Universidade de Aveiro
2015

Departamento de Eletrónica,
Telecomunicações e Informática

Diogo da Cruz Vieira

Evolução de uma Plataforma de Interação Multimodal e Aplicações

Enhanced Multimodal Interaction Framework and Applications

Universidade de Aveiro
Departamento de Eletrónica, Telecomunicações e Informática
2015

Diogo da Cruz Vieira

Evolução de uma Plataforma de Interação Multimodal e Aplicações

Enhanced Multimodal Interaction Framework and Applications

"*Anything that can go wrong will go wrong.*"

— Murphy's law

Universidade de Aveiro
2015

Departamento de Eletrónica,
Telecomunicações e Informática

Diogo da Cruz Vieira

Evolução de uma Plataforma de Interação Multimodal e Aplicações

Enhanced Multimodal Interaction Framework and Applications

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Mestrado Integrado em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Doutor Prof. Doutor António Joaquim da Silva Teixeira, Professor Associado do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro, e do Doutor José Miguel de Oliveira Monteiro Sales Dias, Professor Associado Convidado do Departamento de Ciências da Informação do Instituto Universitário de Lisboa.

Este trabalho é dedicado à minha Avó.

o júri / the jury

presidente / president

Professora Doutora Maria Beatriz Alves de Sousa Santos
Professor Associado C/ Agregação da Universidade de Aveiro

vogais / examiners committee

Professor Doutor Daniel Jorge Viegas Gonçalves
Professor Associado da Universidade de Lisboa - Instituto Superior Técnico

Professor Doutor António Joaquim da Silva Teixeira
Professor Associado da Universidade de Aveiro

agradecimentos /
acknowledgements

Gostaria de agradecer ao Prof. Doutor António Joaquim da Silva Teixeira pelo seu apoio e orientação para o sucesso deste trabalho, assim como aos membros do MLDC, em especial ao Prof. Miguel Sales Dias, João Freitas, Luís Sousa, pela sua ajuda e cooperação.

Um agradecimento especial para a minha namorada Andreia Catarina, pela sua companhia e paciência, por todo o seu apoio e carinho, importantes e essenciais para que eu conseguisse chegar até este patamar.

Aos meus pais por me aturarem e me proporcionarem a oportunidade de frequentar o curso. A pessoa que sou hoje também se deve a vocês.

Aos meus colegas do IEETA, Alberto, Nuno, Ana Isabel e Mário, pela sua companhia e pelos bons momentos passados no departamento.

Ao Joel Santos e Filipe Oliveira que me acompanharam praticamente durante todo o meu percurso académico, sem eles tudo o que aconteceu durante esse tempo provavelmente não teria a mesma graça e tudo seria muito mais difícil.

Também agradeço a todos os meus amigos, aos dos bons e maus momentos, os das aventuras e maluqueiras, os que me apoiaram e fizeram levantar a cabeça e olhar para à frente quando as coisas não corriam bem, e aos que me deixaram memórias para mais tarde poder recordar. Vocês sabem quem são.

Para todos um grande obrigado.

Palavras Chave

DEA, IHC, olhar, modalidade, plataforma multimodal, multi-dispositivo, fusão de modalidades.

Resumo

A investigação em Interação Humano-Computador (IHC) explora a criação de novos cenários e formas de utilização de dispositivos, permitindo que aqueles com menos capacidades ou deficiências possam também utilizar a tecnologia. Projetos como o European AAL PaeLife ou IRIS exploram o uso de múltiplas modalidades para atingir esse objetivo. Esta tese apresenta uma evolução da plataforma multimodal existente e utilizada nos projetos AAL anteriores com os seguintes três principais pontos: adição de um novo componente de modalidade à lista de modalidades já disponíveis que permite a interação com o olhar; a criação de um sistema de pesquisa para encontrar outros dispositivos que executam a mesma plataforma multimodal possibilitando a troca de contexto entre os dois dispositivos; permitir que os componentes de modalidade existentes possam ser usados em conjunto com a nova modalidade de olhar através da criação de um processo de fusão na plataforma. Estas melhorias foram apresentadas em cenários relacionados aos usados no PaeLife e IRIS, para os idosos e para uma criança com uma desordem do espectro autista.

Keywords

ASD, HCI, gaze, modality, multimodal framework, multi-device, modalities fusion.

Abstract

The research in Human-Computer Interaction (HCI) explores the creation of new scenarios and forms of using devices, enabling those with less capacities or impaired to also use technology. Projects such as the European AAL PaeLife or IRIS explore the use of multiple modalities to achieve that goal. This thesis presents an enhancement to the existing multimodal framework used on the previous AAL project with the main three points: addition of a new modality component to the list of available modalities that allow the interaction using gaze; the creation of a search system for finding other devices running the same multimodal framework and enable the exchange of context between the two devices; enable the existing modality components to be used together with the new gaze modality by adding a fusion process to the framework. These improvements were presented in scenarios related to the ones used on PaeLife and IRIS, for the elderly and for a child with an autistic spectrum disorder.

Contents

List of Figures

List of Tables

Glossary

AAC	Augmentative and Alternative Communication
ADSL	Asymmetric Digital Subscriber Line
ALS	Amyotrophic Lateral Sclerosis
ASD	Autism Spectrum Disorder
ASR	Automatic Speech Recognition
BCI	Brain-Computer Interfaces
CMS	Content Management System
DTT	Discrete Trial Training
EEG	Electroencephalography
FIPA	Foundation For Intelligent Physical Agents
HCI	Human Computer Interaction
IM	Interaction Manager
MC	Modality Component
MMI	Multimodal Interaction
PECS	Picture Exchange Communication System
PLA	Personal Life Assistant
UPnP	Universal Plug and Play
UPnP	Universal Plug and Play
W3C	World Wide Web Consortium

CHAPTER 1

Introduction

1.1 Motivation

We live in an era where computers are no longer large machines and with low portability as they were before. Devices such as smartphones and tablets are more popular and used anywhere. The characteristics of this type of devices, among which its portability lead us to seek a way to innovate how people access and use the technology, by creating applications ensuring its ease of use for all types of user, but also capable of supporting the user in his/her tasks during daytime.

There are numerous cases and ways in which their use has been recognized as advantageous. Among them we find the use of technology in supporting the elderly, such as the use of monitoring systems that provide an immediate response to problems [1] or multimodal personal assistants capable to detect different types of interaction and unifying access to various services [2] with the purpose of simplifying the access to the users. In the context of the people with disabilities there are multiple success stories in using computers, assisting the development of the ability to perform tasks in real life [3], or also by using programs on portable devices for education or as an assistant for their levels of interaction with others [4].

The Autism Spectrum Disorder (ASD) is characterized by the impairment in social interaction and communication, as well as in manifestations during repetitive use of certain objects, activities and interests. In the absence of a cure for this type of disorder, the treatment of autistic children is crucial to the reduction of symptoms related to the disease [5], whether in households or educational institution [6].

There are many and different impairments that prevent or hamper the use a computer, therefore seeking new methods of interaction for those who have less or no chance to access the new technologies has been an interesting research theme and the main motivation to develop new solutions.

1.2 THESIS CONTEXT

1.2.1 PAELIFE PROJECT

Paelife (*Personal Assistant to Enhance the Social Life of the Seniors*) [7] is a project aimed at retired people with some knowledge in technology, in order to combat social isolation. For that, a personal assistant called AALFred provides various functionality including sending messages, calendar, weather and news, all that within a single application. In addition, it is also possible to interact with the application in multiple ways, either by touch, uttering voice commands, with gestures, or also by using a keyboard or mouse, allowing users to choose the method they feel most comfortable with. The application was developed for Windows 8.1 and will be available for download on the Windows Marketplace store.

1.2.2 IRIS PROJECT

The IRIS Project [8] aims to provide a communication platform, especially for people with speech problems or elderly. The interaction with the platform can be made through various natural communication interfaces such as speech, gestures, pictograms and animated characters. The services provided by the application will allow easy access to social networks, friends and family who are far away, with the purpose to fight social exclusion of people with special needs or people with disabilities.

1.3 PROBLEM STATEMENT

When speaking of interaction and multimodality we may focus on how the application from PaeLife Project (ALLFred) handled all the diversity of possible methods the user could use to interact with it. The element responsible for that task is a multimodal framework [9], responsible to handle and process all the user interactions with a device. This modular framework consists on a central unit called Interaction Manager (IM) and it is used along with different modules, each one allowing the interaction with different modalities, such as gestures or speech. However, and in order to have a larger number of users capable of using the computer, there are still some modalities that speech or movement impaired users can't use.

Also, the interaction made by users is normally limited to one device scenarios. The multimodal framework has demonstrated examples of multiple devices complementing themselves, by sharing the same multimodal context but with each device showing different information. This implementation did not focused on the use of portable devices commonly used everywhere, turning almost impossible the integration of this device in a completely different multimodal scenario.

1.4 OBJECTIVES

In order to fill the gap described in the previous problems, it is then proposed an enhancement of the current multimodal framework with two new distinct features:

- the addition of a new modality, enabling the use of gaze as a method of interaction. Also, make possible the concurrent use of modalities such as gaze and speech, with the goal of creating new and better methods of interaction in the multimodal framework;
- create an autonomous discovery system for the multimodal framework, allowing two different devices on the same network to automatically connect and start exchanging their multimodal context, extending the current single device scenarios.

The work for this thesis was then planned to start with a literature research to contextualize with other type of solutions, when they exist, with a theme related to our work, and understand if it is possible to create something better and how. Also, and since this would be the root of all the work, conduct a deeper analysis of the multimodal framework and what is its architecture. Then, by using a user centered approach, establish possible user scenarios and goals. Based on these previous points, develop one or more prototypes aligned with the created scenarios and conduct an evaluation to determine if the goals were obtained.

The following list summarizes the thesis work objectives:

- Development of detection capability and establishing communication between multiple devices;
- Add mode(s) to the Multimodal Framework being developed by the University of Aveiro in alignment with the IRIS project. One possibility is the development of an approach based on eye tracking mode;
- Definition of modes of use of an application when running on multiple devices;
- Continue the work in Paelife project in particular the ability of redundant and complementary use of multimodal virtual assistants;
- Development of a "demonstrator" application provided with multimodal interaction and multi-device capability. The development must be based on the application of an iterative process.

1.5 CONTRIBUTIONS

The work developed on this thesis resulted in the following two contributions:

- An article, presented on the ASSETS 2015 conference, with an initial fusion prototype used in a multimodal speech and gaze interaction
  Diogo Vieira, João Dinis Freitas, Cengiz Acartürk, António Teixeira, Luís Sousa, Samuel Silva, Sara Candeias, and Miguel Sales Dias. 2015. "Read That Article": Exploring Synergies between Gaze and Speech Interaction. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15). ACM, New York, NY, USA, 341-342. DOI=http://dx.doi.org/10.1145/2700648.2811369
- A future chapter to be included in the book named "Multimodal Interaction with W3C Standards: Towards Natural User Interfaces to Everything", D. Dahl, Springer, 2016, presenting different approaches to tackle multi-device interaction scenarios by using a multimodal framework

CHAPTER 2

Background and Related work

In this chapter we describe the background context of this thesis, as well as some other related studies that will allow us to realize what was already made to tackle the issues related to the theme of this work. The articles and papers search was focused primarily on publications or other related work on personal assistants, the use of multimodality in applications, existing eye tracking devices and software, and ASD applications from the year 2012, by using searching tools such as *Publish or Perish*, *Google Scholar* and *Mendeley*. Some applications for ASD without any kind of publication were still listed for analysis and comparison, since they were labeled as recommended applications on various web pages with information about ASD.

The Section 2.1 explains what is a personal assistant and presents some of the most popular personal assistants; in the Section 2.2 what work has been done and what types of modalities have been used in multimodal applications; Section 2.3 presents some platforms and concepts based on the use and communication between multiple devices; finally Section 2.5 is a brief explanation of what is the Autism Spectrum Disorder and what kind of applications exist for this type of disease.

2.1 Personal assistants

A personal assistant is a software agent able to help users in their tasks, facilitating the use of specific equipment or device, or even automatically perform some of their tasks in real life. All with the goal of reducing the time that is required in most time-consuming tasks, such as filtering or sorting the e-mail, organizing activities or tasks, making purchases or planning a trip. ”The notion of personal assistant is very wide”, as mentioned in the personal assistants specification document [10] created by the Foundation For Intelligent Physical Agents (FIPA). Figure 2.1 presents the model of a personal assistant defined by FIPA, and it can be seen that in addition to the features that are present in the agent, in this case *Agenda*, *Directory* and *User Profile*, an interface to communicate with other agents is not required .

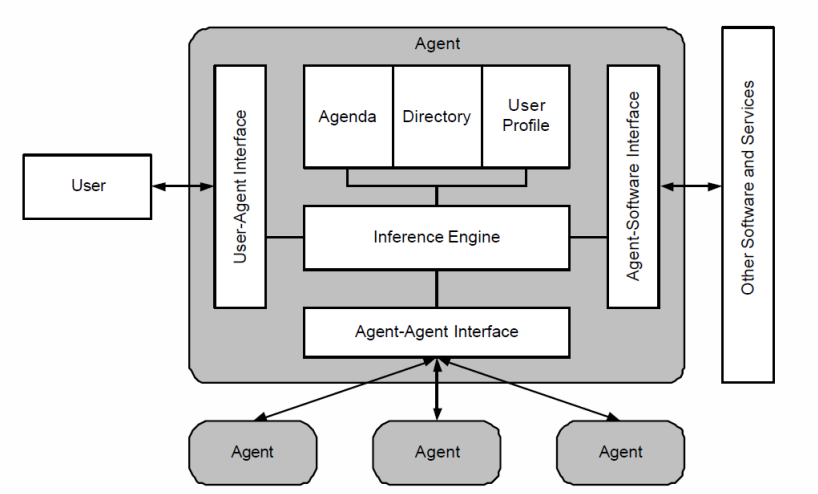


Figure 2.1: The Personal Assistant Reference Model (Source: [10])

The use of personal assistants in computer systems is not recent. Open Sesame! [11] was commercially released as a personal assistant for Microsoft Windows and MacOS 7.0 , capturing and registering the actions the user repeated, asking him/her if, for example, an application on the computer that was opened several times could be added to the Apple menu in order to facilitate its access, as shown in Figure 2.2.

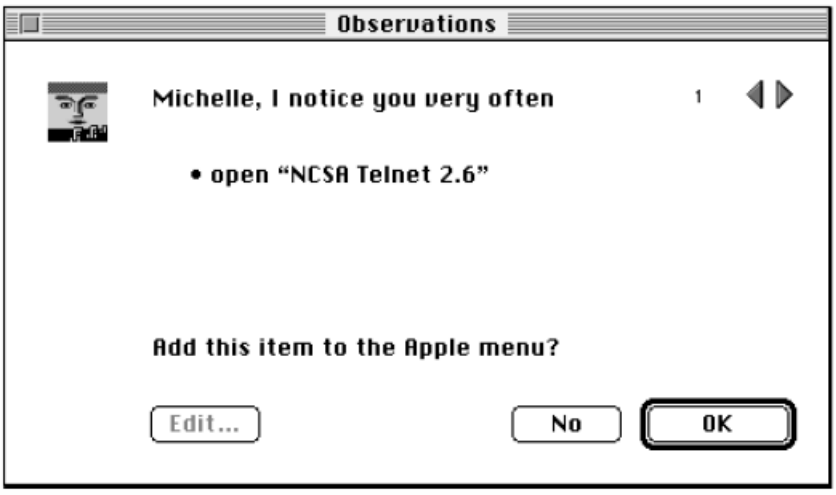


Figure 2.2: An Open Sesame suggestion (Source: [11])

Since then personal assistants had a great evolution both in their capabilities and their use. The ability to interact and communicate with an application by multiple ways (view Section 2.2) made possible certain actions, harder and complicated to complete by some groups of users, more intuitive and accessible. At the same time the computing power evolution has further extended the horizons of a personal assistant. We have the cases of the currently most known assistants, Siri[1], Cortana[2] and GoogleNow[3], capable of executing a large number of tasks, only by uttering voice commands.

[1] `http://www.apple.com/ios/siri`
[2] `http://www.windowsphone.com/en-us/how-to/wp8/cortana/meet-cortana`
[3] `http://www.google.com/landing/now`

It is important to note that for an agent to be labeled personal assistant it does not necessarily have to perform all sorts of tasks or the same type of modalities like any of the assistants described in this section. There is no rule stating which are the minimum required attributes a personal assistant should have, though multiple sources have similar ideas for their basic characteristics [12] [13]:

- Must be independent and operate without guidance from the user;
- Learn from the commands received and after some time increase its operating performance;
- Capable of communicating with other agents through a language recognized by both, when the situation requires several workers to perform different functions;
- Establish a set of characteristics or details related to its user or user group, set accordingly to its functionality;

AALFred, a Personal Life Assistant (PLA) by the PaeLife Project [2] was developed for elderly users with some experience in using computers, offering a variety of services particularly in the areas of social interaction and entertainment, consists of a set of different modules for tasks such as messaging services, calendar, social networking, weather information, news, and others. The PLA also allows interaction by different languages like English, French, Hungarian, Polish and Portuguese, both for input and output, and other ways of interaction - a Microsoft Kinect sensor for gestures, touch (when using a mobile device: tablet), keyboard and mouse. Since the target population of this PLA is a delicate group of people propitious for some difficulties and limitations, these type of applications should take that into account and provide users an easy and more natural computer interaction.

2.2 MULTIMODALITY

2.2.1 MULTIMODAL SYSTEMS

A multimodal system is a computerized system capable to simultaneously interpret stimuli of different methods of interaction. Likewise, this type of system may be also able to provide the result of the user's input by different types of outputs. All the interactions with a multimodal system are directly related to the senses of the human body, in order to recognize natural gestures such as voice, writing, touch, body movements, gaze and lip movements. These systems are much more robust than unimodal systems, as they are only able to interpret one method of interaction and therefore may not obtain the most correct interaction input. For example, uttering a voice command in a noisy environment, when supplemented with other input methods such as writing, can dramatically increase the confidence from the recognized speech command [14] [15].

2.2.2 BACKGROUND

In order to understand what types of interactions are used in multimodal systems, this section presents several multimodal applications, the methods or devices used for detecting the user interactions, as well as the existing frameworks used to exchange messages in a multimodal system.

Laar et al. [16] developed an application named *BrainBrush* that with the use of devices such as Brain-Computer Interfaces (BCI) and Electroencephalography (EEG) wireless devices, allowed users to draw in the computer by reading their head movements, eye blinks, and also to select objects using the BCI device capable of reading P300 cerebral waves. The device used was named "Emotiv EPOC" and captures the three modalities that are used in the application. These kind of devices provide a direct interface between the human brain and a computer without the use of muscles and peripheral nerves, making it possible for disabled people such as amyotrophic lateral sclerosis patients to interact with the computer. In the results obtained in tests made with people with and without disabilities, the order of the modalities from the most used to the least used were respectively the movement of the head, the blink of an eye and finally the selection through the P300 waves. Users characterized the last modality as mentally exhausting, also causing some frustration when not succeeding in the selection task. However, the combination of the three modalities made sense for the users and although there were some difficulties when interacting with the BCI, but in overall it was considered fun and interesting.

Hoste et al. [17] created a multimodal interface named *SpeeG2* to recognize text, in which the speech is used as the main modality for input, and gestures are used to make changes in the results obtained from the voice recognition process. This work was based on other projects that also included a combination of speech input with a different modality used to complement an incorrect word detection, such as *Parakeet* [18] developed for mobile devices and *Speech Dasher* [19]. Unlike these two where speech recognition was made with the CMU Sphinx Toolkit [4], *SpeeG2* uses The Microsoft's Speech Recognition API. This choice was made due to the barrier caused by the CMU Sphinx Toolkit recognizer of requiring an initial training before start using the application, something that was wanted to be avoided. The second recognizer uses a generic recognition model and can easily be used by a greater number of people than the first one. For detecting gestures it was used a *Microsoft Kinect* device. Four different prototypes were created with the following words correction modes:

Scroller Prototype The words move from right to left, and the user by moving hands up and down, must select the correct word column at the position 0. He may also increase and decrease the speed at which the green bar progress at the top moves by bring the hand closer or afar from the body, or go back through the choices made by moving his hand to the opposite side of the body (Figure 2.3).

Scroller Auto Prototype This method is similar to the above, but the green progress bar was removed and the correction is made by the vertical position of a black dot that is controlled by the user's hand.

Typewriter Prototype In this prototype there is no words movement between columns and words are selected with the hand movement of the user who, while moving the hand from left to right will also need to move it up or down, choosing the correct word and going through to the right side of the screen. The weak point of this mode is that the choices made can not be changed.

Typewriter Drag Prototype Created to remove the possible accidental error in the previous prototype, for example when the point reaches the right area and any word was incorrectly selected. In this mode after reaching the area on the right, the user may make a fast move with the hand to the left, as if it were a typewriter, to clear the selection and make a new one.

[4]http://cmusphinx.sourceforge.net

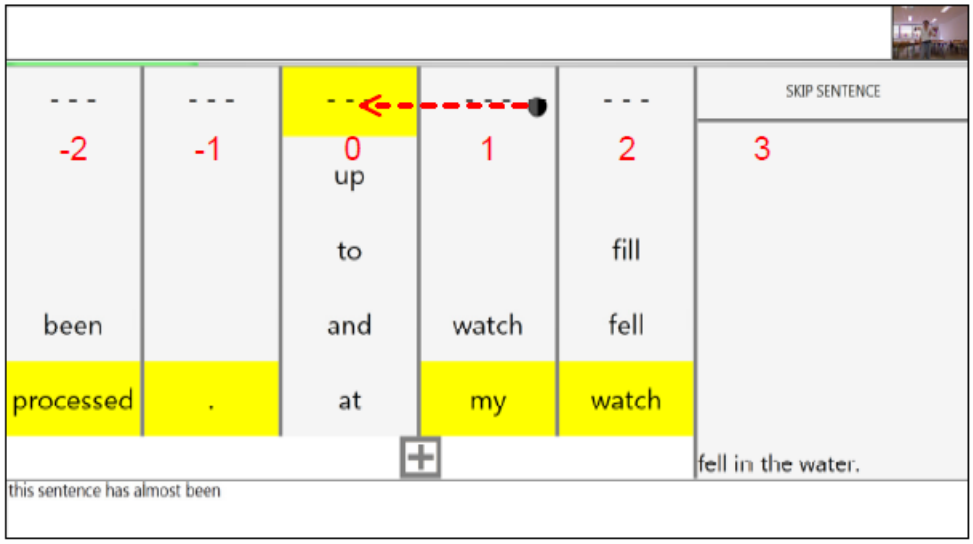


Figure 2.3: Scroller Prototype Interface (Source: [17])

The test results were positive with users preferring the *Typewriter Prototype* method, with an average of 21, 04 words per minute. However one of the recommended improvements by the participants was to add more choices of alternative words to prototypes, but this point was also described as a factor that could reduce and turn more difficult the word readability.

Based in the Multimodal Architecture recommendations by the W3C [20] an IM Framework in Java was developed [21], supporting the communication between other similar IMs and with the available embedded modalities components communicate through HTTP. Some of the projects that included this Framework are the AAL4ALL [5] and Paelife, with the personal assistant AALFred [6]. Its great benefits of it use is the possibility of modularly add new services or new methods in order to attach a new method of interaction to the system.

2.3 MULTI-DEVICE

We can describe such systems as a system including multiple devices and communicating with each other towards the same result or goal. The devices in this context don't need to be of the same type to be considered a multi-device system. With the increasing number of mobile devices usage, their capabilities increasing and their easy portability, these type of devices represent a great beneficial factor for the creation of multi-device systems, with the goal of improving how users interact with other devices around them, and providing solutions that help people get the most out of these technologies [9].

Deborah Dahl, in the W3C Blog [22] describes some of the scenarios or systems that are amenable of being multi-device:

Cars : car's features can be controlled by other devices, such as temperature, radio or GPS;

[5] `www.aal4all.org`
[6] `www.microsoft.com/Portugal/MLDC/paelife`

Houses : home appliances, lights, temperature or any other equipment may be controlled or monitored by a device;

Medical Devices : sensors coupled to a patient may send the data directly to a smartphone or any other device that in turn forwards that information for analysis.

However, multi-device systems must have a method capable of connecting to other devices, by either a physical or wireless connection, to perform the search and detection of a specific device, recognize what kind of device is and which operations it offers. Wiechno et al. [23] labeled Modality Component (MC) as a logical entity responsible for handling the input and/or output, whether voice, writing, video, etc., by devices such as keyboards, microphones, mobile phones, or by software services, and this entity can be associated to a multimodal system. Such entities may also be shaped to complement the functionality of a software on a device. In Figure 2.4 it is shown a MC receiving two different input types. In this kind of scenario, it would also be possible to consider two independent MCs from each other to process the data obtained from a single device. For the construction of a MCs there are a set of requirements defined in order to standardize the messages exchanged, allowing an easier development of components able to dynamically synchronize and communicate between each other. Table 2.1 presents an example of requirements in a use case for smart homes, with their low and high level requirements, and whose audiovisual devices can be considered as an extensions to a smartphone.

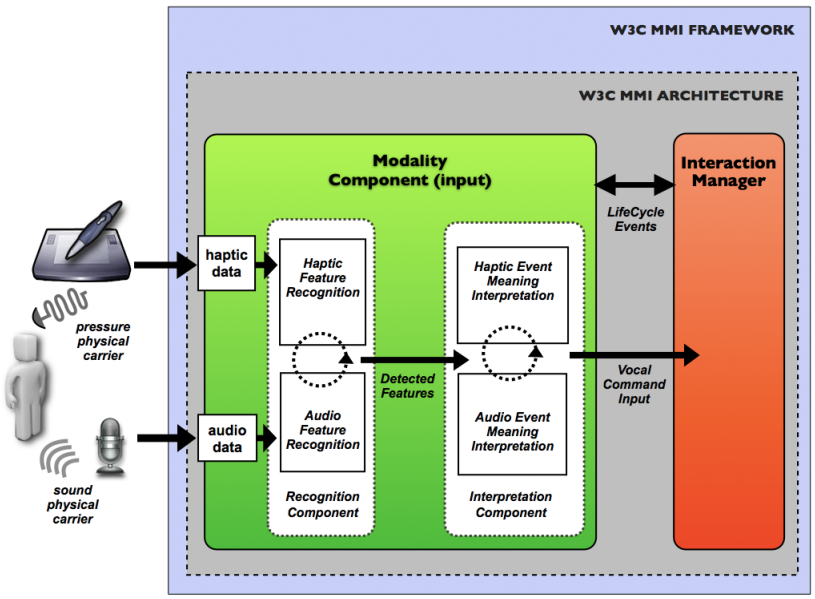


Figure 2.4: A Modality Component (MC) Example (Source: [23])

The integration of multiple devices can be associated with the multimodal framework described in the previous section [21], in which each device can be considered as an independent module. Almeida et al. [9] proposed a solution for a multi-device scenario using the multimodal interaction architecture from W3C (Figure 2.5), where each device runs an instance of the IM presented in previous section [21], and each one of the instances acts as an additional modality component to the remote IMs, thus allowing the share of input and output multimodal events. This multi-device approach was used in the context of the PaeLife project [2], with the development of an application with the current news headlines that interprets the touch, voice and gestures modalities. Two possible scenarios of using the application

Table 2.1: Use cases for audiovisual devices working as smartphone extensions (Source: [23])

Requirements	**Low-Level**	**High-Level**
Distribution	Lan	Any
Advertisement	Devices Description	Modality Component's Description
Discovery	The Application must handle Multimodal Interaction (MMI) requests/responses	Fixed, Mediated, Active or Passive
Registration	The Application must use the Status Event to provide the Modality Component's Description and the register lifetime information.	Hard-State
Querying	The Application must send MMI requests/responses	Queries searching for attributes in the Description of the Modality Component (a predefined MC Data Model needed)

were described, using only one device or two devices simultaneously (e.g.: a television and a tablet). When using only one device, for all user's interactions to communicate with the application (voice, touch gesture, etc.), a message is sent from the modality application that recognized the interaction to the IM which in turn redirects a reply message to the same modality. For example, a voice command recognized by the speech recognition modality results in a message sent firstly to the IM, which in turn delivers it to the application to be processed. In the multi-device scenario each device must run the same version of IM, with an updated version from the single device scenario that communicate with other IM, to share multimodal context between them.

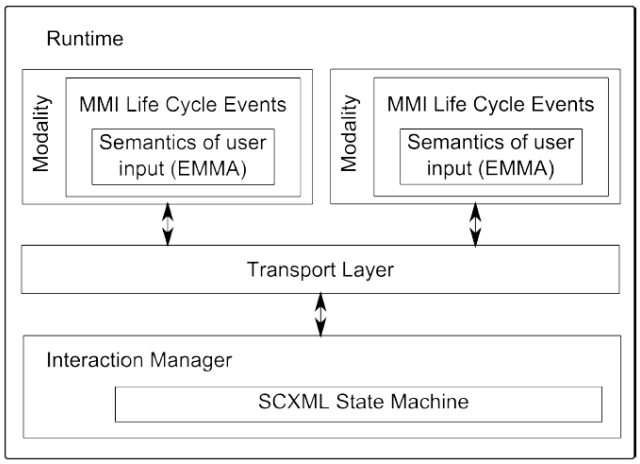


Figure 2.5: The W3C Multimodal Architecture (Source: [9])

The multimodal framework also uses a state machine to control its status, and since the machine is internal to its own IM, it is thus possible, independently of each other, to choose what is displayed in each one of the devices. Therefore, what is presented in one of device's display can be used as a complement of the other by using different views or displaying different information. For example, one device displays the current news photo while the other device only displays the news text. In the event of a device no longer being accessible to the other, both IMs changes a parameter in its state machine

indicating a remote device is not connected and the user may continue to interact with the application through the same device as a single device scenario. A future work for this solution referenced by the authors, was to complement it with a method able to determine if the two applications were able to communicate with each other and only work together when two devices were close to each other.

2.4 EYE-TRACKING

Eye-Tracker is the name of the device that allows the user to interact with a computer or any other equipment using his/her eyes. This type of device include sensors able to register the eye gaze information in order to know the direction of user's gaze, and can be a device coupled directly to a monitor and used as a mounted camera, or used as special wearable glasses able to provide gaze data. The main benefits of using this type of devices are the study of human behavior by observing the user's gaze, but also the possibility of interacting with a device using gaze. Imagining the possible combination of this modality with others such as speech, touch, etc., it turns possible the creation of new and innovative types of interaction. [24]. Figure 2.6 demonstrate how an eye-tracking device works, explaining each step of the gaze processing from the near-infrared micro-projectors and optical sensors contained in the device that create reflections in the user's eyes and capture the required information, to the image and mathematical models that are executed to determine the gaze position on the screen.

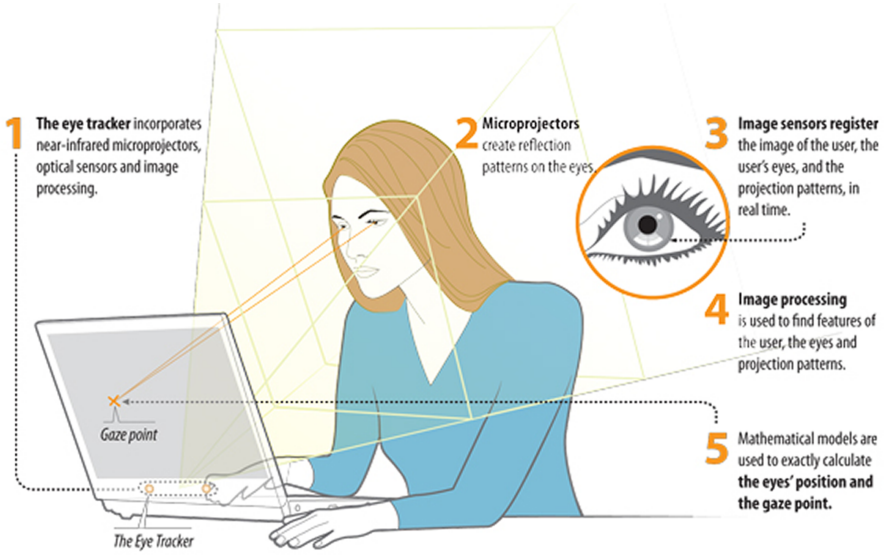


Figure 2.6: Working with an eye-tracking device (Source: [24])

2.4.1 DEVICES

Tobii^{TM}[7] is one of the most known and used brands of eye-tracking devices, offering multiple solutions from hardware to Augmentative and Alternative Communication (AAC) software, for gaming,

[7] http://www.tobii.com

healthcare, research and Human Computer Interaction (HCI). The devices from Tobii ™range from wearable eye-tracking glasses, small portable devices that can be used coupled to a monitor, or even screens with a integrated eye-tracking system.

The Eye Tribe™[8] offers the most affordable eye-tracker in the market. Providing free and open-source SDKs for languages like C#, C++ and Java, this product is the small bridge to take this type of product to common consumers, and allowing developers to easily create new applications or ways of interacting using gaze as a modality.

There are many different eye-tracking devices, and while some may be less intrusive to the user and preferably used for a more natural HCI, others provide higher frequency rates and even more accurate gaze information. This latter group of devices, and due to a more robust technology used in each one of them, have a higher cost price than others and are often only used in research. Table 2.2 presents some devices from the previous brands of eye-tracking devices. The eye-tracker from The Eye Tribe is the easiest to obtain for users, mostly because of its low price. The other devices, although offering higher frequency sampling, have a very high price range and are practically only affordable for research by companies with proper funding.

Table 2.2: Comparing eye-tracking devices specifications and prices

Name	**Company**	**Type**	**Samples Frequency**	**Connection Type**	**Aprox. Price**
Tobii Pro Glasses 2	Tobii	Wearable Glasses	50 Hz	Wireless	11 900 €
Tobii X2-60	Tobii	Small Portable Device	60 Hz	USB 2.0	8 000 €
PCEye Go	Tobii	Small Portable Device	not specified	USB 2.0	1800 €
Tobii EyeX	Tobii	Small Portable Device	60 Hz	USB 3.0	99 €
GP3 Eye Tracker	Gazepoint	Small Portable Device	60 Hz	USB 2.0	450 €
The Eye Tribe Tracker	The Eye Tribe	Small Portable Device	30 Hz/60 Hz	USB 3.0	90 €

2.4.2 EYE-TRACKING APPLICATIONS

Most of the applications from the brands behind the previous or others non mentioned eye-tracking devices are used only for research, reading eye movements with the goal of analyzing the users behavior while performing actions or to evaluate an application usability. As a solo modality in HCI, the eye tracking is commonly used by patients with Amyotrophic Lateral Sclerosis (ALS) or any other type of

[8] `https://theeyetribe.com`

impairment, only able to use gaze to interact with a computer and aid in their communication with other persons. Since the main interest is the use of gaze in HCI context scenarios, the software search was limited to the eye-tracking software used to interact with the computer. As this is a new and emergent type of technology only recently becoming more popular due to the availability of cheaper devices, the available software that uses gaze is either limited to a certain eye-tracker device or is already embedded in a specific system built for impaired users such as ALS patients.

Tobii Dynavox, part of the Tobii Group, developed Windows Control [9], an application that emulates mouse control and provides an on-screen keyboard, and the both options when working together can be used to control a computer only using gaze. Another component of this application is the "Gaze Selection" that displays different types of mouse interactions in a task bar. This way, the user gazes at the desired option, and then to the position that he wants to execute the previously selected mouse operation.

Aside from the commercial products, Sweetland recently released a free and open-source software (2015) [25] that provides the user the ability to fully control a computer using an on-screen keyboard and control the mouse using gaze. This software was designed to use with low cost eye-tracking devices, and it was created with the goal of challenging the high prices required to obtain most of the commercial AAC products.

Samsung Electronics has also been working on a project aimed at people with disabilities that uses eye-tracking as a method to interact with the computer, and was solely composed with company volunteers. Their first project, named EyeCan [10], required the use of custom made glasses to detect gaze movements. The newest updated version uses a specific device that is placed below the monitor like almost all other eye-tracking devices[26]. The software interface offers 18 different commands that requires a blink to select an option while looking to the desired icon.

2.4.3 EYE TRACKING IN HCI STUDIES

Besides the previous sections with more technical details about this type of devices and software, the existing studies related to HCI that use the gaze modality. Therefore, the nowadays available technology and current equipment capabilities benefits researchers seeking better and newer natural forms of interaction.

It is plausible to say in most cases speech is the most preferred way to communicate. If we go back to 1980, the "Put that there" [27] modality system could be interacted using speech and gestures, with the latter used mostly as a pointer. Their main assumption was that Automatic Speech Recognition (ASR) would never be 100% accurate, and explored different types of input that when combined, each one may had redundant information to the other. Nowadays, eye-tracking devices can also be used for pointing, in a more natural and easier way than using gestures. Besides, studies have also shown that gaze and speech are linked during a natural human communication, and when used together they can greatly increase the erroneous confidence obtained when only ASR is used [28, 29].

[9] `http://www.tobiidynavox.com/windows-control/`

[10] `http://www.eyecanproject.org/p/english.html`

2.5 AUTISM SPECTRUM DISORDERS

ASD are the set of neurological disorders characterized by causing difficulties in social interactions, verbal and nonverbal communication problems, and repetitive and common behaviors that are detected during the early years of a child's childhood. The significant communication and interaction differences distinguish the ASDs from the other types of disorders. These disorders can be split in three different groups: Autism, Asperger's Syndrome and Pervasive Developmental Disorder-Not Otherwise Specified, usually abbreviated PDD-NOS. Autism is the most serious case, where a large number of affected patients don't have any kind of verbal expression, have severe motor disabilities and show some indiscipline. Asperger's syndrome patients, referred to as "a mild form of autism" have normal or above average cognitive skills, and may also express certain atypical interests. PDD-NOS is the diagnosis of exclusion that covers cases that do not fit in the other two first types of disorder [30].

Right now there are hundreds or more applications on the market with AAC features [31]. This type of applications is essential for children unable to communicate, and the Picture Exchange Communication System (PECS) is the most widely used system as an alternative method of communication [32], where communication is done with the help of cards with different images, each one with its own meaning. However, the most common problems of this type of system is its lack of portability and organization when the child owns a high number of cards [32], aspects in which some applications for mobile devices are capable of solving, opening different ways to implement new alternatives to assisted communication and with a much lower cost than some type of devices and AAC systems [33].

> ".. such devices are readily available, relatively inexpensive, and appear to be intuitive to operate. These devices also seem to be socially accepted and thus perhaps less stigmatizing when used as assistive technological aids (e.g., as SGDs) by individuals with developmental disabilities."
> Kagohara et al. [4]

2.5.1 SOFTWARE FOR AUTISTIC CHILDREN

Proloquo2go™ [34] is an AAC system developed by *AssistiveWare* for *Apple*'s iOS devices, meant for people with difficulties in verbal communication. Awarded with several prizes in the categories of applications for people with special needs, this is one of the most complete programs to aid in their communication. The application uses a package symbols called *SymbolStixs* [11], but it is also possible to create new symbols by using the device's camera. As defined by PECS each symbol is represented by one picture which can be displayed in a list or a grid. The *Acapela* [12] synthesized voice system is used to reproduce letters and constructed sentences. A major innovation of the *Proloquo2go* is the ability to automatically conjugate verbs and pluralize names 2.7. For example, pressing the symbol of a verb displays other representative symbols of the various verb's conjugations [35].

Buzzi et al. [36] developed a module for the *Drupal* [13] Content Management System (CMS) called *ABCD SW* [14] in order to "facilitate the execution of applied behavioral analysis with low-functioning autistic children", taking advantage of AAC and Discrete Trial Training (DTT), where the responsible

[11] `https://www.n2y.com/products/symbolstix`
[12] `http://www.acapela-group.com/`
[13] `https://www.drupal.org/`
[14] `http://abcd.iit.cnr.it/wordpress/`

Figure 2.7: A verb conjugation in *Proloquo2go* (Source: [35])

tutor can select the type of test to be carried out and their difficulty, and the data of automatically recorded sessions. The *Drupal* was used as the basis for this application as it offers advantages in terms of internationalization and scalability. The system is able to work simultaneously on different devices, a laptop for the tutor and a tablet device for the child, providing the tutor with a real time summary of the actions taken by the child and a simpler and interactive access to the interface. The communication between the two devices is done by placing the exam data in a database and access it by using *Ajax* requests, performed in every second. It has also been equated the usage of *WebSockets* as an alternative, but it was not chosen due to not demonstrate sufficiently stability. In the event of not having two devices available, one for each of the users, an additional tab can also be opened to separately access the child interface. The great advantage of this application described by the authors, it's the usage of a web-browser, which allows its use in various platforms, as well as the ability to control the tests in real time.

Chien et al. [37] created a PECS application for Android tablets named iCAN that, unlike other existing applications such as *Proloquo2go* that supports the communication of autistic children with others, is destined to the child's tutor in order to increase the children motivation to learn and stimulate their senses and communication skills. The application's features were chosen taking into account various problems in the traditional method of using images on physical cards. These problems include the lack of organization of the cards, the difficulty of creating new cards, and the lack of child's interest in using this method. Therefore, the application has been projected to enable the creation and editing of digital new cards to communicate building sentences or phrases that have been re-recorded previously. To create or edit cards the user can draw or select an image, and record the pronunciation of the respective word. While in creation mode the user must drag the desired phrase image for the respective layer in order to produce the desired phrase(Figure 2.8).

The application will read aloud the sentence built after pressing the button to play it. Sentences

can also be saved so that they can be reused again, thus easing the child's expression by using phrases he already knows. To reuse any previous constructed sentences, the user must access the respective mode where all phrases are shown on the left side. After selecting a phrase, their respective images are automatically placed on the track and the user can select how many times he wants the sentence to be read. The application was tested by children aged between 5 and 16 years, with the help of tutors who had previously used PECS.

In the final evaluation questionnaire all the answers were marked as extremely positive, with the exception of the question on the desire to let children use the application without any supervision. The children really demonstrated willingness to continue using the application by themselves, but were rarely left alone when using a device as a precaution, as they could accidentally damage them. Educators have left a positive opinion on improving children's learning ability, as it was especially notable the difference in the will to learn, the difference in the knowledge level obtained that even though different from child to child it was always visible their evolution, and cognitive growth of children while they were learning using the application.

Figure 2.8: iCAN: Images saved by categories. Phrases are build by draggind or selecting images (Source: [37])

Gay et al. [38] created a prototype (*CaptureMyEmotion*) that uses data received from sensors to allow the autistic children to learn and know the emotion (happy, sad, angry,...) they are having at the current time. The two *Bluetooth* sensors are used to measure the excitation (*Affectiva Q sensor*) and the child's stress level (*Zephyr BioHarness*^{TM}), however the application is independent of both sensors and does not need the two to operate, since both sensors are only used because all children have their own expression, and therefore could be happy but without expressing any smile. Thus, the application allows autistic children to take pictures, record a video or sound and describe how they feel. When taking a picture, it is also taken a photograph using the device's front camera to capture the child's face at that moment. In the end, it is presented a list of emotions to the child choose. After this process is completed, the data can be sent to a online repository such as *Dropbox*.

Muñoz et al. [39] state that more and more children with special needs use haptic devices, such as *tablets* and *smartphones*, but these type of devices lack solutions for the training of empathy of autistic children. They decided to create an application, *Proyect@Emociones* [15], with the goal of developing their independence and to increase the children social skills and confidence. Supported by a tutor,

[15] `https://play.google.com/store/apps/details?id=air.Proyectoemociones`

they are faced with different problems and situations. When the answer selected is correct they are presented with an audible and visual signal used to stimulate the child's confidence. A heuristic test and evaluation of the prototype was made to collect its usability issues, with participants highlighting the lack of audio to explain the context of the current activity and the existence of an error message not feasible for the child's learning process due to it express too much negativity. A final version was posteriorly tested by teachers and therapists at a school for autistic children. The application contained 5 activity levels with increasing difficulty, each one with 3 steps, and the child would choose the option that they thought would be correct depending on the type of emotion that was described to him, both by voice and by text. However, they reported that students who participated in the early levels demonstrated some inability to solve the more difficult ones. Based on the observations made during the tests, they concluded that children with lower levels of autism benefit from using the application as they appear to have less issues understanding emotions and feelings than children with more difficulties.

Table 2.3: List of software with methods used in autistic children communication or learning

Application	Languages	Year	Ages	Devices	Multi-Device	Multimodalities	Features
Proloquo2Go	English	2009	All	iPod, iPhone, or iPad	No	No	Phrases construction using PECS
ABCD SW	Multilingual	2012	?	Web	Yes	No	Applied behavior analysis
iCAN	English, Chinês	2014	5-16	Android	No	No	Phrases construction using PECS
CaptureMyEmotion	English	2013	7+	Android	No	Yes (Excitement and Stress)	Emotion capture
Proyect@Emociones	Spanish	2013	8-11	Android	No	No	Empathy training
Autismate [40]	English	2009	All	Android, iPod, iPhone, or iPad	No	No	Phrases construction using PECS, stories, ambient creation, agenda
OlaMundo [41]	English	2014	All	iPad	Yes	No	Remote chat using pictures
Pictello [42]	15 different languages	2011	All	iPhone, iPad	No	No	Story creation
Aaron [43]	English	2013	2-12	Windows Phone	No	No	Basic PECS

2.6 CONCLUSION

In this chapter the background context of this work was presented, with an analysis of personal assistants, multimodal systems, multi-device systems and applications for autistic spectrum patients. We can say that, although there are several applications and methods for the teaching and support of autistic children, few extend beyond the basic modalities that are allowed by the system to which they were developed. Moreover, the applications are focused solely on the development of only a stimulus, such as emotions or social interaction. Although children often require observation and monitoring during the use of devices, in some cases they demonstrate willingness to use them alone. The level of autism varies from child to child, and not all cases require constant monitoring, and the use of certain learning methods could be a great benefit in the child's education.

The applications analyzed and presented in Table 2.3 demonstrate the short supply of applications that provide multiple methods of interaction, but also the lack of applications for the *Windows* or *Windows Phone* operating systems. In addition, many applications are used as an aid in communication, and with little to no diversity of the available languages, which often is limited only to English, hinders its use to people who do not speak the language of the application.

It is evident that in certain cases, the existence of a multimodal solution capable of integrating multiple devices for different types of users, as well as the possibility to use different types of modalities for interaction, may be beneficial to train and stimulate the capacities of an autistic child.

In the next chapter we will present the architecture and methods adopted for the solution of the problem previously described in the Introduction (Chapter 1).

CHAPTER 3

Architecture and Adopted Methods

This chapter describes the architectures used to support the thesis work. Section 3.1 shows the base multimodal architecture of this work and how different modalities and devices are connect between them, including the new eye-tracking modality; Section 3.2 is a brief explanation of the multi-device component and how different devices connect with each other; finally Section 3.3 describes the methodology used during the development of all the work.

3.1 Multimodal framework architecture

Interacting with a system using multiple modalities enriches the way it works, for example, by combining different actions. However, the biggest advantage this type of interaction can bring is to allow people with disabilities to execute certain actions by an alternative modality, enabling them to use the application. The multimodal architecture adopted in the construction of this system is based on the architecture presented by Almeida et al. [9] in Section 2.3, and it enables not only to couple different interaction as component modules to the multimodal framework but also the use of other devices to also work as a component in that same multimodal framework context.

The multimodal architecture, shown in Figure 3.1, displays how the modalities and the two devices are connected within the multimodal framework context. The IM is the application responsible for the process and distribution of interaction commands received from each one of its modules. Each modality or component that is attached to the system has the job to exchange the messages between the input device associated with their functionality and the IM, who therefore delivers the message to the correct destination. Exceptions to this type of interactions are the ones native to the devices where the application is running, such as touch, keyboard or mouse, whose interaction detection is done directly in the application and no message is exchanged between modules and the IM.

The Figure 3.1 displays the existing components in the multimodal framework and how they connect. Each modality is connected directly with the IM, the root of all multimodal communications. When including a new modality component, for example the eye-tracking modality, the device used to

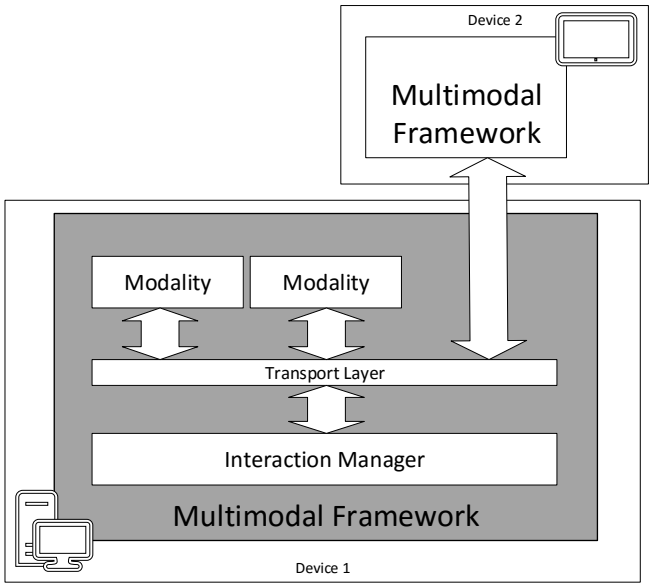


Figure 3.1: Multimodal Architecture

obtain the gaze data is linked to the same device where it is connected. By taking advantage of the modular architecture, the modality has its own independent application that processes all the data obtained from the eye-tracking device and sends all the recognized interactions to the IM. On the other hand, to explore more complex scenarios, we may also have the multimodal framework from Device 2 acting as a modality of the first IM in the Device 1. Because this implementation is performed in separate machines and not locally like the other existing modalities, the existence of another type of communication platform between devices, such as a local area network is required, so that both IMs can exchange messages between them.

3.2 MULTI-DEVICE ARCHITECTURE

As it was identified in the previous chapter, there are many solutions or applications which support the simultaneous and collaborative use by more than one device. The proposed system architecture for a multi-device interaction is based on the same multimodal framework as the one described in Section 3.1, with different devices running the same IM application version. In structural terms, as shown in Figure 3.2, the home network may be used as the platform for the communication between devices.

The component responsible for the multi-device tasks is embedded in the IM. Therefore, when the multi-device use is requested, each IM executes a set of actions that allow the search and pairing of a remote IM. When there is a connection between the two devices, and in order to share multimodal context between devices, each IM replicates all the messages messages to the remote IM, exchanging the interactions made through any modality such as speech recognition or voice synthesis, or by using external connected devices such as an *eyetracker* and a *Microsoft Kinect*. This allows not only the creation of multi-device scenarios with distinct or complement outputs in different devices, but also to send specific application data to enable the collaborative use with a remote device.

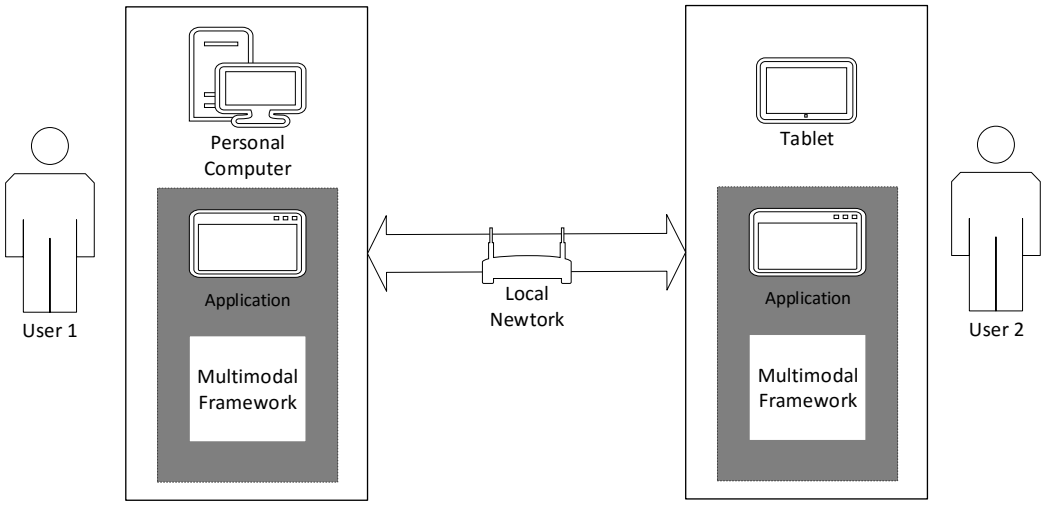


Figure 3.2: Conceptual Architecture

3.3 ITERATIVE METHOD

Before starting working in this matter, it was required to get a better understanding of the current context and roots of work background as there are many partners involved and many different goals. Therefore, the collaboration with external partners was crucial and allowed to realize what are the existing problems in a multimodal interaction and what is the best path to take in order to achieve the desired goals.

For the methodology process of this work it was adopted an iterative process, composed by the following stages:

1. Evaluation of existing problems and creation of scenarios
2. Analysis of studies related to the context
3. Development of a prototype
4. Testing the application(s) with users
5. Data comparison and evaluation
6. Analysis of Iteration's Results

As defined by the iterative method, the stages must be executed in cyclic process and repeated until obtaining a version with a reduced number of issues. For the sake of available time to this thesis only a first iteration of the method was executed, but since this work is currently included as a small part of its research projects, it may be later reevaluated and reanalyzed in order to create a new and updated prototype.

CHAPTER 4

Requirements Analysis

This chapter is composed by the requirements analysis for this thesis, describing the Personas and goals defined for this work. Firstly in Section 4.1, it is explained the method used to elaborate a set of requirements. Regarding the initial studies made for this work, the set of requirements were split in two different groups in order to better comply with capabilities from different Personas in using computing devices and interacting using gaze. Therefore, Section 4.2 presents all the information about the elderly scenario, focusing in the usage of a eye-tracking modality. Section 4.3 describes the autistic child scenario, where two different devices are used and lastly, Section 4.4 presents all the requirements established for the goal of this work.

4.1 Personas, Scenarios, Goals

To define the set of requirements for this dissertation it was adopted the PSG method (Personas, Scenarios, Goals) [44], identifying the characteristics of the Personas group and its relationship with the features of the system, the set of scenarios where their use is possible, and goals in terms of usability, also evaluated for the impact they have during the use of the system.

As stated in the Introduction Chapter 1, this work was built on top of the two projects with different groups of target users, and while the PaeLife was designed for the elderly, the IRIS is more focused in an abroad approach to all the family. Thus, two different Personas, scenarios and goals were created to target each project's context, fulfilling the needs in each of the projects.

4.2 The Elderly

4.2.1 The Persona

Persona is the name given to a hypothetical user, where it is associated with a model describing information about his/her lifestyle, attitudes and the relationship with the product or service [44]. Based on the research conducted for this study and following the same Personas group set to PaeLife

Project [2], the Personas of this scenario have more than 60 years, are retired at home and without any kind of occupation. Although with little computer experience they are able to use some of the common features alone, such as social networking and reading news.

FERNANDO

Table 4.1: Fernando's Persona Description

Fernando		
	Idade	66 years old
	Disabilities	None
	Job	Retired
	Computer Skills	Few
	Preferred Device	Tablet
	Frequency of Use	3 times/day
	Use Duration	2 Hours
	Education	Elementary School

4.2.2 THE SCENARIO: USING THE NEWS MODULE

This scenario explores new types of interaction for the PLA AALFred, using gaze as the only type of input modality or using together with speech. The main purpose of this scenario is to demonstrate that it is possible to control and navigate the news module by only using the new eye modality, but also to use it together with other existing modalities, such as speech.

> Fernando, retired with 66 years, likes to be aware of the news by reading or even by just hearing the reading of the news through his tablet while having breakfast. For this, after opening the PLA he utters "News" to open the news module. The list of news is then displayed on the screen. He also likes to select the news through voice, reading the news name that he wants to open, and although his pronunciation wasn't perfect, the PLA successfully interpreted his command by using his gaze information from the eye tracker. After opening the first news title, and to avoid "talking" to or touching the tablet while he does not end the meal, Fernando uses only gaze to control the navigation between the various news titles, fixating his look into the left or right arrows to navigate between the other news.

Table 4.2: Timeline from Scenario 1

	1	2	3	4	5	6	7
Input	Speech			Gaze+Speech		Gaze	
Output	Open the news module			Open the news details		Navigate to the next news title	
User	Utters "News"	Starts breakfast	Views available news titles	Gazes a news title and utter "Open news"	Reads the news content	Fixates the navigation button	Reads the news content

4.2.3 GOALS

Considering the type of users and usage scenarios, the following goals were established:

1. Create a new eye tracking modality for the modality framework used in AALFred
2. Update the AALFred's news module to allow gaze interaction
3. Add a process of modalities fusion to the modality framework to allow the combined use of modalities such as speech and gaze

4.3 THE AUTISTIC

The work for autistics described in this section was done in collaboration with Ana Leal, a Master's student of Science in Speech and Hearing, to establish a persona, a scenario, and goals that would be valid in a real autistic scenario.

4.3.1 THE PERSONAS

Based on the research conducted for this document, the following Personas for this scenario were defined as the autistic child being the primary persona, and all persons directly connected somehow related with the monitoring and education of autistic children, such as family and educators or teachers as the secondary Personas. The child may have difficulty interacting and communicating with people, which implies some difficulties in teaching and in the family relationship.

NUNO

Table 4.3: Nuno's Persona Description

Nuno Rocha		
	Age	10 years old
	Disabilities	Light ASD
	Job	Student
	Computer Skills	Few
	Preferred Device	Tablet
	Frequency of Use	5 times/day
	Use Duration	3 Hours
	Education	In Special School

4.3.2 THE SCENARIO: USING THE TABLET

This scenario explores the possibility of a child using the tablet device as tool for school learning and to develop his communications skills, that can be used together with another tablet in a multi-device scenario. Also, the parents or any other family person friend may check his Facebook status and view what he's doing, as the children by using a restricted interface to access some of the Facebook functionalities, is able to create posts and reply to his wall.

Nuno just finished his activity in the speech therapy session and wants to take a picture of his work to save and share the moment. When he uses the tablet, the main menu is composed of four options: "Take a picture", "Gallery", "Quiz" and "View my Diary".

Touching the "Take a picture" option, the tablet displays the current view obtained from the tablet's background camera, and after pointing it to the top of the table to capture his work, he presses the button to capture the photo, that it is stored on the device.

Thereafter the application displays the edit menu so that Nuno can select between select an emotion to link it with the current photo, add a comment to the photo or share it in his diary so that Nuno's family and friends can be aware of what he is doing at school. Pressing the first option, 6 different emotions are presented and Nuno's pick the one laughing. Going back, he wants to add a small text explaining what he was doing, and after that he chooses to share it in his diary.

Nuno then goes to the Structured Teaching Classroom for a new activity. He has great difficulty establishing eye contact with others, and the tablet is used as a teaching method and to train his dialogue and communication. In this situation, both Nuno and teacher use different tablets. He access the "Quiz" item at the main menu, while at the same time the teacher setups her tablet and a group of questions to send to the Nuno's tablet. The questions are then made and read by a computerized character shown in his tablet, that selects one of the answers with the help of the teacher. While Nuno is thinking, the teacher may also control the character to make it talk, helping Nuno with the question or even to stimulate his communication.

After finishing the lesson he decides to check on his diary. The photo he previously had shared already had a comment from his mother congratulating him on his work, and he quickly replied to it expressing thanks.

Often, his mother is faced with the dilemma of knowing what his son's homework and how she can help, which leaves her a little anxious. For this reason, the application have a section that allows the communication between the parents and the school.

4.3.3 GOALS

Considering the type of users and usage scenarios, the following goals were established:

- Create two different modules to two different types of user(child and teacher), able to assist in the child's educational process
  - The child module should use some of the existing modalities such as speak and the new eye tracking modality
- Integrate the module in the existing multimodal framework
- Allow the automatic remote IM discovery in the multimodal framework in a Local Area Network
- Permit the exchange of modality command messages between the two devices when used in the multi-device scenario

4.4 REQUIREMENTS

Next we present the set of requirements established for each one of the previous analysis.

THE FIRST SCENARIO

The two main points to meet this scenario goals are the new eye tracking module for the existing framework, and the creation of a method to fusion the new modality with the speech. Therefore:

- The modality should capture all the gaze information from an eye tracking device when the user is looking to the screen and send it to the multimodal framework;
- The modality should be able to receive the information about the current items displayed on the current application interface, along with their multimodal commands;
- The modality should send constant updates to the framework with the gaze information, on whether the user is looking to an item in the interface, or if he's looking at the screen or not;
- The framework should be able to use and join the information received from both the eye tracking and speech modality;
- All previous points should be functional on the AALFred PLA.

THE SECOND SCENARIO

For this scenario, we planned the creation of a modular application for children with some level of autism that is reflected by their lack of social interaction, to fill the low offer of this type of applications in some mobile platforms and operative systems, allowing the use of one or more people simultaneously in two independent devices and adapted to different the different locations of use, such as school and home. The following are thus considered as the main requirements for the application:

- Should be able to take photos;
- Should be able to save all the photos taken;
- Should be able to delete the photos;
- Should be able to view and edit the photos, associating it with an emotion or a comment;
- Should be able to connect to another device;
- Should be able to receive information (questions and emotions) from other devices;
- Should grant an easy and limited access to his Facebook's wall;
- Should be able to share the photos in his Facebook;
- Should allow logging information by others, such as teachers or parents, to allow the trade of information between them;

CHAPTER 5

Development

This chapter describes which were the steps taken to fulfill all the requirements components listed in the previous Chapter 4.

Section 5.1 describes the new eye tracking modality, how it was developed and what are the messages exchanged with the multimodal framework. The Section 5.2 describes the changes on the multimodal framework, for fusing modalities and for the multi-device use. The last Section 5.3, shows the prototypes used to demonstrate all the work on the first Sections of this Chapter.

5.1 The eye-tracking modality

Eye-Tracking, as stated previously in Section 2.4, requires the use of a specific type of device capable of that. For this work and in order to obtain the data for the eye tracking modality, the device chosen was the *EyeTribe* [45]. This device is the least expensive among all the existing ones (view the device's comparison in Section 2.4.1), which makes it the "easiest" obtainable device for this purpose, and although not having the higher frequency like other eye tracking devices needed for a more complex research, it can be used for simple interaction tasks. Also, its working programming library eases the creation of an application that enables the interaction by gaze using the data obtained from the device. The structure of this new modality was aimed to be as independent as possible from this library so that in the future it could be possible to integrate with more libraries to use other different eye-tracking devices.

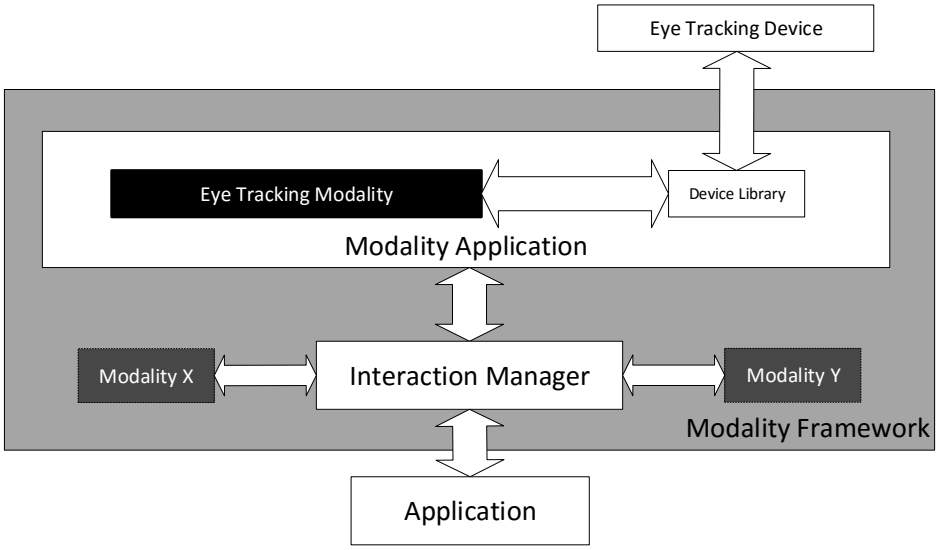


Figure 5.1: The Eye Tracking modality in the modality framework

5.1.1 CONTROLLING THE MODALITY

The new modality, besides reading the information from the eye tracker and send data to the IM, allows the configuration of some of its parameters relevant for interacting using the eye movements. Therefore, and like all the messages exchanged in the multimodal framework, the messages between the modality and the framework IM are based on the W3C Multimodal Architecture [20] [46]. Also, all interactions detected generate a message that is sent from the eye tracking modality, this time based on the multimodal messages like Emma W3C Markup Language [47], which embed specific interaction parameters.

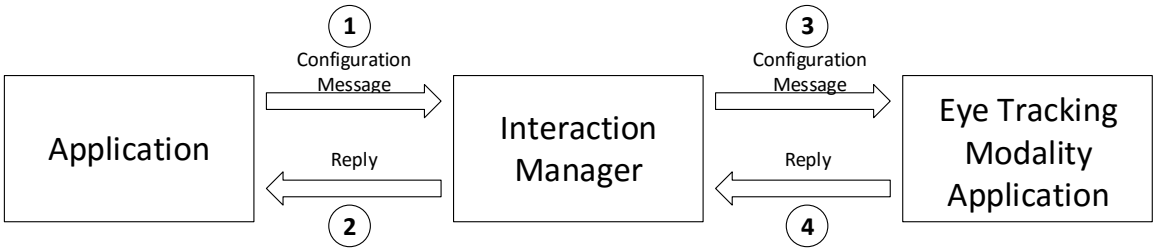


Figure 5.2: Configuring the modality, using the multimodal platform as a bridge for exchanging messages

When exchanging messages to execute any configuration on the modality, the order of actions is shown in Figure 5.2. All the interactions that target the modality are received by using an internal HTTP server running in the modality that processes POST methods. On the other side, when a gaze interaction is detected (see Section 5.1.2), a POST message is sent to the HTTP server running in the IM. These messages should be encapsulated in the same W3C Multimodal architecture used in the multimodal framework, as shown in Table 5.1. To understand how to configure the modality, the following items present and describes all the possible configuration methods available and their respective XML messages:

Configure the modality with the available interface items

When the application changes its interface layout, the modality context must be updated with the current available items the user can interact with. An item should be matched with one of the two different region types available (squared or circular) and configured with its respective screen position and size. As a result of an eventual gaze fixation detected, an item can also have a multimodal command that it is then sent to the IM. The application should add all elements, along with their configurations, and send it inside a XML element 'TrackingGroup'. This element should have a name property to allow later page configurations, and the same name should be used when updating the page layout that is automatically selected as the current page. Example:

```
<TrackingGroup Name="page_name">
        <RadialTrackerObject>
                ...
        </RadialTrackerObject>
        <SquareTrackerObject>
                ...
        </SquareTrackerObject>
            ...
</TrackingGroup>
```

When a gaze fixation is detected in a region common to more than one item, the modality always selects the latter object obtained from the configuration group. Thus, smaller objects that are displayed on top of bigger ones should always be placed after them in the items configuration list.

Select a page

A previously added page can be selected directly without having to update the modality context with the same items. Example:

```
<SelectPage Name="page_name" />
```

Remove a page

A previously added page can be removed when it is no longer needed. Its objects and configurations are also destroyed. Example:

```
<RemovePage Name="page_name" />
```

Clear all pages and objects

The modality context can be emptied and all pages and objects removed without specifying pages with the following message:

```
<ClearAllObjects />
```

Configure the fixation time needed to execute an item modality command

When a fixating hover an item region is detected a multimodal command message is sent to the IM to be saved as the current gaze selection. The IM do not send it to the application when the fixation starts, but only when gaze is kept on the same item after a certain amount of time. To configure the time required to fixate items and its messages be sent to the application, the following element can be used:

```
<SetFixationTimeTrigger Value="seconds" />
```

As stated before, there are two different regions types that can be used to configure the modality and the available fixation zones. Besides the item's center coordinates on the screen, given in pixels (the screen left-top is the (0,0) position), the configuration should tell the item's radius in a circular zone, or the item's width and height in a squared region, also in pixels. The following list presents an example with all the properties available for each region type:

Circular Region

```
<RadialTrackerObject>
        <MiddleX>100</MiddleX>
        <MiddleY>100</MiddleY>
        <Radius>40</Radius>
        <CommandTriggerEnter>
                <string>[ACTION]</string>
                <string>[NEWS]</string>
        </CommandTriggerEnter>
</RadialTrackerObject>
```

Squared Region

```
<SquareTrackerObject>
        <MiddleX>100</MiddleX>
        <MiddleY>100</MiddleY>
        <Width>80</Width>
        <Height>80</Height>
        <CommandTriggerEnter>
                <string>[ACTION]</string>
                <string>[NEWS]</string>
        </CommandTriggerEnter>
</SquareTrackerObject>
```

Table 5.1: Example of a message received in the modality that identifies a zone on the screen, its position and the respective multimodal command.

```
<mmi:mmi xmlns:mmi="http://www.w3.org/2008/04/mmi-arch" version="1.0">
        <mmi:startRequest mmi:source="APP" mmi:target="EYETRACKER"
                mmi:context="ctx-1" mmi:requestID="eyetracker-1">
                <mmi:data>
                        <TrackingGroup Name="page_name">
                                <RadialTrackerObject>
                                        <MiddleX>100</MiddleX>
                                        <MiddleY>100</MiddleY>
                                        <Radius>40</Radius>
                                        <CommandTriggerEnter>
                                                <string>[ACTION]</string>
                                                <string>[NEWS]</string>
                                        </CommandTriggerEnter>
                                </RadialTrackerObject>
                        </TrackingGroup>
                </mmi:data>
        </mmi:StartRequest>
</mmi:mmi>
```

5.1.2 MESSAGES SENT FROM THE MODALITY

The eye tracking device constantly provides the information on where the user is looking at by capturing the user's eye movement. When the modality detects interactions, i.e. the gaze fixating an object on the screen, a process is triggered in order to send a command to the multimodal platform with the parameters relating to the interaction.

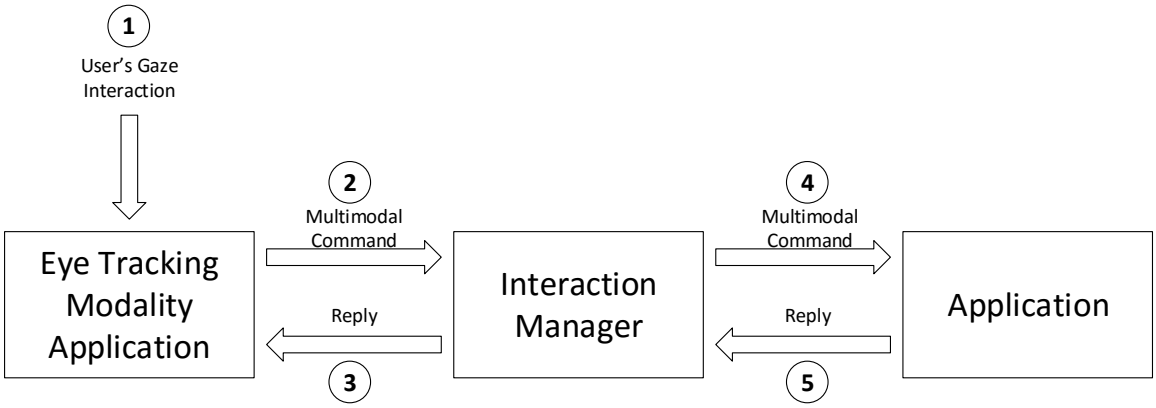


Figure 5.3: Sequence of events and messages resulting from a multimodal interaction triggered by user's gaze

The following list describes all the messages that are originated from a possible gaze interaction detected in the modality. The eye tracking modality's commands use the same format as the ones from the other existing modalities, with a JSON object containing two properties, *recognized* and *text*.

Eyes Detected : When the user's eyes are detected or not by the device, this message is sent to the IM.

```
{ "recognized" : ["ACTION", "EYETRACKER", "ISLOOKING" ],
"text" : "true/false" }
```

Started/Ended Fixating an Object : This message is sent when the user gazes an object on the screen. When gaze is no longer fixating that object, the text value is empty.

```
{ "recognized : ["ACTION", "EYETRACKER", "SELECTITEM" ],
"text" : "-object name-" }
```

Fixation duration : After started gazing an object and if the fixation is kept on the same object's region, in each second a message is sent with the information on how many seconds the user is fixating the object.

```
{ "recognized" : ["ACTION", "EYETRACKER", "FIXATIONTIME" ],
"text" : "-time-" }
```

Actual coordinates : This message contains the information on the current monitor coordinates the user is looking at, and it is used to display some feedback on where his eye position is on the application.

```
{ "recognized" : ["ACTION", "EYETRACKER", "POSITION" ],
"text" : { "PosX": 100.00, "PosY": 200.00} }
```

5.1.3 USING GAZE AS AN INTERACTION MODALITY

The diagrams presented in this section show the events that result from a possible user interaction with the multimodal system using gaze as a single mode of interaction (Figure 5.4) or together with speech (Figure 5.5). In both diagrams it can be seen that prior of all interactions it all starts with the application sending a configuration message to the modality with a list of objects that are presented in the application.

In the first diagram (Figure 5.4), the user interacts with the application only by using gaze. For this, the user fixates his gaze on the desired interface object, i.e. a button, and if the user keeps his eyes position on the same object for the amount of time is was previously configured to trigger this type of interaction (with the command *SetFixationTimerigger*), the modality will send the multimodal command assigned to that object, that was also received in the configuration message.

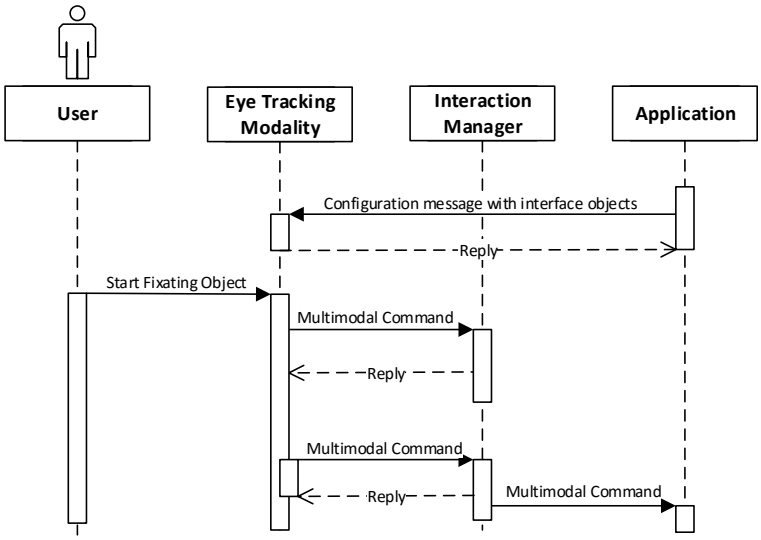


Figure 5.4: Interacting with gaze as the only used modality

In the second diagram (Figure 5.5) the user interacts with the system using multiple modalities. The main objective of this dual use of modalities is to increase the confidence level obtained from the speech interaction, as this type of recognition is easily conditioned with environment elements (i.e noise), as well as the pronunciation of certain words can vary from person to person, leading to an incorrect recognition result or a low result of recognition's confidence. In a real scenario the user would have to look at an object and at the same time utters the voice command to perform the object's action he was looking at. The multimodal platform through current reading of the user's gaze, performs the fusion of both commands to ensure that the resulting multimodal command that is sent to application has a higher confident and is more complete than the speech command. The fusion process is explained with more detail in Section 5.2

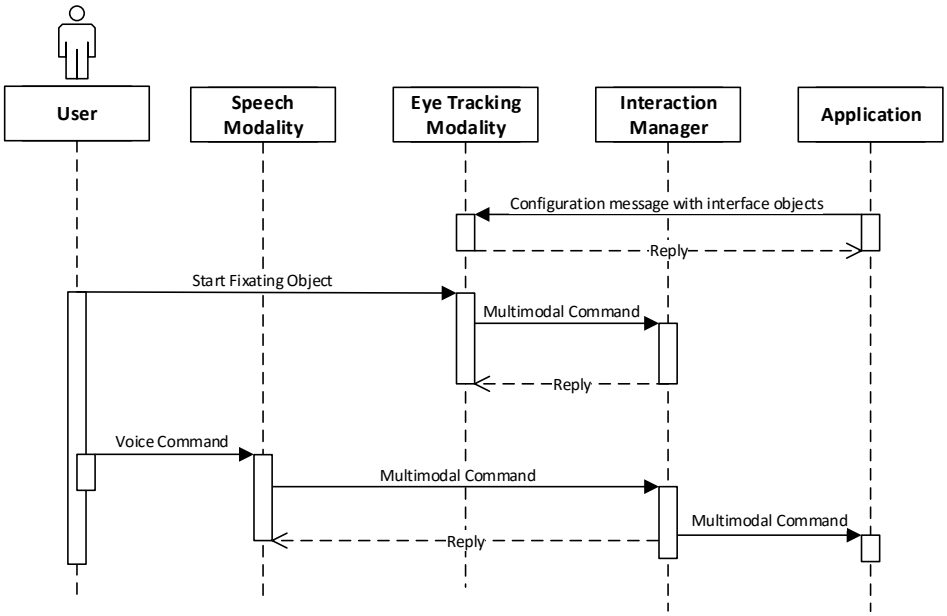


Figure 5.5: Using gaze and speech modalities simultaneously

5.2 MULTIMODAL FRAMEWORK UPGRADES

Despite having a modular architecture, the multimodal framework lacks some processes that allow a more different types of interaction, like the fusion of modalities. Therefore and without changing the current framework architecture, new functionalities were added without disrupting how the existing modules and modalities currently work.

5.2.1 SPEECH AND GAZE FUSION

One of the requirements previously described for this work (see Section 4.4) and already described in this chapter as a possible type of interaction, is the need to fusion of voice modality with gaze. This new functionality was built as a new IM component and embedded directly in its message processing task. The fusion process occurs if after receiving a command with the object he is looking in the interface, the IM also receives a multimodal command from the speech modality, with the latter command being evaluated in order to check whether it is possible to execution a fusion between the two commands. In the example previously shown in Figure 5.5, the fusion process occurred upon the arrival of the speech command, and the order of events that occur is shown in Figure 5.6.

The fusion algorithm is executed when there are two commands in the multimodal platform, one of which originated from the eye tracking modality, and both commands are analyzed in order to increase the multimodal command's confidence by joining the existing information on both commands. The JSON object contained in each multimodal command permit understand what kind of interaction the user wants to execute in the system and the element that the user is interacting with, and the fusion process is based on the analysis of these parameters. With two commands received the algorithm may

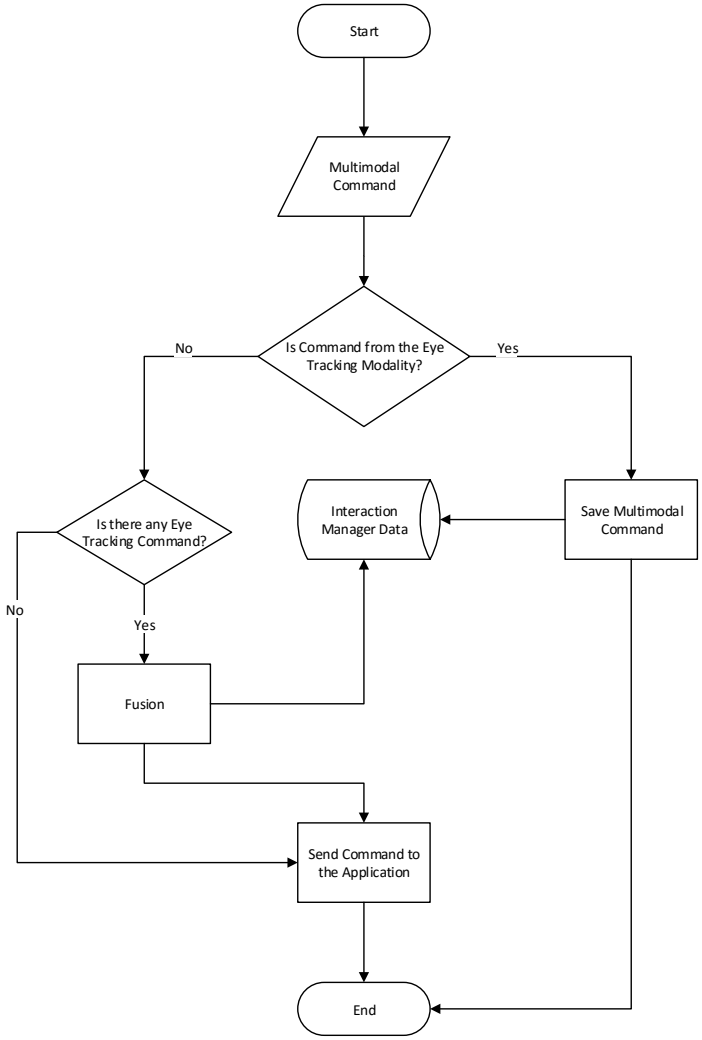


Figure 5.6: Processing sequence occurring upon arrival of a multimodal command to the platform

face two possible situations:

Identical Commands This situation arises when we are faced with two interactions with the exact same command, and the fusion will result in a new command with the same operation but a new confidence recalculated with the average sum of each command's confidence. Example: the user's gaze fixated on a news title and the user utters the respective voice command to open it.

```
Speech Modality:    ACTION NEWS OPEN TITLE X (Confidence: 0.75)
Eye Eracking Modality: ACTION NEWS OPEN TITLE X (Confidence: 1.00)

Fusion Command Result: ACTION NEWS OPEN TITLE X (Confidence: 0.88)
```

Different Commands Both commands are analyzed in order to determine if it is possible to create

a new and coherent command with both the interaction that the user performed. Thus, the fusion is executed when the number of parameters contained in the multimodal command is lower than that the one previously received the type of eye tracking, and their initial parameters are identical. Example: the user gazing on a news title and utter the voice command ”open news”.

```
Speech Modality:       ACTION NEWS (Confidence: 0.75)
Eye Eracking Modality: ACTION NEWS OPEN TITLE X (Confidence: 1.00)

Fusion Command Result: ACTION NEWS OPEN TITLE X (Confidence: 0.88)
```

5.2.2 MULTI-DEVICE INTEGRATION

With two IM instances running in two different devices, a multi-device connection can be used to share different or use complementary types of output modalities by exchanging interaction command messages between each IM. Also, by using the same format for messages used in the multimodal framework, by using a specific set of commands for that purpose, it makes possible to remotely control a application from a different device. As previously stated in Section 2.3, there is already a multi-device proposal for the multimodal framework [9] where the commands received in any IM would also be also sent to the other, but it did not include a process to execute the search for other devices running the same interaction manager, so that they could connect automatically without the need of a previous manual configuration. Therefore, and although for this work the device search was added as a required part of the autistic module scenario, it can also be used in any other multimodal framework multi-device scenario like the one presented by Almeida et al. [9].

The protocol used for automatically search and connect IMs was the Universal Plug and Play (UPnP), and the library used to provide an API for using the protocol was a Java library named Cling [1], supporting the advertise and discover of a UPnP service in a local network using the standard protocol communications. This way, we may use one IM to act as an available service while the other tries to search for the service, and when there is a connection between the two IMs, all the multimodal command messages received in one IM are always replicated to the other.

In order to control the multi-device interaction, the following multimodal commands may be sent from the application to control the device discovery, either for starting the service (server) or to search for remote devices (client):

```
{ "recognized" : ["ACTION", "MULTIDEVICE", "SERVER/CLIENT"],
"text" : "true/false" }
```

5.3 PROTOTYPES

This section describes the prototypes used in the evaluation.

[1] http://4thline.org/projects/cling/

5.3.1 AALFRED BIG EYES

For using the new eye-tracking modality to interact with the application it was required some modifications. As the PLA AALFred was built by multiple partners, with each one contributing with different modules for the assistant, the cooperation from Microsoft MLDC was crucial to make the modality fully operational in AALFred.

As described in Section 5.1, the eye-tracking modality can only be used to interact after the application sends a configuration message to the IM with the elements displayed on the screen. Therefore, every interface update must result in a new configuration message. Also, a new element was added to the interface, giving the user feedback on where he is looking at. Aiming at keeping the interaction natural and instinctive, a small and transparent white circle helps the user to use gaze as a pointer to interact with the application and without causing much disruption.

5.3.2 "CONTA O TEU DIA"

The application for ASD was built on top of the same multimodal framework that was used for AALFred, not only to use the available modalities to provide multiple types of interaction, thus enhancing the application usage, but also to develop a new multimodal application that can be used for ASD children. This group of target users followed the intentions from Iris Project to build a multimodal application targeting different types of users that includes ASD children, and thus our application may be used as a module for it.

Considering the objectives and requirements, the application was built with 4 main functions: a camera to take photos, a gallery to view and edit photos or images, a quiz game to be played with a tutor, and a "diary", a minimalist way to access to Facebook. All the sections were built to have some impact in the teaching and the development of capabilities of ASD, for children with a similar description to the persona in Section 4.3. Although the framework allows the use of different languages for speech recognition and text synthesis, since this was the first developed prototype, we focused only the Portuguese language as the first evaluation would be made only with Portuguese users.

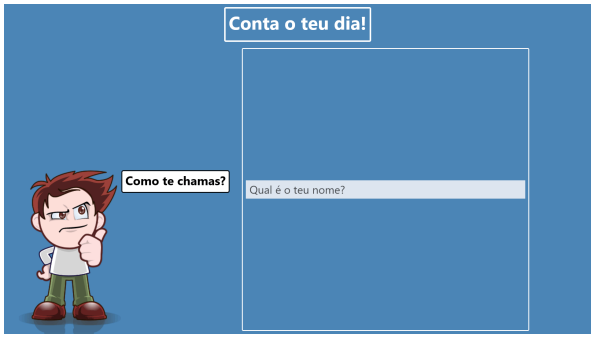
When running the application for the first time, a configuration panel is shown (Figure 5.7) and the children tutor should set the parameters for the application, writing the child name, select whether he may use the Facebook access or not, and what is the login type the child must successfully execute to access the menu. A password may also be set to prevent unauthorized access to this panel the next times accessing it.

Before describing each section of the application, we must note on the character that is displayed on the left side of the interface, because it will be used throughout the application. Instead of using simple output dialog messages, this cartoon represents a small kid that is more visually appealing and is used in an attempt to train the child communication skills by expressing emotions and using the voice synthesis modality to generate the sound of the message also displayed in a speech balloon. As shown later in this section, there's also the possibility of the tutor to take control of this character in a remote device (using the multi-device framework capabilities) in order to create a conversation between the character and the child.

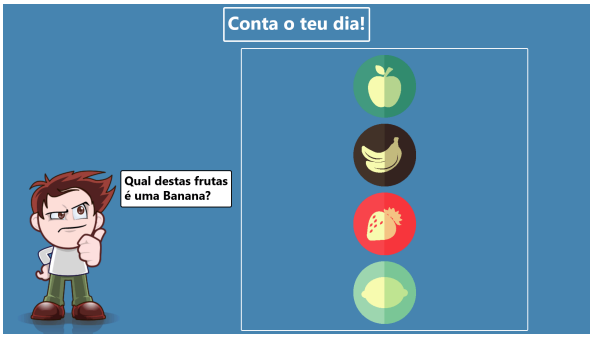
Figure 5.7: The Configuration Panel, where the tutor may set the child name, select the login type and configure the Facebook access.

THE LOGIN

The login used is not used as a method of authentication like normally used, but instead as a method to train children interactions by using touch, speech, gaze, or even using a keyboard to write. The two different type of login screen, displayed at Figure 5.8, allow the interaction with two modalities in each mode. In the first (Figure 5.8a), the character asks the child what is his name. Then, the name previously configured by the tutor on the configuration panel should be typed on the box using a keyboard, or simply using speech and utter the name. On the second screen (Figure 5.8b), a set of four fruits are randomly displayed and the child must select the correct fruit that is asked by the character. In this mode, selecting the correct answer can be also achieved in two different ways, either selecting the correct fruit using touch in a tablet or a mouse when on a desktop, or by looking at the correct fruit. This last type of interaction works when using the eye tracking modality developed in this work.



(a) Login by using speech or typing the name.



(b) Login by touching or looking at the fruit.

Figure 5.8: The two different login modes: the first asks for the child's name, and the second requests to pick the correct fruit.

MAIN MENU

Figure 5.9 displays the application main menu interface, with different buttons routing to each one of the application functions. Along with a text describing the action, each button also include a pictogram image, helping users who may have difficulty reading but also those who already use this method of AAC. Besides, this is one of the most used methods for communication used in ASD persons and is vastly used by other applications that target this same user group. This section of the application also presents the character on the left side that can be controlled by the tutor to interact with the user.

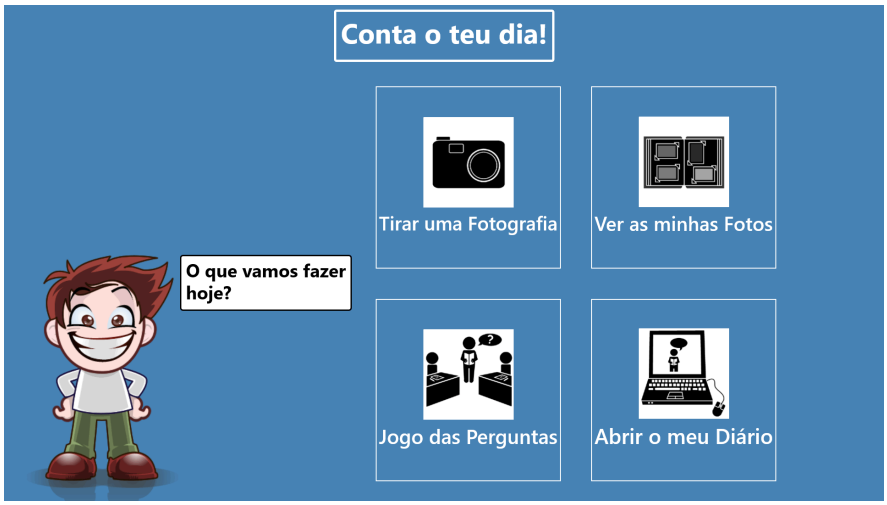


Figure 5.9: The application main menu.

CAMERA AND GALLERY

While using the application either at school or at home, the camera is a very appealing feature that provides children the possibility to take photos of something they saw or of their school work, and all the captured photos can be later viewed in gallery. Figure 5.10a is the interface while displaying the captured scene by the device's back camera. When pressing the shoot button on the bottom of the screen to capture the current preview, the image is stored in the device. The application then displays the photo taken and a edit menu, as shown in Figure 5.10b. This menu can be also accessed by selecting an image in the gallery viewer, and each button contain a pictogram that describes its own action: pick an emotion, edit the photo comment, share the photo in the diary, or simply delete the photo.

EMOTION PICKER

The emotion picker (Figure 5.10c) enables the child to attach a sentiment to the current selected photo. As studies shown, many children have difficulties in expressing their feelings during social interaction. Thus, the inclusion of this section may be useful for developing this topic. The same character is used here to express six different expressions that can be selected - sad, laugh, think, wave, anger, surprise - and are then shown next to the photo in the image gallery, reminding the user how he felt when he took the photo.

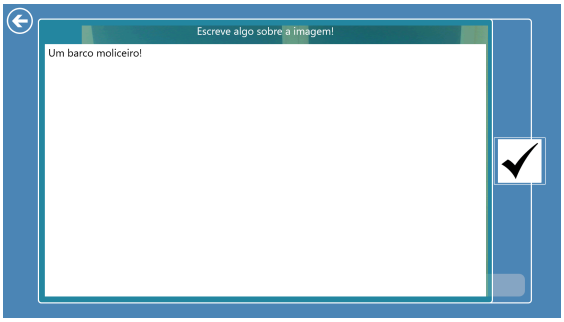
PHOTO COMMENTARY

All the images or photos shown in the gallery can be associated with a commentary, a simple method that can be used to develop the child skills in writing. The text written may also be used in order to describe the photo, so that later he remember what he did in that moment. Furthermore, there is also the possibility to share the photo along with its comment directly in the diary, and its family and friends may see what the child is doing when he is away at school.

(a) The camera allow the child to take photos of his surroundings.

(b) The edit menu, presented after taking a photo or selecting a photo in the gallery.

(c) Selecting an emotion to associate to the current photo.



(d) A comment can be added for the current selected photo.

Figure 5.10: Taking a photo and editing, selecting emotions and adding a comment.

THE QUIZ GAME

Based on the fact that children like games, the goal of this feature is not only to provide a different way to teach but the ability to at the same time develop the child's communication skills. Studies shown that many children have difficulties in making eye contact with other persons, and the immersion with computer and tablet devices may be used as a tool for developing that issue.

The quiz demonstrate the multi-device usage, and it requires the use of two devices simultaneously, one by the child presenting the quiz and the other by a tutor controlling it. The remote device can control the quiz sending questions, or speak with the child using the character. Figure 5.11 displays the screen when a question is presented. The character on the left is the element used to ask the question, and then all the four answers are displayed. After, an answer can be selected either by touching a button or using speech and uttering the selected answer. If the tutor wants to, he can use the character to speak to the children and give tips to help picking an answer.

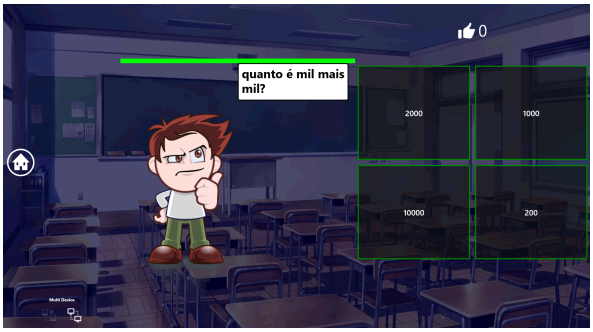


Figure 5.11: The quiz game, with a question currently in display.

DIARY

Allowing the child to share something on a social network let parents and family to remotely keep tracking his actions while he is in school. Considering the child's safety and the risks of children while using an online social network, it was used a minimalist interface that only allow a small set of actions related to the child profile, such as posting or replying with a comment. Nevertheless, regarding the child's security, the parents must keep control of the social network use by accessing the website, only allowing a restricted group of known friends and family see or reply to the child's comments. The diary interface as shown in Figure 5.12, follows the same pattern used in Facebook with a list of posts.

Figure 5.12: The diary displaying the child's facebook wall.

CHAPTER 6

Evaluation Results

An evaluation for each one of the prototypes presented in the previous chapter was conducted to evaluate the usability for the eye-tracking modality and the ASD application, in order to determine the assessment of its use by a group of people. For that, users were asked to perform a set of tasks and to answer a questionnaire and ascertain if the goals previously described in the Requirements (view Section 4.2.3 and Section 4.3.3) were achieved.

6.1 The eye-tracking modality evaluation

For evaluating the gaze modality usability, all the tests were conducted on a calm environment, and all participants used the same desktop computer with a 22 inches screen running the Windows 10 operating system. The device used to capture the gaze context was an "Eye Tribe" eye-tracker. Before starting executing the tasks, and since any had ever used eye gaze as an interaction method for this or other application, participants received a small explanation of how this kind of devices work, and were also told how they could interact with the application, either using speech and gaze. A generic information on how to use the application was also given, so that all would had similar knowledge.

6.1.1 Test description

The participants chosen for this evaluation were 5 adults with no previous experience in gaze interaction but with some experience in using computers. Before starting the test, each participant executed an eye-tracker device calibration. This procedure was made by using the "Eye Tribe UI" application, where users need to keep looking at a small circle on the screen and follow it while it moves around. In order to obtain a better gaze estimation, the highest possible number of positions for the device calibration was used.

The next table (Table 6.1) presents the list of tasks participants were asked to perform, and were split into three sections: on the first, to get used on the current interaction system, participants could only use the speech modality; secondly, being their first experience using eyes to interact and to learn

how the application would react to it, participants could only use the gaze modality; lastly, the two previous modalities could be used together. All tasks were meant to be executed by the users, while the evaluator took notes whether the tasks were successfully fulfilled or not and also the approximate time needed to finish them. All participants started the test with all the framework modalities enabled.

After executing the tasks the participants were given the questionnaire (Figure A.1). With the intention of evaluating their feelings and opinions from the interactions with the application, the questionnaire is composed by questions that primarily target the user's opinion about the gaze modality and compare it to other modalities such as speech.

Table 6.1: The tasks for the eye-tracking modality evaluation.

speech	1. Access the news module using speech 2. Open any news title by uttering the title name 3. Scroll to the next news title using speech 4. Navigate back to the news list using speech
gaze	5. Open a news title by fixating at any news title 6. Scroll to the next news title 7. Navigate back to the news list
both	8. Open a news title by fixating at any item and uttering "open news" 9. Scroll to the next news title by fixating the right and uttering "right" 10. Navigate back to the news list by fixating and uttering "go back"

6.1.2 RESULTS

PERFORMANCE EVALUATION

Before starting the evaluation process, it was decided to set the time required to activate a button using gaze modality to 2 seconds of fixation, as the default time of 5 seconds was found to be very tiring and frustrating when the gaze accuracy was not perfect and harder to use for small buttons. Overall, the participants were all able to easily interact with the application only using gaze. As their first time experience with this type of modality, after finishing the evaluation participants kept using only gaze to interact with the application, and progressively got used to it and easily used and navigated in the news reader module. The evaluation results are shown in Table 6.2.

The first tasks were easily completed by all the participants, and the only issues with using speech were related to the sound capture not detecting the participants voice when uttering too quietly or the recognition engine incorrectly guessing the user's input.

The second group of tasks (5 , 6 and 7) required the use of gaze as a single interaction modality. Participants easily used gaze to open a news item in the first task, but for others some stated the feedback position was not exactly in the same place as they were looking. Thus, the tasks 6 and 7 were those which users have had more difficulty in executing as they required looking and fixating the buttons to navigate right and back. Participants started to demonstrate some frustration by not being able to execute the tasks because the buttons were small and the gaze feedback was not perfect, and asked to skip it.

When returning to tasks with speech, this time to use it together with gaze, participants felt more confident and rapidly opened a news item using only gaze. As the task required the use of both modalities, we asked participants to redo the action. The tasks were similar to the ones used in the gaze only interaction, and again, for the tasks 9 and 10 participants did not try to put a lot of effort

making sure the feedback was hovering the navigation arrows but successfully executed the tasks with the correspondent speech command.

Table 6.2 presents the results from the prototype evaluation. The average time needed to complete the tasks that used gaze is somehow high when comparing to the ones using speech, but after the evaluation process participants completed the same tasks in much less time. Also, participants said that if the application interface was different and with specific layout to help the use of gaze with bigger buttons, more accessible and easier to look at, it would be much more appealing to use with gaze as a main method.

The issues with feedback and gaze accuracy was the greatest concern for participants, and only one rapidly adapted his gaze position to correctly focus the interface buttons to successfully complete the gaze tasks. This mainly could be explained that the participants could have moved during the evaluation and the eye-tracker lost the calibration, or the device used is was the most precise eye-tracker and was unable to answer to this type of use.

After finishing the test, a participant stated he felt that using gaze as a single method to interact with a computer is not necessary even if the eye-tracker had greater precision and the gaze position was detected correctly, unless the user had some kind of disability and could not interact using any other method, because using gaze for a while turns out to be somehow uncomfortable, and while there are other methods available they will always be preferred. The use of gaze was also considered to be restrictive, because if the user moved in front of the device the calibration weakened and the gaze interaction turns harder to use correctly.

Table 6.2: The evaluation of participants performance in the gaze modality evaluation, with the average time from the successfully executed tasks.

	Task Nr.	Success / Failure	Time Average (seconds)
speech	1.	5/0	$9,2 \pm 4,16$
	2.	5/0	$9,2 \pm 1,44$
	3.	5/0	$12,4 \pm 5,44$
	4.	5/0	$20,2 \pm 3,44$
gaze	5.	5/0	$17,4 \pm 4,32$
	6.	1/4	$8,5 \pm 2,5$
	7.	1/4	$12,0 \pm 0$
both	8.	5/0	$24,4 \pm 9,12$
	9.	5/0	$17,2 \pm 4,64$
	10.	5/0	$22,2 \pm 7,12$

QUESTIONNAIRE

The full questionnaire used to evaluate the eye-tracking modality is presented on the Appendix (Figure A.1). The results obtained are shown on the following figures and their results will be analyzed separately.

The first chart (Figure 6.1) depicts the choices on how difficult participant felt to use the modalities to interact with the application, whether if they were easy or difficult to use. Aligned with what was previously described in the performance evaluation section, the values show that users considered speech much more easier to use than gaze, mostly because it was difficult to fixate small icons, as the gaze accuracy was not perfect. Still, if perhaps it was used a more accurate device, the modality would be much more appealing to use.

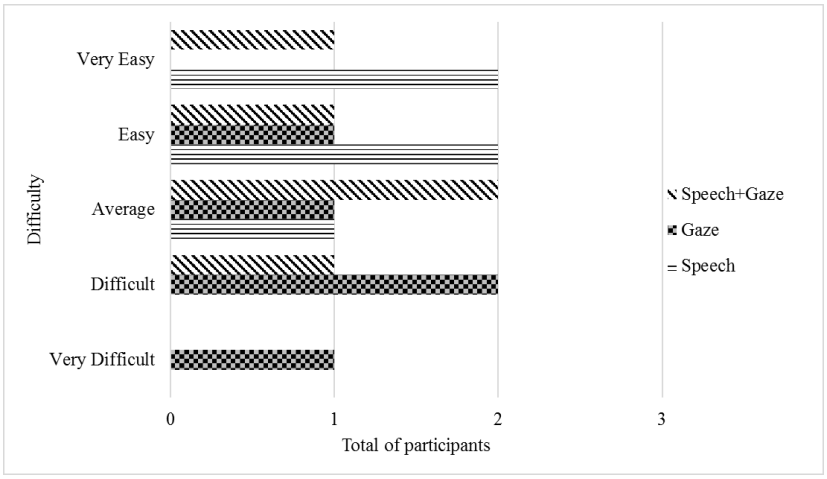


Figure 6.1: The chart with how difficult users felt to use each method of interaction.

The chart from Figure 6.2 show the participants opinions for each modality, and if they liked or not to use it. It can be seen that users preferred most when it was not required the use of gaze. Again, this was because the gaze modality precision was not the greatest and frustrated the participants that were not able to fixate the buttons using gaze and finish the tasks successfully.

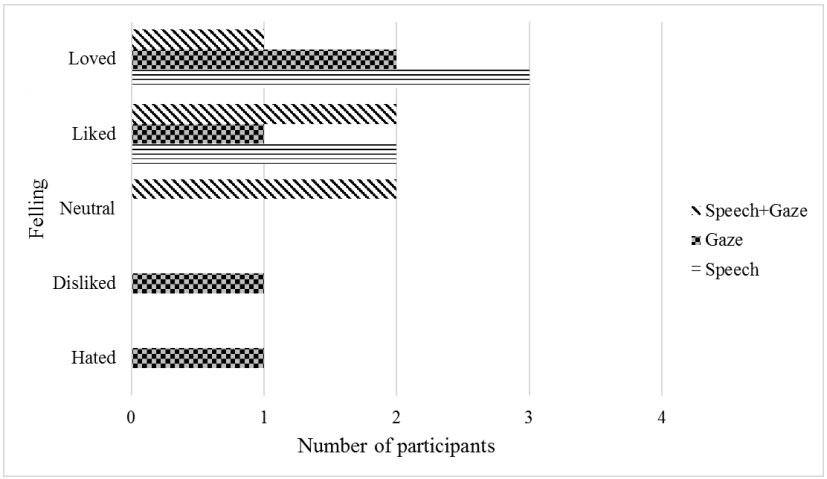


Figure 6.2: The chart with participant's feelings when interacting with the modalities.

The questionnaire also had a section inquiring certain matters relative to the use of the gaze modality, where each statement could be rated with they veracity with a scale of 5 different values, from totally false to totally true. In order to obtain a value for simpler analysis, each option is rated from -2 to 2 and the score for each statement is obtained by the sum's average from all participants option, and the values are presented in Figure 6.3. The values are somehow aligned with the opinions mentioned before, and although participants found interesting the use of the gaze modality, the possible use of it together with speech was considered easier and a more natural way to interact.

Regarding the participants opinions concerning the last question, they mostly enlighten the issues previously depicted when using the gaze modality, stating the need of "a more accurate device or a specific interface", mainly based in the inability to select the small arrows, but also referring to the possible benefits of using gaze and speech simultaneously. The gaze could also be used with eye

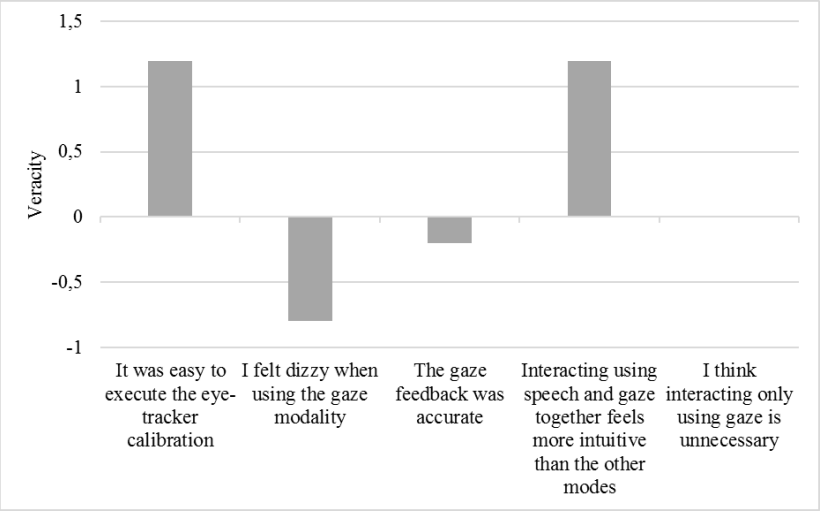


Figure 6.3: The chart with the questionnaire statements.

commands such as blinking instead of fixating interface objects for some time in order to select items or execute any other type of command.

6.2 THE MULTI-DEVICE APPLICATION FOR AUTISTIC

For evaluating the ASD application, all the tests were conducted on a calm environment, and all participants used the same Surface Pro device. Other desktop device, managed by a test operator, was also used in the multi-device tasks. This operator conducted the "tutor" persona, while the participants took the role of the ASD child, similar to the child presented on the scenario described in Chapter 4. Participants also received a small explanation on how the application worked, as this was the first time they were using it.

In order to get a broad set of values from different types of users but also to inquire if the application can be or is relevant for a real use scenario, the participants chosen for this evaluation were 1 child without ASD, 2 special education teachers, plus 2 other persons. This evaluation was made with Ana Leal and the values obtained were also used for her Masters degree thesis [48].

6.2.1 TEST DESCRIPTION

To evaluate the participants performance in their ability to execute the following enumerated tasks for this evaluation (Table 6.3), it was registered for each task the user success or failure while performing, the time needed, the total number in errors and unforeseen events if any occurred.

Table 6.3: The tasks for the ASD application.

1. Log in the application.
2. Take multiple photos.
3. Assign an emotion to a photo.
4. Add a description to a photo.
5. Publish a photograph in the diary.
6. Access the quiz game and answer to a question, using voice as the method of interaction.
7. See the latest publications in the diary and place a comment.

After finishing the test, all participants answered an ICF-US test [49] used to obtain an overall usability evaluation and a PSSUQ test, an instrument used to evaluate the user's satisfaction with the system usability, composed by 19 questions. According to the participants main language, it was used a validated Portuguese version of PSSUQ [50]. Also, the special education teachers answered an additional quiz in order to verify the usefulness of using an improved version of the application in the educational process of children with ASD, as well as which points or functionalities could be added, changed or removed for a future prototype version.

6.2.2 RESULTS

PERFORMANCE EVALUATION

Regarding the participant's ability to successfully complete the previously enumerated tasks, the evaluation results are presented in Table 6.4. The prototype interface can be considered accessible due to the fact that the tasks were successfully executed by the most part of the participants. Only one of the participants had some difficulties performing the tasks 3,4 and 5 and required help, but during the evaluation he stated he had low experience in using tablet devices. The child was the

quickest participant of them all and demonstrated joy when using the quiz game, and after finishing the evaluation, promptly asked to answer more questions.

Still, while performing the tasks some unforeseen events occurred that caused difficulties to the participants, but few of them required the help of a technician to resolve the problem. A participant was not able to login in the application after writing the child's name because a space was mistakenly inserted not allowing to finish the task. Also, all participants in the quiz game task found trouble when using speech to answer to the questions, with promptly uttering the correct answer but the interaction was not being recognized. Therefore, as seen in the Table 6.4, the task number 6 took longer than others to finish because participants had to wait a moment while the speech recognition engine was reloading the new grammar when the application received a new question from the remote device. Apart from the technical problems, some users found difficulties when navigating back and forward between the application sections and recognizing the meaning of pictograms in the buttons which had no caption associated.

Table 6.4: The evaluation of participants performance in the ASD application evaluation.

Task Nr.	Success / Failure	Time Average (seconds)	Number of Errors Average
1.	5/0	6, 6	0, 4
2.	5/0	23, 4	0, 2
3.	4/1	33, 8	0, 6
4.	4/1	47, 4	0, 4
5.	3/2	49, 6	0, 6
6.	5/0	63, 2	0, 4
7.	5/0	51, 4	0, 0

PSSUQ TEST

The PSSUQ items are rated from strongly agree (1 Point) to strongly disagree (7 Points). Therefore, the lower the score the better the participant's overall satisfaction when using the application, with a maximum value of 7. Also, the 19 items can be subdivided in subgroups to rate specific values such as the interface quality (16 to 18), system usefulness (1 to 8) and quality of information (9 to 15). The full PSSUQ questionnaire is presented in appendix Figure A.2.

The following Figure 6.4 presents the values obtained in the evaluation.

Analyzing the previous test results and considering the scale used in the PSSUQ, the questions with the best ratings were "8. I believe I could become productive quickly using this system" and "17. I liked using the interface of this system" with an average rating of $1, 2$ points. The worst questions, thus with a higher value, were "2. It was simple to use this system" and "11. The information provided with this system was clear.", both rated with 3 points. These results suggest that participants were more satisfied with the prototype quality than the easiness and quality of information. Evaluating the PSSUQ subgroups score, the interface quality was the best result with an average of $1, 73$ points. With higher ratings, the system usefulness scored an average of $2, 23$ points whereas the information quality scored $2, 4$ points. In overall, the total average score from all the test items was $2, 15$, indicating a high prototype usability and that participants felt satisfaction while using the prototype.

The test points with the number 9, 10 and 14 were classified as not applicable and were omitted, since the prototype version used in the evaluation did not had any implementation for error message's feedback , or because the item never occurred during any of the evaluations.

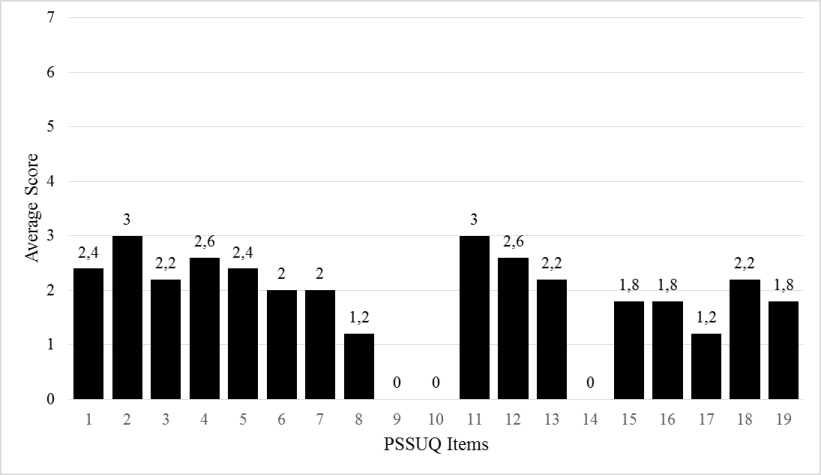


Figure 6.4: Average score values from the PSSUQ test items.

ICF-US I TEST

For each statement of the ICF-US I test, the user must rate whether if he/she considers it a barrier or a facilitator while using the application in evaluation. The test range values goes from -3 points when the item is considered a complete barrier, to 3 points when the user thinks it is a complete facilitator. The final score is then calculated by summing the scores from all the items, and a value above 10 points can assume that the system have a good usability. The results from ICF-US questionnaire items used in the evaluation are presented in Figure 6.5. A full example of the questionnaire is shown on appendix Figure A.3.

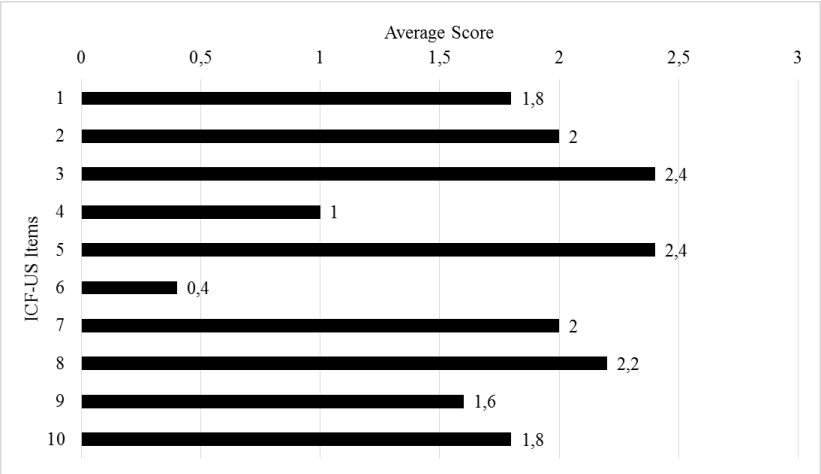


Figure 6.5: Average score values from the ICF-US test items.

From an overall perspective, the prototype had a good reception from participants, and all found the application as a facilitator. Also, the average total score was $17,6$, stating that in general terms the prototype is a facilitator.

Examining each item individually, "1. The ease of learning" and "5. The similarity of the way it works on different tasks" were the items with higher score $(2,4)$, confirming that the use of a simple interface with a similar layout for each page was a good decision. However, "6. the ability to interact in various ways" had the lowest score $(0,4)$, and being a multimodal application this value is somehow

intriguing since it differs greatly from the others. Although the prototype accepts interactions by using the speech modality beyond the regular use of touch or a mouse, the only evaluation task that demanded participants to use speech was the quiz game section. But, and as stated before in the *Performance Evaluation*, almost all participants were not able to use speech and promptly answer the quiz question and finish the task and opted to use touch. Therefore, the difficulty in using speech may be the explanation to the low usability score obtained for the item number 6.

6.2.3 PARTICIPANTS RECOMMENDATIONS

During the evaluation, while observing and by talking to the participants, all the recommendations and personal opinions were registered so that in future works these points can be worked on in order to obtain an easier to use application. Thus, the considered most important suggestions are then listed:

1. The system should also allow the use of a frontal camera when the device contains one;
2. The icons included in the prototype should be more appealing and intuitive;
3. In order to facilitate its recognition, the icons in the edit panel should have captions such as the ones in the main menu;
4. The system should give feedback when an emotion is selected, placing the image corresponding to the emotion selected in the upper left corner of the current photo;
5. After editing a photo that was just captured with the camera and clicking the back arrow, the system should return to the main menu instead of returning to the camera device;
6. When viewing the pictures stored on the device and in addition to the arrow keys, the system should allow swipe to scroll between photos;
7. The time required to use the speech modality to answer a question in the quiz game should be smaller;
8. In the diary, the section for entering comments should be called the "comment" instead of "answer";
9. Accessing the notes panel should be more intuitive;
10. The main menu should contain a button to enable the application shutdown.

CHAPTER 7

Conclusion

7.1 Thesis work analysis

Analyzing the objectives for this thesis, the majority of points were successfully completed, with most effort on the development of a new modality, and the enhancement of the multimodal framework to enable the fusion of modalities and the multi-device detection. Since any other work explored the combination of gaze and speech recognition in European Portuguese within a multimodal framework, these goals may be considered as a small step in order to a more intuitive and complete HCI, and further research on this matter is certainly the right choice. Also, the complementary use of multiple devices in a small multimodal environment is a subject with little exploration on possible and useful scenarios, and the expansive growth of the number of electronic devices used simultaneously in teaching or even at house may turn to be a motivation to captivate more attention to establish goals to explore.

The following list presents the items executed :

- development of a new module for the multimodal platform that enables the interaction using gaze;
- inclusion of a new functionality in the multimodal framework that allows the fusion between the new gaze modality and the existing speech modality;
- embedded the possibility of a multi-device scenario for the existing multimodal framework;
- integration of the new module in the AALFred application;
- development of an application, "Conta o Teu Dia" (Report your day), that being multi-device and capable of multimodal interaction, can act as an assistant in the development and communication of the autistic children, but also explore a different connection between the child's family and his school;

7.2 Future work

From what was collected during all this work process, there are many possible and suitable branches for an evolution of the gaze interaction, but also to explore the multi-device capabilities of

the multimodal framework. Although the results obtained from the gaze modality evaluation were not ideal and its challenges such as the required device calibration or movements restriction turn its use more difficult, its existence is of real importance for a HCI scenario, creating completely different and innovative methods or giving access even for those with more difficulties.

The gaze modality was started for using with just a device, since this was the only device available for testing and develop a prototype. The market of the low cost eye tracking devices is on an imminent grow, thus enabling the use of more devices in the modality could be something to be made in a near future, and assert if the difficulties when using it remain. If not, using speech and gaze together is an interesting point to explore even further to create a more complete method of modalities fusion.

For the multimodal framework, the autonomous search of other devices was a very interesting topic that followed the existing idea of exchanging multimodal context in a multi-device scenario to allow different and complementary outputs. In this work, instead of sharing the context using the same application in two different devices, both devices had different applications but each towards the same objective of working cooperative. The tutor and child scenario does not completely fit in a real life environment, since a school commonly never has only two persons (a teacher and a student) in the same classroom, and this is something that should be accomplished in the future.

References

[1] A. Costa, J. C. Castillo, P. Novais, A. Fernández-Caballero, and R. Simoes, "Sensor-driven agenda for intelligent home care of the elderly", *Expert Systems with Applications*, vol. 39, no. 15, pp. 12 192–12 204, 2012-11, ISSN: 09574174. DOI: 10.1016/j.eswa.2012.04.058. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417412006550.

[2] A. Teixeira, A. Hämäläinen, J. Avelar, N. Almeida, G. Németh, T. Fegyó, C. Zainkó, T. Csapó, B. Tóth, A. Oliveira, and M. S. Dias, "Speech-centric multimodal interaction for easy-to-access online services - a personal life assistant for the elderly", *Procedia Computer Science*, vol. 27, pp. 389–397, 2014, ISSN: 18770509. DOI: 10.1016/j.procs.2014.02.043. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050914000453.

[3] S. Ramdoss, R. Lang, C. Fragale, C. Britt, M. O'Reilly, J. Sigafoos, R. Didden, A. Palmen, and G. E. Lancioni, "Use of computer-based interventions to promote daily living skills in individuals with intellectual disabilities: a systematic review", *Journal of Developmental and Physical Disabilities*, vol. 24, no. 2, pp. 197–215, 2011-10, ISSN: 1056-263X. DOI: 10.1007/s10882-011-9259-8. [Online]. Available: http://link.springer.com/10.1007/s10882-011-9259-8.

[4] D. M. Kagohara, L. van der Meer, S. Ramdoss, M. F. O'Reilly, G. E. Lancioni, T. N. Davis, M. Rispoli, R. Lang, P. B. Marschik, D. Sutherland, V. A. Green, and J. Sigafoos, "Using ipods(®) and ipads(®) in teaching programs for individuals with developmental disabilities: a systematic review.", *Research in developmental disabilities*, vol. 34, no. 1, pp. 147–56, 2013-01, ISSN: 1873-3379. DOI: 10.1016/j.ridd.2012.07.027. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0891422212001941.

[5] S. M. Myers and C. P. Johnson, "Management of children with autism spectrum disorders.", *Pediatrics*, vol. 120, no. 5, pp. 1162–82, 2007-11, ISSN: 1098-4275. DOI: 10.1542/peds.2007-2362. [Online]. Available: http://pediatrics.aappublications.org/content/120/5/1162.

[6] S. E. Levy, D. S. Mandell, and R. T. Schultz, "Autism.", *Lancet*, vol. 374, no. 9701, pp. 1627–38, 2009-11, ISSN: 1474-547X. DOI: 10.1016/S0140-6736(09)61376-3. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673609613763.

[7] Paelife Project, *Paelife: personal assistant to enhance the social life of seniors*. [Online]. Available: http://www.microsoft.com/portugal/mldc/paelife/ (visited on 2015-03-02).

[8] Marie Curie IAPP, *Iris project - towards natural interaction and communication*. [Online]. Available: http://www.iris-interaction.eu/ (visited on 2015-03-02).

[9] N. Almeida, S. Silva, and A. Teixeira, "Multimodal multi-device application supported by an scxml state chart machine", in *Proceedings of the 1st EICS Workshop on Engineering Interactive Computer Systems with SCXML*, 2014, pp. 12–17. [Online]. Available: http://tuprints.ulb.tu-darmstadt.de/4053/1/Proceedingsofthe1stEICSWorkshoponEngineeringInteractiveComputerSystemswithSCXML.pdf\#page=12.

[10] FIPA, *FIPA personal assistant specification.* [Online]. Available: http://www.fipa.org/specs/fipa00083/PC00083.pdf (visited on 2015-02-03).

[11] M. a. Hoyle and C. Lueg, "Open sesame!: a look at personal assistants", *International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM '97)*, no. Paam 97, pp. 51–60, 1997.

[12] M. Huhns and M. Singh, "Personal assistants", *Internet Computing, IEEE*, 1998. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=722236.

[13] H. S. Nwana, "Software agents: an overview", English, *The Knowledge Engineering Review*, vol. 11, no. 03, p. 205, 1996-07, ISSN: 0269-8889. DOI: 10.1017/S026988890000789X. [Online]. Available: http://journals.cambridge.org/abstract_S026988890000789X.

[14] S. Oviatt, "Advances in robust multimodal interface design", *IEEE Computer Graphics and Applications*, 2003. [Online]. Available: http://www.computer.org/csdl/mags/cg/2003/05/mcg2003050062.pdf.

[15] B. Dumas, D. Lalanne, and S. Oviatt, "Multimodal interfaces: a survey of principles, models and frameworks", *Human Machine Interaction*, 2009. [Online]. Available: http://www.springerlink.com/index/65J39M5P56341N49.pdf.

[16] B van de Laar, I. Brugman, and F. Nijboer, "Brainbrush, a multimodal application for creative expressivity", *ACHI 2013, The Sixth International Conference on Advances in Computer-Human Interactions*, 2013. [Online]. Available: http://www.thinkmind.org/index.php?view=article\&articleid=achi_2013_3_40_20409.

[17] L. Hoste and B. Signer, "Speeg2", in *Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13*, New York, New York, USA: ACM Press, 2013-12, pp. 213–220, ISBN: 9781450321297. DOI: 10.1145/2522848.2522861. [Online]. Available: http://dl.acm.org/citation.cfm?id=2522848.2522861.

[18] K. Vertanen and P. O. Kristensson, "Parakeet: a continuous speech recognition system for mobile touch-screen devices", *Proceedings of the 14th international conference on Intelligent user interfaces*, pp. 237–246, 2009. DOI: 10.1145/1502650.1502685. [Online]. Available: http://eprints.pascal-network.org/archive/00005606/.

[19] ——, "Parakeet: A continuous speech recognition system for mobile touch-screen devices", IUI '09, pp. 237–246, 2009. DOI: 10.1145/1502650.1502685. [Online]. Available: http://doi.acm.org/10.1145/1502650.1502685.

[20] D. Dahl, "The w3c multimodal architecture and interfaces standard", *Journal on Multimodal User Interfaces*, 2013. [Online]. Available: http://link.springer.com/article/10.1007/s12193-013-0120-5.

[21] A. Teixeira, P. Francisco, N. Almeida, C. Pereira, and S. Silva, "Services to support use and development of speech input for multilingual multimodal applications for mobile scenarios", in *ICIW 2014, The Ninth International Conference on Internet and Web Applications and Services*, 2014-07, pp. 41–46, ISBN: 978-1-61208-361-2. [Online]. Available: http://www.thinkmind.org/index.php?view=article\&articleid=iciw_2014_2_40_20133.

[22] D. Dahl, *Discovery for multimodal interaction with multi-device systems | w3c blog*, 2012. [Online]. Available: http://www.w3.org/blog/2012/07/discovery-for-multimodal-inter-1/ (visited on 2015-02-10).

[23] P. Wiechno, D. Dahl, K. Ashimura, and R. Tumuluri, *Registration & discovery of multimodal modality components in multimodal systems: use cases and requirements*, 2012. [Online]. Available: http://www.w3.org/TR/mmi-discovery/ (visited on 2015-02-10).

[24] *Tobii: this is eye tracking.* [Online]. Available: http://www.tobii.com/en/about-tobii/what-is-eye-tracking/ (visited on 2015-09-29).

[25] J. Sweetland, *Optikey*, 2015. [Online]. Available: http://www.optikey.org (visited on 2015-10-01).

[26] Samsung Electronics, *Samsung electronics introduces eyecan+, next-generation mouse for people with disabilities*. [Online]. Available: http://global.samsungtomorrow.com/samsung-electronics-introduces-eyecan-next-generation-mouse-for-people-with-disabilities/http://www.theverge.com/2014/11/25/7279849/samsung-eyecan-plus-eye-mouse (visited on 2015-10-01).

[27] R. A. Bolt, ""put-that-there": voice and gesture at the graphics interface", *Proceedings of the 7th annual conference on Computer graphics and interactive techniques - SIGGRAPH '80*, pp. 262–270, 1980, ISSN: 00978930. DOI: 10.1145/800250.807503. [Online]. Available: http://portal.acm.org/citation.cfm?doid=800250.807503.

[28] D. Hakkani-Tur, M. Slaney, A. Celikyilmaz, and L. Heck, "Eye gaze for spoken language understanding in multi-modal conversational interactions", 16th ACM International Conference on Multimodal Interaction, 2014. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=230315.

[29] M. Slaney, R. Rajan, A. Stolcke, and P. Parthasarathy, "Gaze-enhanced speech recognition", in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 3236–3240, ISBN: 978-1-4799-2893-4. DOI: 10.1109/ICASSP.2014.6854198. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6854198.

[30] E. B. Caronna, J. M. Milunsky, and H Tager-Flusberg, "Autism spectrum disorders: clinical and research frontiers.", *Archives of disease in childhood*, vol. 93, no. 6, pp. 518–23, 2008-06, ISSN: 1468-2044. DOI: 10.1136/adc.2006.115337. [Online]. Available: http://adc.bmj.com/content/93/6/518.short.

[31] M. Armanda Quintela, M. Mendes, and S. Correia, "Augmentative and alternative communication: vox4all® presentation", in *Information Systems and Technologies (CISTI), 2013 8th Iberian Conference on*, Lisbon, 2013, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6615718.

[32] H. Sampath, B. Indurkhya, and J. Sivaswamy, *Computers Helping People with Special Needs*, K. Miesenberger, A. Karshmer, P. Penaz, and W. Zagler, Eds., ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012-07, vol. 7383, pp. 323–330, ISBN: 978-3-642-31533-6. DOI: 10.1007/978-3-642-31534-3. [Online]. Available: http://dl.acm.org/citation.cfm?id=2363956.2364009.

[33] Rehabilitation Engineering Research Center on Communication Enhancement, *Mobile devices & communication apps*, 2011. [Online]. Available: http://aac-rerc.psu.edu/index.php/pages/show/id/46.

[34] AssistiveWare, *Proloquo2go : symbol-based aac for ios | assistiveware*. [Online]. Available: http://www.assistiveware.com/product/proloquo2go (visited on 2015-02-24).

[35] S. Sennott and A. Bowker, "Autism, aac, and proloquo2go", *Perspectives on Augmentative and Alternative Communication*, vol. 18, no. 4, p. 137, 2009-12, ISSN: 1940-7475. DOI: 10.1044/aac18.4.137. [Online]. Available: http://sig12perspectives.pubs.asha.org/article.aspx?articleid=1765980.

[36] M. C. Buzzi, M. Buzzi, D. Gazzé, C. Senette, and M. Tesconi, "Abcd sw", in *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility - W4A '12*, New York, New York, USA: ACM Press, 2012-04, p. 1, ISBN: 9781450310192. DOI: 10.1145/2207016.2207037. [Online]. Available: http://dl.acm.org/citation.cfm?id=2207016.2207037.

[37] M.-E. Chien, C.-M. Jheng, N.-M. Lin, H.-H. Tang, P. Taele, W.-S. Tseng, and M. Y. Chen, "Ican: a tablet-based pedagogical system for improving communication skills of children with autism", *International Journal of Human-Computer Studies*, vol. 73, pp. 79–90, 2015-01, ISSN: 10715819.

DOI: 10.1016/j.ijhcs.2014.06.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S107158191400086X.

[38] V. Gay, P. Leijdekkers, J. Agcanas, F. Wong, and Q. Wu, "Capturemyemotion: helping autistic children understand their emotions using facial expression recognition and mobile technologies", 2013. [Online]. Available: https://domino.fov.uni-mb.si/proceedings.nsf/0/5aabd320c05ae58dc1257b8a00337d03/$FILE/32_Gay.pdf.

[39] R. Muñoz, T. Barcelos, R. Noël, and S. Kreisel, "Development of software that supports the improvement of the empathy in children with autism spectrum disorder", in *Proceedings - International Conference of the Chilean Computer Science Society, SCCC*, 2013, pp. 223–228, ISBN: 9781479929375. DOI: 10.1109/SCCC.2012.33.

[40] SpecialNeedsWare, *Autismate365 a comprehensive app for autism*. [Online]. Available: http://autismate.com/AutisMate-Comprehensive-App-For-Autism/ (visited on 2015-03-13).

[41] OlaMundo, *Ola mundo | near or far, always close*. [Online]. Available: http://www.olamundo.com/ (visited on 2015-03-13).

[42] AssistiveWare, *Pictello : visual social stories creation for ios | assistiveware*. [Online]. Available: http://www.assistiveware.com/product/pictello (visited on 2015-03-13).

[43] TeckieGirl, *Aaron*. [Online]. Available: https://www.autismspeaks.org/autism-apps/aaron (visited on 2015-03-16).

[44] M. Aoyama, "Persona-scenario-goal methodology for user-centered requirements engineering.", in *15th IEEE International Requirements Engineering Conference, RE 2007, October 15-19th, 2007, New Delhi, India*, 2007-01, pp. 185–194. [Online]. Available: http://www.researchgate.net/publication/221222470_Persona-Scenario-Goal_Methodology_for_User-Centered_Requirements_Engineering.

[45] The Eye Tribe, *The eye tribe*. [Online]. Available: http://theeyetribe.com/ (visited on 2015-05-27).

[46] W3C, *Multimodal architecture and interfaces*. [Online]. Available: http://www.w3.org/TR/mmi-arch/ (visited on 2015-05-28).

[47] ——, *EMMA: extensible multimodal annotation markup language*. [Online]. Available: http://www.w3.org/TR/emma/ (visited on 2015-05-28).

[48] A. Leal, "Contributos para o desenvolvimento de aplicações destinadas a crianças com perturbação do espectro do autismo", Master's thesis, Universidade de Aveiro, 2015.

[49] A. I. Martins, A. F. Rosa, A. Queirós, A. Silva, and N. P. Rocha, "Definition and validation of the icf – usability scale", in *6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, 2015-06. [Online]. Available: http://www.researchgate.net/publication/281204129_Definition_and_Validation_of_the_ICF__Usability_Scale.

[50] A. F. Rosa, A. I. Martins, V. Costa, A. Queiros, A. Silva, and N. P. Rocha, "European portuguese validation of the post-study system usability questionnaire (PSSUQ)", in *2015 10th Iberian Conference on Information Systems and Technologies (CISTI)*, IEEE, 2015-06, pp. 1–5, ISBN: 978-9-8998-4345-5. DOI: 10.1109/CISTI.2015.7170431. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7170431.

Appendix

The information contained in this appendix was considered somehow relevant to the main subject matter, therefore included in a separate section for consulting.

1) Rate your interaction with the application with each of the following modalities:

	Very Difficult	Difficult	Medium	Easy	Very Easy
Speech					
Gaze					
Speech + Gaze					

2) Rate how you felt when using the following modalities:

	Hated	Disliked	Average	Liked	Loved
Speech					
Gaze					
Speech + Gaze					

3) Rate the following statements accordingly to their veracity:

	Totally False	False	Neutral	True	Totally True
It was easy to execute the eye-tracker calibration					
I felt dizzy when using the gaze modality					
The gaze feedback was accurate					
Interacting using speech and gaze together feels more intuitive than the other modes					
I think interacting only using gaze is unnecessary					

4) Please describe the parts, if any, that you most liked, didn't liked and would like to see when using the gaze modality:
(Free answer question)

Figure A.1: Questionnaire for the eye-tracking modality usability test

	strongly agree ... strongly disagree
1. Overall, I am satisfied with how easy it is to use this system	
2. It was simple to use this system	
3. I could effectively complete the tasks and scenarios using this system	
4. I was able to complete the tasks and scenarios quickly using this system	
5. I was able to efficiently complete the tasks and scenarios using this system	
6. I felt comfortable using this system	
7. It was easy to learn to use this system	
8. I believe I could become productive quickly using this system	
9. The system gave error messages that clearly told me how to fix problems	
10. Whenever I made a mistake using the system, I could recover easily and quickly	
11. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear	
12. It was easy to find the information I needed	
13. The information provided for the system was easy to understand	
14. The information was effective in helping me complete the tasks and scenarios	
15. The organization of information on the system screens was clear	
16. The interface of this system was pleasant	
17. I liked using the interface of this system	
18. This system has all the functions and capabilities I expect it to have	
19. Overall, I am satisfied with this system	

Figure A.2: PSSUQ Evaluation Items used in the ASD application evaluation.

	barrier ... facilitator
1. The ease of use	
2. The degree of satisfaction with the use	
3. The ease of learning	
4. The obtain obtainment of expected results (e.g. I wanted to write a text and I did)	
5. The similarity of the way it works on different tasks (e.g. to confirm an action is always equal)	
6. The ability to interact in various ways (e.g. keyboard, touch or speech)	
7. The understanding of the messages displayed (e.g. written or audio)	
8. The application responses to your actions	
9. The knowledge of what was happening in the application during its use	
10. Overall, I consider that the application was	

Figure A.3: ICF-US Evaluation Items used in the ASD application evaluation.