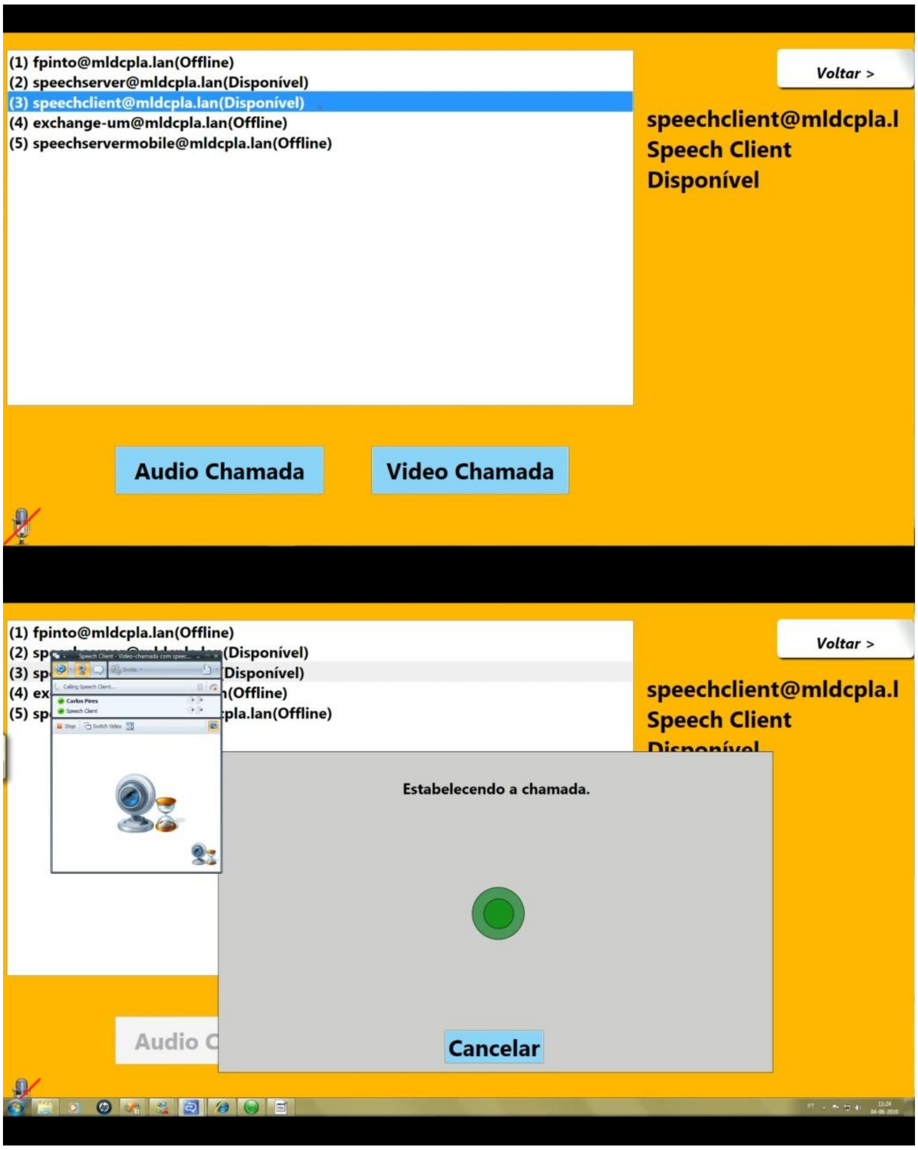
- return to agenda main screen ("OK");
- create a new appointment ("Nova marcação");
- delete a selected appointment ("Desmarcar");
- edit a selected appointment ("Editar");
- cancel actions on selected appointment ("Cancelar");



Appointment details screen - this screen appears whenever user wants to edit an appointment or create a new one

User can:

- return to selected day screen ("Cancelar");
- edit location ("Local"), subject ("Assunto"), description ("Descrição") and duration in hours ("Duração");
- edit start hour ("Hora início") by using buttons to increase or decrease time values or by saying "N horas" to set hours and "N minutos" to set minutes (where N is a number)
- use sidebar to insert symbols;
- save the appointment ("Marcar");



Conference screen - this screen appears when user selects Conference on main menu

Considerations: speech is not available in this screen.

User can:

- return to main menu (“Voltar”);
- select a contact;
- start an audio call (“Audio Chamada”) or a video call (“Video Chamada”) with a selected contact;
- cancel a dialing call (“Cancelar”);
- stop a started call (“Terminar Chamada”);

D.2 Mobile

General considerations:

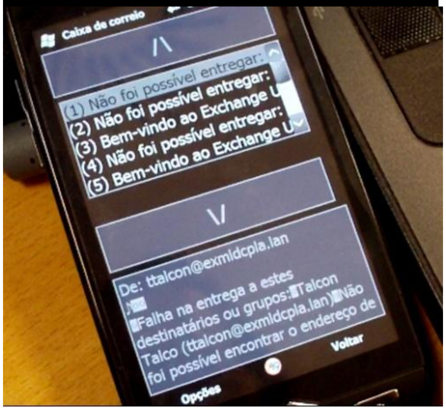
- In order to activate speech on mobile, user must press smartphone's middle button in order to start listening mode, and press again to stop it.
- Dictation is not available.
- To use virtual keyboard, user must press the red button presented on screen's footer.

Main screen - this screen appears after *PLA* (mobile) starts

User can:

- exit application ("Sair");
- select Email, Agenda or Conference;

NOTE: "Redes Sociais" is related with Fernando's work (see [68])



Email main screen - this screen appears after user selects Email on main screen

User can:

- back to main menu ("Voltar");
- select an email message (user can say its number);
- navigate to next or previous message (user can say "próxima" or "anterior" respectively; 3D gestures are also available);
- delete ("Eliminar"), reply ("Responder") or forward ("Reencaminhar") a selected email message - under options ("Opções");
- create a new email ("Novo email") - under options ("Opções");

Edit email message screen - this screen appears after user selects Email on main screen

User can:

- back to email main menu (“Voltar”);
- edit recipients (“Destinatários”);
- edit subject and body text (“Assunto” and “Mensagem”);
- send the email message (“Enviar”);

Recipients screen - this screen appears after user selects recipients option on edit email message screen

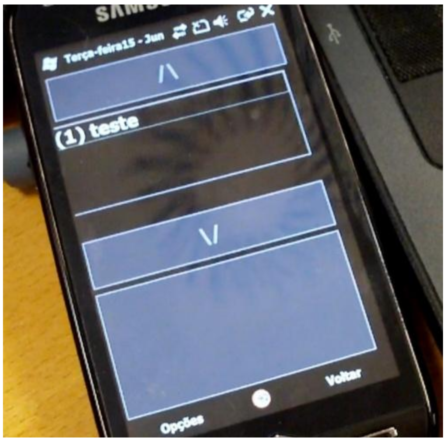
User can:

- back (which also saves recipients) to edit email message screen (“OK”);
- insert a new email address -first user must: select “Inserir email” text box, then use virtual keyboard to insert text, and finally select “Inserir”
- change which contacts will be on To, Cc or Bcc list, by selecting them (user can say contact’s number)

Agenda screen - this screen appears after user selects Agenda on main menu

User can:

- return to main menu (“Voltar”);
- select a day (user can say its number or its weekday);
- change to month view (“Vista mensal”), where user can see and select another month (user can say its name)
- navigate to next or previous week (user can say “próxima semana” or “semana anterior” respectively; user can also use 3D gestures);



Agenda day screen - this screen appears after user selects a day on Agenda

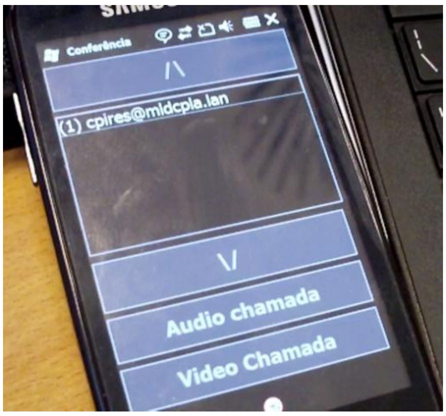
User can:

- back to Agenda screen ("Voltar");
- select an appointment (user can say its number);
- navigate to next or previous appointment (user can say "próxima" or "anterior" respectively; 3D gestures are also available);
- delete ("Eliminar") or edit ("Editar") a selected appointment - under options ("Opções");
- create a new appointment ("Nova marcação") - under options ("Opções");

Appointment screen - this screen appears whenever user wants to create a new appointment or edit an existent one (on Agenda day screen)

User can:

- back to Agenda day screen ("Voltar");
- edit subject ("Assunto"), description ("Descrição"), duration ("Duração") or location ("Local");
- edit appointment start hour, using hour/minutes increase or decrease buttons;
- save the appointment ("Guardar");



Conference screen - this screen appears when user selects "Conferência" from main menu

Considerations: this interface works only has desktop's remote controller.

User can:

- back to main menu ("Voltar");
- select a contact (user can say its number);
- navigate to next or previous contact;
- start an audio call ("Audio chamada") or video call ("Video chamada") with a selected contact;
- stop a started call ("Terminar chamada");

Appendix E

Social Mobile Web 2010 Paper

A paper was submitted and accepted for presentation at the Social Mobile Web 2010 Workshop (http://thesocialmobileweb.org):

Galinho Pires, C., Miguel Pinto, F., Mendes Rodrigues, E., Sales Dias, M. **Improving the Social Inclusion of Mobility Impaired Users**. Proc. of the Social Mobile Web 2010 Workshop, at 12th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2010, Sep. 2010