University Politehnica of Bucharest

Automatic Control and Computers Faculty

Computer Science and Engineering Department

PHD THESIS SUMMARY

Multimodal Interfaces for Active and Assisted Living Systems

Thesis Supervisor:

Prof. Dr. Eng. Adina Magda FLOREA

Author:

Imad Alex AWADA

Bucharest, 2019

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
AAL	Active and Assisted Living
AmI	Ambient Intelligence
API	Application Programming Interface
ASR	Automatic Speech Recognition
AU	Action Units
CAMI	Companion with Autonomously Mobile Interface
CAMI	Companion with Autonomously Mobile Interface
DM	Dialog Management
FACS	Facial Action Coding System
GPS	Global Positioning System
GUI	Graphical User Interface
HMI	Human-Machine Interaction
ICT	Information and Communication Technology
JSON	JavaScript Object Notation
K-NN	K-Nearest Neighbors
NLP	Natural Language Processing
NLU	Natural Language Understanding
RFID	Radio Frequency Identification
SDK	Software Developer's Kit
SVM	Support Vector Machine
SVR	Support Vector Regression
TTS	Text-to-Speech
UPB	University Politehnica of Bucharest
VGB	Visual Gesture Builder

Introduction

Multimodal interfaces offer multiple ways of interaction between the users and the systems while at the same time making the interactions more natural for the users and, consequently, easier and more attractive. These interfaces contribute to dissolve the barriers between technology and non-technical users or users with special needs such as the elderly users.

In addition to the traditional ways of interaction, such as the interactions with the system through a graphical user interface by using a mouse and/or a keyboard, multimodal interactions allow the user to interact with the systems through natural ways of interaction that are based on speech, touch-based gestures and touch-free gestures as well as any other modality which comes natural to human users.

Motivation

The United Nations World Population Ageing 2017 report illustrates that the number of persons above 60 years of age has reached 962 million persons in 2017. This is more than twice than the number registered in 1980 and this number is expected to double by 2050, to reach 2.1 billion or nearly [1]. The report also illustrates that the number of elderly people that live independently at their residence is increasing.

The growing number of elderly people is accompanied by a massive pressure on the different social services and healthcare systems, which are already hustling with challenges such as the costs of the health and well-being services as well as the availability of professional personnel. To face these challenges, caregivers, health professionals, policy makers and computer engineers are looking for new solutions that enable elderly people to live at their residence with an acceptable degree of independence while ensuring their quality of life and their safety. One of the envisaged solutions is offered by the so called Active and Assisted Living (AAL) systems.

To make the interaction with an AAL systems accessible and much easier for elderly people as well as to maximize the benefits of these systems, any AAL system should integrate a multimodal interface. The multimodal interface plays an essential role in determining the acceptance of the system by the users, by consequence the success of the system. Since each user of the AAL system comes with different backgrounds, budgets, knowledge and preferences, as well as the possibility that each user uses the system for different reasons and in different environment conditions, it is important that the multimodal interface integrates some adaptive features (automatic adaptive capabilities and customizable features).

Due to the huge need for AAL systems and the crucial role of multimodal interfaces, the Ambient Intelligence (AmI) and the multimodal interaction fields are currently highly active research fields. They receive a lot of attention from both the academic and the industry sectors.

For all those reasons the work presented in the thesis addresses the interaction issues between the elderly people and technology, and proposes a multimodal interface that integrates different adaptive capabilities, for an AAL system which targets the elderly people as its main users. as its main users.

Objectives

The aim of the research was to identify the best modalities adequate for elderly people interaction with computers and/or mobile devices, to identify good design practices in creating multimodal interfaces with adaptive features, and to develop a multimodal interface that can be integrated into an active and ambient assisted living system.

Specifically, this interface was developed for CAMI, a system for assisting users with special needs in their daily life, including interactions with a humanoid robot, but can be easily integrated in any other AmI application.

The proposed multimodal interface is composed of four modules: the graphical user interface (GUI), the voice module, the gesture module and the emotion module. The voice and the emotion modules are formed from five and two components respectively. The gesture module is composed of two submodules: the touch-based gesture and the touch-free gesture submodules. Each submodule is composed of two components.

While the graphical user interface and the gesture module work without the presence of an active internet connection, the voice and the emotion modules have two modes of working: Online, which requires the presence of an active internet connection, and Offline, which does not require the presence of an active internet connection.

To give the user the flexibility to smoothly interact with the system from any device that the user already has or prefers to use, the multimodal interface works independently of any device or any operating system.

The design of the multimodal interface observed the guidelines for the user interface design and its design, and all of its features and commands are consistent across the different devices and system modules.

The multimodal interface is multilingual and integrates different customizable features and automatic adaptive capabilities, such as the adaptation to the profile of the user (e.g. user's preferences, health issues, etc.), to the activity of the user, to the emotional status of the user or to the configuration of the system, as well as to other parameters.

Thesis Structure

The thesis is structured in four main parts:

The first part of the thesis is a literature review and it is composed of two chapters. Chapter 1 makes a general overview about the ambient intelligence field, the active and assisted living sector and the multimodal interaction field. Chapter 2 analyzes the works that are related to the human-machine interaction field and the active and assisted living sector.

The second part of the thesis introduces the CAMI system and presents a general overview about the multimodal interface process design. It is composed of two chapters. Chapter 3 presents a short overview of the CAMI system and highlights the different multimodal interaction facilities of the system. Chapter 4 delivers an overview about the design of the multimodal interfaces and illustrates the challenges that should be taken into consideration during the design of a multimodal interface for elderly people, as well as the ergonomics guidelines that should be observed during the design of an interface. It presents a model for the multimodal interactions.

The third part of the thesis presents the design and the implementation of the multimodal interface and its modules. It is composed of two chapters. Chapter 5 delivers an overview about the design of the CAMI multimodal interface. It illustrates the different functionalities of the interface, as well as the ergonomics and the adaptive features of the interface. It presents the architecture of the interface and illustrates the design of each module of the interface. Chapter 6 delivers an overview about the implementation of the multimodal interface and illustrates the implementation of its different modules, aspects and features.

The fourth part of the thesis presents the different experiments and evaluation of the interface together with the obtained results, as well as the conclusions and the planned future works. Chapter 7 describes the different experiments and evaluation that took place and illustrates the obtained results. Chapter 8 concludes the work presented in the thesis and presents the different conclusions of the work and lists the envisaged future work.

Chapter 1

Problem Definition

This chapter presents a general overview about the Ambient Intelligence field, about the Active and Assisted Living sector and about the multimodal interaction field.

Ambient Intelligence refers to a world in which computing devices are embedded everywhere into the environment in a transparent way and form a network of hidden intelligent devices that are capable to recognize and interpret human events, to communicate with each other (and/or with the user), and to adapt the environment to the inhabitant's needs, habits, gestures and emotions.

Therefore, any Ambient Intelligent system should be capable to deal simultaneously with a big number of devices and to process in real time a huge amount of data. Meanwhile, it should work without much intervention from users, it should interact with users in the most natural way possible and it should be capable to take some autonomous context-based decisions. To achieve such a complex goal, an AmI system, is particularly identified by the following characteristics: sensitive, responsive, adaptive, transparent, ubiquitous and intelligent.

The applications of AmI are so diverse ranging from sensors and wireless communications to personal health care and new human-machine interactions. Some AmI areas of applications are: domotics or home automation, health care, work, sports, tourism, cultural heritage, vehicles, transports and human-machine interactions.

In the health care domain, AmI offers numerous solutions that not only improve health care but it also changes the way on how health care is practiced. Active and assisted living systems are one of those solutions.

Active and assisted living is a concept in which solutions that include one or more Information and Communication Technology (ICT) components help elderly people live with a degree of independence in their preferred environment for as long as possible [2]. Those solutions are of direct value for elderly people, their families and their caregiver to improve their health and well-being as well as their independence [3]. In addition to the previous benefits, the AAL systems play an important role in reducing the number of the solicitations toward the health care systems. According to the AAL Association, the AAL sector covers eight service areas: health and care, information and communication, living and building, safety and security, mobility and transport, vitality and abilities, leisure and culture, and lastly, work and training.

An AAL system integrates mainly multiple health services that allow the system to monitor the health parameters of the user, to manage the medication of the user, to monitor the emergency situation (such as the falls-cases situations), to continuously monitor the cognitive and physical status of the user, to communicate the health data of the user to a caregiver or health professional, to assist the user during the rehabilitation program and to provide real-time assistance for the user in case of need, as well as other health related services. Usually, an AAL system integrates, in addition to health services, a navigation service (based on GPS and/or RFID tags) and multiple home automation services that allow the system to monitor the status of the windows/doors, to monitor the air quality of the environment, to monitor the light level of the environment and to adjust the light level in the environment, as well as other related services.

Since the elderly people are the main targeted users of the AAL systems, the system should interact with them in the most natural way possible. Multimodal interfaces support multiple ways of interaction between the user and the system such as speech, gesture and touch interactions, in addition to those supported by the traditional interfaces. In those interfaces, the traditional input and output media, such as keyboards and mice are disappearing and are being replaced by devices that allow more natural and direct communication between the user and the system. Multimodal interfaces make the interaction, flexible, easier and faster for the users. They allow the user to combine multiple input and output ways or to alternate between them, as well as they are effective in preventing and handling errors.

Chapter 2

STATE OF THE ART

This Chapter analyzes the works within the human-machine interaction (HMI) field and the active and assisted living sector, that are related to the work that is presented in the thesis.

The evolution in the HMI has not been limited to improve the quality of the HMI, neither to improve the design and the implementation of the regular interfaces. Multiple concepts such as multimodal interfaces, adaptive interfaces and smart interfaces appeared.

Different solutions that are related to the human-machine speech already exist. Some are commercial solutions and others are open source solutions. Some of those solutions are ready to be used, others are used as a base to build solutions over, while others are research projects. A comparison between the different solutions is illustrated in Table 2.1.

Speech recognition integrates three models that work together: the acoustic, the pronunciation and the language models. The acoustic model crops the waveforms of the speech into small fragments to figure out each sound that was made by the user, then the pronunciation model makes words for the sounds that were identified by the acoustic model and, lastly, the language model puts the words together to construct sentences.

The speech synthesis process is composed of three stages, the first stage is the preprocessing stage in which the system goes through the different words of the text to find out the most appropriate way to read the text, then in the second stage the words are transformed into phonemes which are the sound components that compose any spoken word and, lastly, the phonemes are converted into sounds using one of the following approaches: concatenative (using phonemes extracted from recorded human voices), formant (using phonemes generated by the machine) or articulatory (makes the machines speak by modeling the human complex vocal apparatus, which is the most complex approach).

Table 2.1. Speech-Interactions related solutions comparison -: free testing period & some free monthly transactions, ‡: some features are free others require payment, °: prebuilt language(s), S-t-T: speech-to-text service and T-t-S: text-to-speech service*

<table>
<tr><th rowspan="2">Solutions</th><th rowspan="2">Serv.</th><th colspan="2">Services</th><th colspan="8">Languages</th><th rowspan="2">Comments</th></tr>
<tr><th>Free</th><th>Paid</th><th>EN</th><th>FR</th><th>RO</th><th>SW</th><th>DA</th><th>PL</th><th>IT</th><th>Add.</th></tr>
<tr><td rowspan="2">IBM Watson</td><td>S-t-T</td><td rowspan="2"></td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td></td><td rowspan="2">✓</td><td rowspan="2">Internet dependent</td></tr>
<tr><td>T-t-S</td><td>✓</td></tr>
<tr><td rowspan="2">Dragon NaturallySpeaking</td><td>S-t-T</td><td rowspan="2"></td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">Internet dependent</td></tr>
<tr><td>T-t-S</td></tr>
<tr><td rowspan="2">Lexix</td><td>S-t-T</td><td rowspan="2"></td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>T-t-S</td></tr>
<tr><td rowspan="2">Windows Desktop Speech Technology</td><td>S-t-T</td><td rowspan="2">✓</td><td rowspan="2"></td><td rowspan="2">✓</td><td rowspan="2">✓</td><td></td><td></td><td></td><td></td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">Microsoft Windows dependent</td></tr>
<tr><td>T-t-S</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td rowspan="2">Azure Speech Services</td><td>S-t-T</td><td rowspan="2"></td><td rowspan="2">✓*</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td></td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">Internet dependent</td></tr>
<tr><td>T-t-S</td><td>✓</td></tr>
<tr><td rowspan="2">Google Cloud Speech Services</td><td>S-t-T</td><td rowspan="2"></td><td rowspan="2">✓*</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td>✓</td><td rowspan="2">✓</td><td>✓</td><td>✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">Internet dependent</td></tr>
<tr><td>T-t-S</td><td></td><td></td><td></td></tr>
<tr><td>Alexa Voice Service</td><td>All Services</td><td>✓</td><td></td><td>✓</td><td></td><td></td><td></td><td></td><td></td><td></td><td>✓</td><td>Internet & Compatible device dependent</td></tr>
<tr><td>Julius</td><td>S-t-T</td><td>✓</td><td></td><td>✓°</td><td></td><td></td><td></td><td></td><td></td><td></td><td>✓°</td><td>Any language can be supported</td></tr>
<tr><td>Cmusphinx</td><td>S-t-T</td><td>✓</td><td></td><td>✓°</td><td>✓°</td><td></td><td></td><td></td><td></td><td></td><td>✓°</td><td>Any language can be supported</td></tr>
<tr><td rowspan="2">HTK</td><td>S-t-T</td><td rowspan="2">✓</td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2"></td><td rowspan="2">✓</td><td rowspan="2">Any language can be supported</td></tr>
<tr><td>T-t-S</td></tr>
<tr><td>Kaldi</td><td>S-t-T</td><td></td><td></td><td>✓°</td><td></td><td></td><td></td><td></td><td></td><td></td><td>✓</td><td>Any language can be supported</td></tr>
<tr><td>Jasper</td><td>S-t-T</td><td></td><td></td><td>✓°</td><td></td><td></td><td></td><td></td><td></td><td></td><td>✓°</td><td>RPi board dependent
Any language can be supported</td></tr>
<tr><td>Apple Speech Framework</td><td>S-t-T</td><td>✓</td><td></td><td>✓</td><td>✓</td><td></td><td>✓</td><td>✓</td><td></td><td>✓</td><td>✓</td><td>iOS & internet dependent</td></tr>
<tr><td rowspan="2">Ispeech</td><td>S-t-T</td><td rowspan="2"></td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2"></td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">✓</td><td rowspan="2">Internet dependent</td></tr>
<tr><td>T-t-S</td></tr>
<tr><td>eSpeak</td><td>T-t-S</td><td>✓</td><td></td><td>✓°</td><td>✓°</td><td>✓°</td><td>✓°</td><td>✓°</td><td>✓°</td><td>✓°</td><td>✓°</td><td></td></tr>
<tr><td>ResponsiveVoice.JS</td><td>T-t-S</td><td>✓‡</td><td>✓‡</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>Internet dependent</td></tr>
<tr><td>LUIS</td><td>NLP</td><td></td><td>✓*</td><td>✓</td><td>✓</td><td></td><td></td><td></td><td></td><td>✓</td><td>✓</td><td>Internet dependent</td></tr>
<tr><td>DialogFlow</td><td>NLP</td><td>✓‡</td><td>✓‡</td><td>✓</td><td>✓</td><td></td><td>✓</td><td>✓</td><td></td><td>✓</td><td>✓</td><td>Internet dependent</td></tr>
<tr><td>RASA NLU</td><td>NLP</td><td>✓</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>Any language can be supported</td></tr>
<tr><td>Wit.ai</td><td>NLP</td><td>✓</td><td></td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td><td>Internet dependent</td></tr>
<tr><td>Azure Bot Service</td><td>DM</td><td></td><td>✓*</td><td colspan="8">N/A</td><td>Internet dependent</td></tr>
<tr><td>Bot Builder SDK</td><td>DM</td><td>✓</td><td></td><td colspan="8">N/A</td><td>Microsoft Windows dependent</td></tr>
</table>

In addition to the solutions that appear in Table 2.1, many researches addressed the different aspects of the human-machine voice interactions and they offered solutions for the different identified issues, such as in [4, 5, 6 and 7] for the speech recognition, [8, 9 and 10] for the speech synthesis, [11, 12 and 13] for the natural language understanding.

Regarding the Dialog Management (DM) of the human-machine speech interactions, it has been addressed through two main approaches: knowledge-based DM and principle-based DM. Many systems that use knowledge-based DM approach were proposed, such as the SUNDIAL system proposed in 1993 by Peckham [14] and the ARISE system proposed in 1999 by Lamel et al. [15]. The principle-based DM approach is used in different works such as in the work of Chu-Carroll presented in 1999 [16], in the work of Seneff and Polifroni presented in 2000 [17], and in the work of Zue et al. presented in 2000 [18] as well as in the systems presented by Lemon and Liu in 2006 [19] and by Varges et al. in 2008 [20].

Multiple works investigated the interactions between humans and machines through touch-based gestures. From a side, the single touch-based gesture in which the users use a finger to interact with the system through a touch screen such as in [21, 22 and 23] where the users used one of their fingers to either swipe or pan, in [24 and 25] where the users used one finger to either drag or move an object and in [22 and 26] where the users used one finger to draw gestures pattern. On the other side, multiple works investigated the multi touch-based gesture in which the users use multiple fingers, from the same hand or from both hands, to interact with the system through a touch screen such as in [22, 24, 25 and 27] where users used multiple fingers to either rotate, resize or steer. Some applications that support multi-touch gestures and target elderly people as their main users have been developed such as the multi-touch training games from the HERMES project [28] and the email application designed by Hollinworth and Hwang in [29].

Multiple works investigated the touch-free gesture in which the users interact with the system through their body movement. The hand gestures are probably the most investigated touch-free gestures form, each touch-free hand gesture can be divided in five phases: the rest position, the preparation, the gesture stroke, the holds, the retraction phases, based on the work presented in [30 and 31]. The rest position is the phase in which the hand is stable in a position from where the interaction is initialized, the preparation phase in which the hand is moved from the rest position, the gesture stroke in which the hand executes the gesture command, the holds phase in which the hand is almost immobile after the end of the gesture execution and, lastly, the retraction phases in which the hand is moved to the rest position [32]. Some works proposed additional phases to the hand free-touch gesture such as in [33] where a recoil phase was proposed. In 2003, Gut and Milde proposed to break down each gesture into several morphological features such as the hand direction, the hand shape, the movement type, etc. [34]. Another investigating touch-free gesture form is the gesture through the movement of the head such as in [35 and 36]. The touch-free gestures are currently used during the training and the rehabilitation programs, in which, the system tracks the movements of the users, and reproduce them on the screen by an avatar in real time such as the "Voracy Fish" [37] and the "Hammer & Planks" [38] games. The touch-free gestures are currently also used for entertainment and social activities such as in different Microsoft Xbox and Nintendo Wii games.

Multiple studies investigated the human-machine interactions through the emotions of the user. The user's emotions can be recognized by the perception of the user's facial expressions, the user's body language or the user's voice. The works that recognized the emotion of the user through the facial expressions of the user, divided the recognition into three major phases. The first phase is the "face detection" phase that is followed by the "facial feature extraction" phase, then by the "facial expression classification" phase. In [39], Vineetha et al. used a Microsoft Kinect Sensor and its related techniques to detect the facial expressions of the user from which they recognized the emotion of the user. In [40] Yousef et al. used the Microsoft Kinect Sensor and its related Windows SDK to detect the facial expressions of the user from which they recognized the user's emotion. They used the support vector machine (SVM) and the k-nearest neighbors (k-NN) classifiers to classify the obtained facial expression. In [41] Lemaire et al. proposed an automatic 3D approach to detect the facial expressions of the user using histograms of oriented gradients and curvature maps. Some works recognized the emotion of the person through his/her body language as in [42 and 43] while other works recognized the emotion of the person through his/her voice such as in [44].

Multiple systems that integrate a multimodal interface already exist, such as the Memphis Intelligent Kiosk Initiative [45], MATCHKiosk Multimodal Interactive City Guide [46], the Canesta [47] system, the Touch'n'Speak [48] system, the Put That There [49] system and Gaze detection pointing system for people with disabilities [50] as well as the personal assistants that appeared recently, such as Siri from Apple, Cortana from Microsoft and Google Assistant from Google.

As a result of the great efforts put by the AAL community to encourage the development of systems that pack together several solutions that were developed or proposed in the field of health management, well-being and aging at home, different AAL systems and projects appeared, such as the inCASA [51] project, Persona [52] platform, NITICS [53] project, Sociable [54] platform, PersonAAL [55] project, CareWell [56], healthy@work [57] project, Wellbeing [58] project and EldersUP! [59] project.

Regarding the analyzed solutions that are related to the voice module, the majority of solutions need to be combined with other solutions in order to work as a functional voice module that can interact with the user, while some integrate different functions that allow them to interact with the user without the need of any other solution. Some of those solutions are cloud-based and others can be deployed locally. Most of the solutions support the English language together with other languages. The performance varies between different solutions and each solution performance varies also depending on the language which is used for. The performance of the cloud-based solutions is better than the performance of the locally deployed solutions. However, they depend on the availability of an internet connection.

Regarding the analyzed solutions that are related to the gesture module, different works proposed solutions for the touch-based gestures (single-touch and multi-touch) while other works proposed solutions for the touch-free gestures. Different works found that gestures represent a suitable way of interaction for the elderly people.

Regarding the analyzed solutions that are related to the emotion module, they used different approaches to determine the emotion of the user, such as determining the user's emotion through the user's facial expression or the user's body language or the user's voice.

Regarding the analyzed multimodal systems, it should be noted that the systems integrate the speech with another modality as their input modalities. They support only a single language and most of those systems are device-dependent, not compatible across platforms and they do not offer any adaptive capabilities.

Regarding the analyzed AAL systems, the majority of them integrate different services to support the elderly people in their daily life and to help them in maintaining a healthy lifestyle. The majority of those systems allow the users to interact with them through multimodal interactions. With the exception of the Sociable platform, neither solution among the set of the reviewed systems integrates a supervised physical exercise module that helps the elderly people exercise at their residence which is very important in order to maintain a healthy life. With the exception of the PersonAAL and EldersUP! projects, neither solution takes into consideration the importance of integrating adaptive capabilities for their user interfaces. All the mentioned solutions (being commercial solutions or research projects) have limited functionalities and do not cover all the functionalities that can be targeted by an AAL system. Moreover, from an architectural perspective view, the reviewed solutions are bound to one dominant technology that restrict the implementation and requires a lot of efforts to integrate new functionalities and modules.

Chapter 3

CAMI – A System for Active and Assisted Living

A significant part of the work presented in the thesis took part within the European project "CAMI: Artificial intelligent ecosystem for self-management and sustainable quality of life in AAL" in the framework of the Active and Assisted Living Programme of European Union, from 2015 to 2018 [60].

CAMI is an artificial intelligent ecosystem for self-management and sustainable quality of life in AAL. The project consortium is composed of eight partners from five European countries and the University Politehnica of Bucharest (UPB) is the coordinator of the consortium.

CAMI incorporates the major functionalities of the AAL systems such as activity planners, fall detection, supervised physical exercises, health monitoring, home monitoring and automation [61]. Moreover, CAMI includes a robotic telepresence unit and integrates a multimodal interface that allows the user to interact smoothly with the system. The target users of the system are the elderly people aged fifty-five to seventy-five. The caregivers and health professionals can remotely access the system. They can visualize the user's health data and activities in real time, as well as the user's activity and health records. Further, the caregivers and the health professionals, depending on their position and specialty, can adjust the medication, the daily exercises and some daily tasks of the user.

CAMI has multiple key features. One of those key features is that the system can be tailored to the user's needs. Another key feature is that the system can be easily extended at any time with more functionalities and devices. The modular architecture of CAMI represents another key feature of the system, it is composed of 6 modules: the sensor unit, the robotic telepresence, the CAMI gateway, the CAMI cloud, the multimodal user interface and the 3rd party platforms unit.

The CAMI system offers different services to the user. The user is able to add or remove, easily, any service whenever he/she wants. The main services of CAMI system are the following: health monitoring, fall detection and alarm, home and environment management, physical activity monitoring, program management, multimodal interactions, and robotic telepresence.

Chapter 4

Designing Multimodal Interfaces

This chapter presents a general overview about the design of the multimodal interfaces. Before the beginning of the design process, it is very important to identify who are the targeted users of the system and where will the users use the interface (mainly and occasionally), as well as what are the different functionalities of the system. Only after identifying the previous information, the design process of the multimodal interface can begin.

Usually designing an interface for a complex system is a very delicate task, especially for an active and assisted living system in which the interface plays a decisive role in the acceptance of the system by the users, subsequently the success of the whole system. The challenges that should be taken into consideration during the design process of a multimodal interface for any system can be divided into two main groups: the challenges that are related to human factors and the challenges that are related to technical factors. The challenges that are related to human factors vary depending on the targeted users of the system and can be divided into three high level categories depending on the causing factor of the challenge: physical issues, cognitive issues and computer experience [62]. Besides the challenges that are related to human factors, technical related challenges exist, such as the simultaneous contradictory commands, the fusion and fission of data, as well as the security and privacy issues.

To create an easy and pleasant interaction experience for the users, the design of any interface should follow some guidelines. Before talking of the guidelines that should be followed, it is very important to mention that, in addition to those guidelines, different techniques should be used to ensure the logical flow of information such as the use of the storyboarding techniques.

Regarding the content of the interface that will be displayed on the screen, some aspects of the interface should behave in a consistent way throughout the different screens (or pages) of the interface. The interface terminology as well as the used font, font-size, icons and colors should be consistent throughout the different screens of the interface. The interface should be simple, the icons used should be familiar to the users, the long sequences and complex tasks should be broken into separate steps and simple tasks. The interface design should take into consideration the limitations of the human memory. Thus, the displayed information should be well organized into a limited number of sections, the interface should always provide a navigation clue which allows the user to know on what screen he/she is, the interface should always provide the user with an informative feedback on what has just happened or what it is happening and to generate user appropriate warnings when they are needed. Moreover, the quantity of information should be minimized, the input data fields should be automatically formatted and validated to help the user, as well as to avoid the simple errors and for the information that are flashed for a specific period on the screen, it is important that the flash period is not too short and that it allows the user to clearly receive the information. Lastly, the interface design should minimize the mental transformations of the displayed information and it should have appropriate visual cues not only to help the user navigate through the interface, but also to make the important information easier to be found. The interface should look similar as much as possible (and even identic when possible) through the different devices and it should preserve the used terminology, colors and icons.

Regarding the different phonetic and gesture (touch-free and touch-based) inputs, the commands should be consistent throughout the different devices, but also across the different services and modules of the system. The phonetic outputs, and especially in the case of the phonetic feedback, should also be consistent through the different screens of the interface.

If the interface has different adaptive capabilities, it is very important that each adaptive feature will introduce itself clearly to the user on the first use. It is also very important that the user has an option to turn off easily each adaptive feature as well as the whole adaptive capability.

Any interaction between a user and a machine can be divided in eight states: four states for each side [63] as illustrated in Figure 4.1. The following multimodal human-machine interaction model is inspired by the Norman's action cycle [64]. From the side of the user, there are the decision, action perception and interpretation states while from the system side there, are the perception, interpretation, computation and action states.

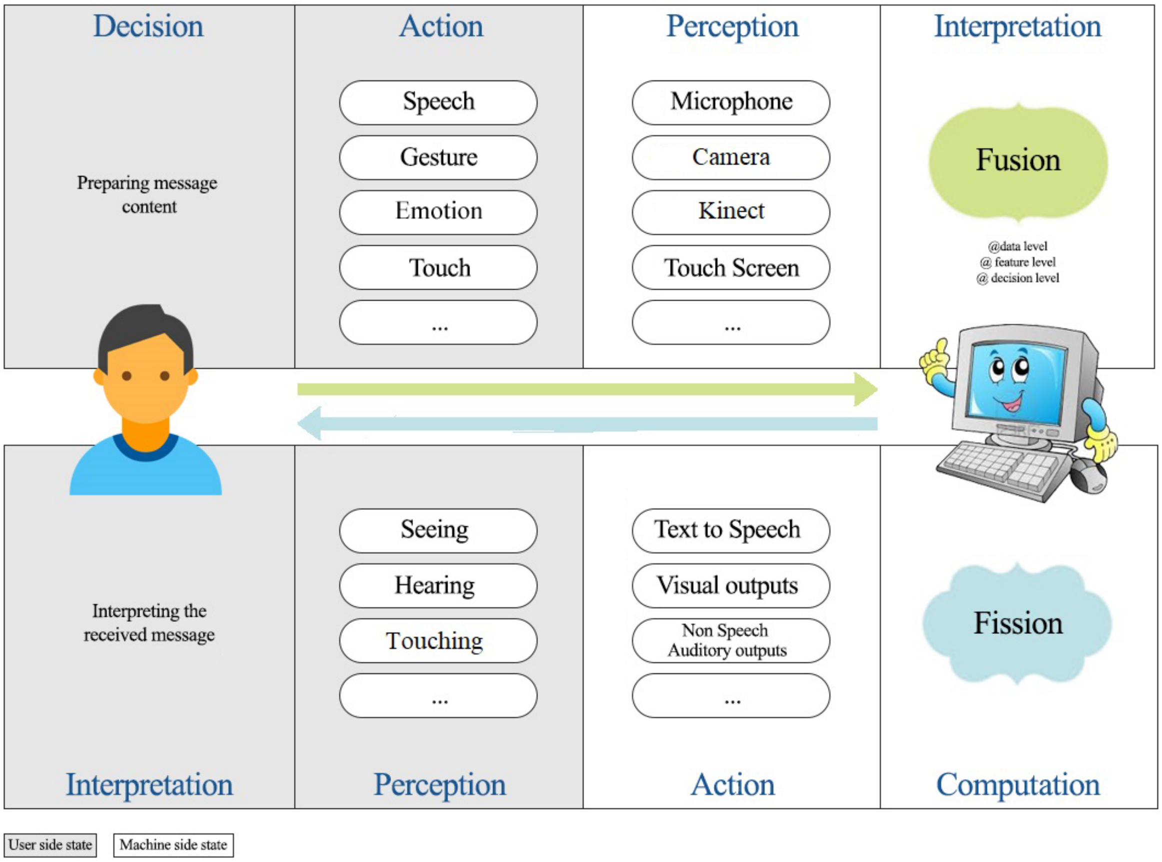


Figure 4.1. A Model for Multimodal Human-Machine Interaction

Chapter 5

An Adaptive Multimodal Interface for CAMI

This chapter presents a general overview about the design of the multimodal interface for the CAMI system.

Since the CAMI system targets the elderly people as its main users, the interface is designed to fulfill the needs and the requirements of the elderly people such as big font size, big button size, easy navigation, sound volume level as well as more natural ways of interaction with the system. At the same time, the interface allows an easy access to the different features of the system.

Since the targeted users of the system are from different countries, they come from various background and culture with different preferences, habits and knowledge, as well as different issues. Therefore, the interface has different adaptive capabilities and customizable features. It is multilingual, it supports the different languages of the CAMI consortium (English, French, Romanian, Swedish, Danish, Polish and Italian) with the exception of the voice module that supports only the English, French and Romanian. More languages can be added easily to the interface including to the voice module.

The interface works across different platforms and devices. It supports multimodal interactions and it is composed of four modules: graphical user interface, voice, gesture and emotion. Although the different modules are intended to work together, each module is designed to function independently of any other module. The architecture of the interface is illustrated in Figure 5.1.

The design of the interface ensures that the interface is consistent across all the devices and also ensures that the functionalities of the different modules of the interface are consistent across the different devices. However, if a device or an operating system does not support a specific module of the interface, this module will be disabled automatically and the user will be notified about it.

The interface has multiple functionalities. It allows the user to interact smoothly with the system through multimodal interactions using the GUI or the voice, touch-based gesture and touch-free gesture commands. In addition, it allows the system to generate visual and phonetic outputs and to track the emotional status of the user, as well as the evolution of the user during the different physical exercise sessions. The interface allows the user to access easily from any device his/her health data (real time data or health record), appointments, reminders, and smart notifications as well as the physical exercises and the different features of the system. It allows the user to customize different aspects of the interface (e.g. used colors, module display order, feedback form, interface language, etc.) and to switch on or off the different modules of the system, the different adaptive capabilities of the interface as well as the different input and output modules of the interface. The interface can communicate the preferences of the user to 3rd party applications. In addition, it allows the caregiver to visualize from any device the real-time health data of the user and the health record of the user as well as the different health reports of the user. The interface allows the official caregiver to customize the medication plan for the user and the normal range of values for the health measurements for the user. It also allows the official caregiver to adjust the health profile of the user and the exercises that should be practiced by the user, as well as some other options such as the height of the user. In very specific cases, the interface allows the official caregiver to customize everything in the user’s system and interface.

To ensure a user-friendly interface for elderly people, the design of the multimodal interface observed the guidelines for the user interface design and took into consideration the requirements of the elderly people as well as the issues and solutions that were mentioned in the previous section. The different aspects of the interface behave in a consistent way throughout the different pages and modules of the interface, the icons, the fonts, the colors and the buttons design are preserved throughout all the pages of the interface. The voice, gesture and touch commands are also preserved through the interface and through all the system services. The key elements of the interface have the same position all times for the different pages of the interface. The interface design is simple, the used icons are familiar for almost every user and they are easy to be interpreted. The default colors used to illustrate a message are also easy interpretable since the used colors for this matter are the colors used in the traffic lights: red to symbolize an alarmed situation or high severity, yellow to symbolize a situation that needs attention or medium severity, and green to symbolize that the situation is good or low severity. Those colors can be easily modified by the user from the setting page of the system. The interface is easily navigable through big buttons or through simple speech and gesture (touch-based or touch-free) commands. It organizes the information by category and highlights the important information to catch the attention of the user. Lastly, the interface always provides appropriate feedback for the user through visual and phonetic clues.

The interface is designed to have different adaptive features to better satisfy the need and preferences of each user.

The first adaptation feature of the interface is the personalization capability. From the side of the caretaker, the interface in designed in a way that allows the user to choose what modules should be displayed in the home page and in what order. The user can easily customize the different colors of the interface such as the background color, the text color, the color of buttons and the colors of the messages. The user can choose the default language of the interface by choosing a language from the seven available and to adjust the font sizes. The user can customize the order of the buttons in the menu and the units of measurement used by the interface. In addition, the user can switch on or off the avatar, the different modules of the system as well as the different modules of the interfaces. The user can customize the form of the system's feedback, turn on and off each of the automatic adaptation features and adjust some other preferences such as the preferred input modality, the preferred time frame for exercising, the multi-touch gestures, the speed of the speech synthesized and its accent and many other preferences. From the side of the caregiver, the interface is designed in a way that allows the official caregiver to customize the normal range of values for the health measurements for each user, which allows the system to analyze correctly the results of the health measurements. Some of those values are: the minimum and maximum values accepted for the systolic and diastolic blood pressure, the minimum and maximum values accepted for the heart rate, the daily recommended number of steps and the daily recommended time to sleep. The official caregiver can also adjust the medication plan for the user, the health profile of the user, the exercises that the user should practice and some other options.

The second adaptation feature of the interface is the automatic adaptation capability. The interface is designed in a way that allows the system to adapt automatically the interface according to the profile and the status of the user, but also to the activity of the user and the system configuration, as well as to other parameters such as the environment conditions.

Lastly, the interface is designed to work independently from any device or any operating system. Therefore, it works cross-platform which enables the user to interact smoothly with the system through any device that the user already has or that he/she prefers to use, as well as to interact smoothly with the system through different devices (at the same time or whenever the user wants to use another device to interact with the system).

To insure multiple ways of interaction between the user and the system, the system integrates a multimodal interface that is composed, as mentioned before, of four modules: graphical user interface, voice, gesture and emotion.

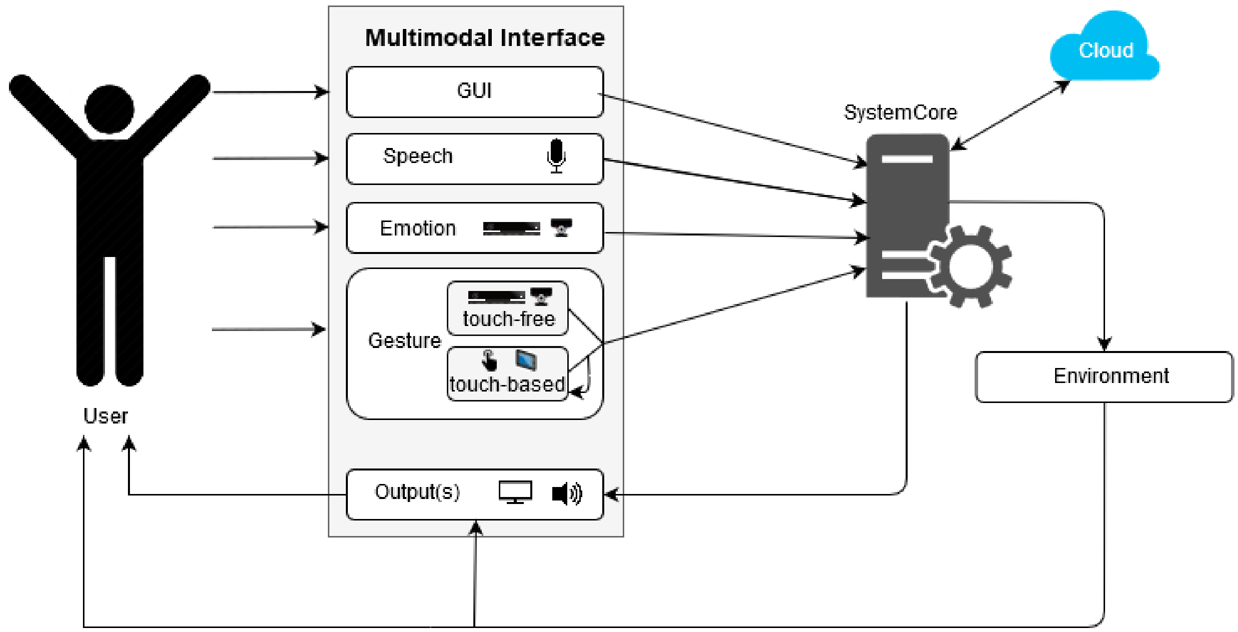


Figure 5.1. The Interface Architecture

The graphic user interface offers to the user a classical alternative for the other interface modules to interact through it with the system.

The voice module is composed of five components: Audio Preprocessing, Automatic Speech Recognition (ASR), Natural Language Understanding (NLU), Dialog Management (DM) and Text-to-Speech Synthesis (TTS), as illustrated in Figure 5.2. It has two modes of working: Online and Offline. This module has as inputs the speech commands of the user or the speech synthesized commands from the system (when it is initiated by the system or by the user using another input module) and as outputs the speech synthesized audio and/or a visual output or an order that is sent to the system.

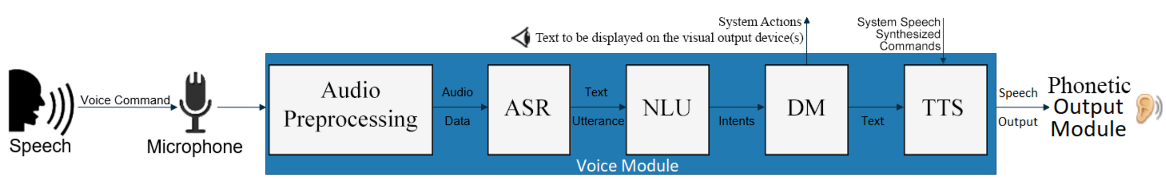


Figure 5.2. Voice Module Architecture

The gesture module is composed of two submodules: a submodule for the touch-based gestures and another submodule for the touch-free gesture, as illustrated in Figure 5.3. The first submodule is composed of two components: data acquisition and data processing, and it has as input the touch command of the user and as outputs the corresponding output for the user command (visual and/or phonetic) and/or an order that is sent towards the system. While the second submodule is composed of two components: data acquisition and data transformation, and it has as input the touch-free gesture command of the user and as output an order that is sent to the system or the 2D coordination of the user movement to be reproduced on the system screen during the physical exercise sessions.

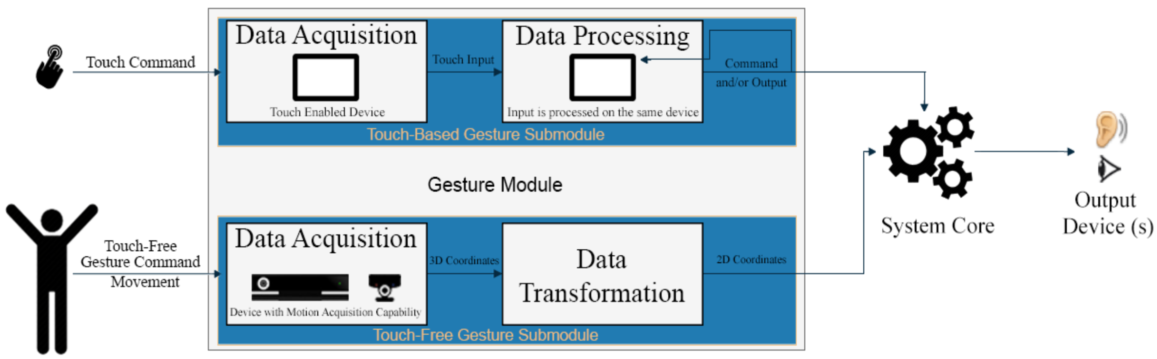


Figure 5.3. Gesture Module Architecture

The emotion module is composed of two components: data acquisition and data transformation, as illustrated in the Figure 5.4. It has two modes of working: Online and Offline. This module has as input the emotion of the user and as output the emotional status of the user that is sent to the system.

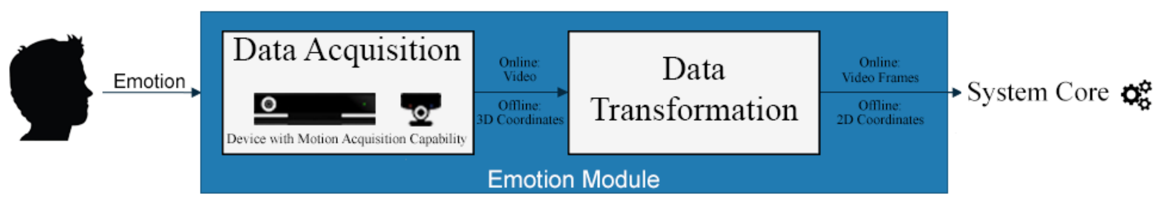


Figure 5.4. Emotion Module Architecture

Chapter 6

Interface Implementation

This chapter presents a general overview about the implementation of the multimodal interface.

6.1. Graphical User Interface Implementation

To ensure a good functionality of the interface across the different devices and platforms, the interface was developed mainly using the HTML5, CSS3 and JavaScript languages. However, some modules of the interface may have been developed using other languages.

The interface checks the configuration of the systems and, if a module of the system is not available or turned off, all the information regarding this module will be hidden in the graphical user interface and all the commands that are associated to that module will be disactivated. To implement the previous feature, each system module has a global Boolean variable that is equal to true, if the system module is available and activated. It is equal to false if the system module is deactivated or unavailable within the system. If the value of the Boolean is true, the information of the corresponding module is visible and its associated commands are activated, while if the value of the Boolean is false, the corresponding information is hidden and the associated commands are deactivated.

Using Bootstrap, the GUI adapts itself to the screen size of the device. The Bootstrap grid system helps to align the different elements of the page based on rows and columns. The page is composed of one or more containers. Each container is composed from one or more rows and each row can be decomposed in up to twelve columns. According to the interface design needs, the interface uses the bootstrap grid system that has four predefined grid sizes that correspond to: phones or extra small devices, tablets or small devices, medium desktop devices and large desktops devices.

The interface extracts from each user profile information that is related to the GUI preferences of the user and adapts the GUI to satisfy the user preferences. To implement this feature, each group of elements has their color associated to a variable. The variable value is extracted from the color selected by the user for this group which is stored in the user profile. Each module position depends on a variable. The variable value is extracted from the modules order selected by the user which is stored in the user profile. The home-page of the GUI is illustrated in Figure 6.1. It displays the last health measurement results, the daily number of steps and the sleep duration in the current day. It also displays the number of reminders, the daily appointments of the user and some smart notifications regarding the user's health status and activity. It should be mentioned that all communications with the system core takes place through JSON Web Tokens.

Each measurement result is displayed together with the time when the measurement was taken, a colored icon that reflects the type (by its shape) and the result of the measurement (by its color). Two smart notifications appear in the notification zone for each measurement. The color of the icon varies according to the body mass index result for the weigh measurement and according to the measurement result for the blood pressure and heart rate. While for the number of steps and sleeping time, the icon color varies according to the addition result of the different measurements that were taken on the same day. As mentioned before, the accepted range of values for each type of measurement can be customized for each user by the caregiver, while the colors that reflect the measurement result status can be customized by the user.

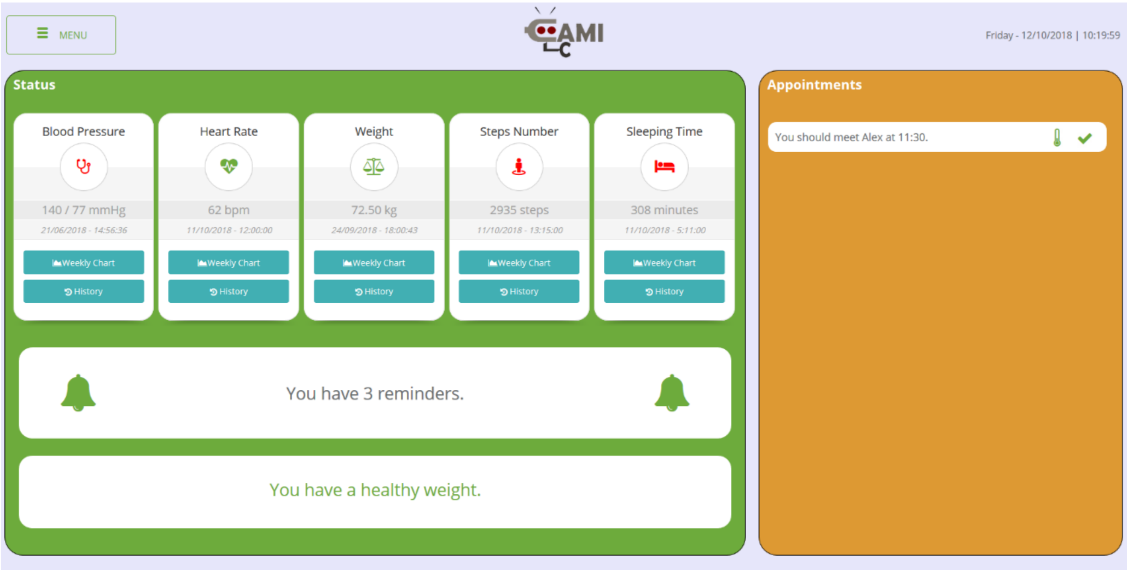


Figure 6.1. CAMI Home-Page

The interface extracts from each user profile information the user's height, the acceptable range for each of the health measurements value, the daily required number of steps and the daily required sleep duration. Then, the system adapts the notifications according to the extracted information to fit each user's health requirements. In what follows, the accepted range of values represents the default values used by the interface.

Regarding the weight measurement, the BMI can be obtained by dividing the weight of the person to his/her height squared. If the BMI result is inferior to 18.50 Kg/m^2 the person is underweight. If the BMI result is between 18.50 Kg/m^2 and 24.99 Kg/m^2 the person has a normal weight. If the BMI result is between 25.00 Kg/m^2 and 29.99 Kg/m^2 the person is overweight. Lastly, if the BMI result is superior the 30.00 Kg/m^2 the person is obese [65].

Regarding the blood pressure measurement, the accepted range of value is between 90 mmHg and 120 mmHg for the systolic blood pressure and between 60 mmHg and 80 mmHg for the diastolic blood pressure, as recommended by the American Heart Association [66]. Regarding the heart rate, the accepted range of value is between 60 bpm and 100 bpm, as recommended by the American Heart Association [67]. Regarding the required daily walking steps, the accepted range of value is between 2000 steps and 9000 steps [68]. Regarding the required daily duration of sleep, the accepted range of value is between 420 minutes and 480 minutes [69].

Two icons are displayed next to each appointment/reminder. The first icon illustrates the importance of the appointment reminder. There are three levels of importance: low, medium and high importance or severity. The second icon allows the user to acknowledge the appointment or the reminder.

The interface allows the users to visualize quickly weekly health charts that illustrate the variation of their health measurements during the last seven days. In addition to the weekly health charts, the users have the possibility to visualize their health measurements history for an interval of time that they desire under a chart format or a list format. The health measurements can be visualized combined or each measurement alone.

6.2. Voice Module Implementation

As previously mentioned, the voice module is composed of five components: Audio Preprocessing, Automatic Speech Recognition (ASR), Natural Language Understanding (NLU), Dialog Management (DM) and Text-to-Speech (TTS) Synthesis. The voice module is multilingual. It supports English and French in both modes of working: Online and Offline. In addition, it supports Romanian for the online mode of working.

Audio Preprocessing: the audio preprocessing component extract data that is useful for the voice module, from the profile of the user and from the setting of the system. The data is as follows: the speech language which is extracted from the user's preferred language that is stored in each user profile, the audio bit depth, the proper frequency of the audio and the number of channels of the audio are extracted from the voice module settings. In addition to the previous data, the automatic punctuation preferences, audio encoding, the TTS solution that should be used, the ASR solution that should be used, the ASR model and the preferences regarding the word time offsets as well as other options are extracted from the voice module settings.

Automatic Speech Recognition (ASR): For the online mode of working, Microsoft speech-to-text service and Google speech-to-text service are used as ASR solutions for French and English. The user can easily choose which ASR solution should the voice module use in the settings page, as well as to switch between them. For Romanian, the Google speech-to-text service is used as ASR solution. The system captures the voice command of the user and sends the audio file to the cloud of the used service, both services return the text transcription in real time. The default configuration of the interface is to use the Microsoft speech-to-text service as ASR solution for French and English; and to use the Google speech-to-text service as ASR solution for Romanian. The voice module sends the audio file that contains the recorded voice command of the user to the cloud of the chosen ASR service using a JSON file. The result of the recognition is obtained through a JSON file. For the offline mode of working, the number of the recognized commands is reduced. The Windows Desktop Speech Technology provided by Microsoft is used as the ASR solution.

Natural Language Understanding (NLU): For the online mode of working, the Language Understanding Intelligent Service (LUIS) provided by Microsoft is used as the NLU solution for French and English. LUIS allows the system to understand what the users want when they express themselves with their own words. It uses machine learning to improve the extraction of meaning of the user's input in natural language. Therefore, by sending the resulted text from the ASR component to LUIS, the system will receive back relevant and detailed information regarding the request of the user in a JSON file. To cover the different interactions between the user and the system, different intents were created, each intent has multiple entities. Some entities have one or more children. For Romanian, wit.ai API is used as the NLU Solution. The workflow and the outputs are similar to those explained for LUIS. For the offline mode of working, RASA NLU provided by Rasa Technologies GmbH is used as the NLU solution. It extracts the meaning of the user's input in natural language and sends back to the system relevant and detailed information regarding the request of the user in a JSON file. Similar to the online mode of working, different intents were created and for each intent multiple entities have been associated.

Text-to-Speech (TTS) Synthesis: For the online mode of working, Microsoft text-to-speech service and Google text-to-speech service are used as TTS solutions for French and English. The user can easily choose which one the voice module should use in the settings page, as well as to switch between them. ResponsiveVoice.JS API is used as TTS solution for Romanian. The DM sends the text that should be spoken to the cloud of the used service through a JSON and both services return the audio file that contains the artificial spoken text in real time. The default configuration of the interface is to use the Microsoft text-to-speech Service as TTS solution for French and English; and to use the ResponsiveVoice.JS API as TTS solution for Romanian. For the offline mode of working, the Windows Desktop Speech Technology provided by Microsoft is used as the TTS solution.

6.3. Gesture Module Implementation

As mentioned previously, the gesture module is composed of two submodules: the touch-based gesture submodule and the touch-free gesture submodule. In its turn, the touch-based gesture submodule is formed of two components: data acquisition and data processing. The touch-free gesture submodule is also formed of two components: data acquisition and data transformation.

The touch-based gesture submodule supports single-touch and multi-touch gestures. It collects the data of a user input through the touch screen of the device from which the interface is accessed. The touch screen represents the data acquisition component. The collected data is sent to the data processing component where it is processed by the event system of the device. Some commands are executed directly by the device itself while other commands are sent to the system core that executes them. To implement this submodule, Hammer.js, which is a JavaScript library, was used. Hammer.js library allows to easily create touch gestures. It has certain default multi-touch gestures (e.g. Pinch, Rotate, etc.) and certain single-touch gestures (e.g. Pan, Press, etc.). However, only the multi-touch gestures were used from the Hammer.js library since it was preferred to keep the single-touch input as simple as possible for the users. The command of each gesture was customized easily through a JavaScript code.

The touch-free gesture submodule tracks the user's movements through a Microsoft Kinect V2 sensor which provides RGB data, depth maps, and the user tracking under the form of a 3D skeleton that indicates the most important 25 body joints. In comparison with the first version of the Kinect sensor, the Kinect V2 sensor has a wider field of view and integrates an enhanced depth sensor and multiple enhanced microphones. In addition, it provides an improved skeleton tracking and a significantly improved tracking in the environments with reduced light conditions. The data acquisition component tracks the body movements of the user through a Kinect V2 sensor and receives from it the coordinates of the body movements as 3D coordinates. Then, those coordinates are sent to the data transformation component in which they are converted into 2D coordinates using classical viewport and affine transformations. Then the obtained 2D coordinates are sent to the core system. During any physical exercise session, the system uses the received 2D coordinates to reproduce, in real time, the movements of the user on the screen through the user's avatar. The exercise games were developed using the Unity 3D engine and the Microsoft Kinect for Windows SDK 2.0 together with the Unity Pro Packages. Besides the physical exercise sessions, the system analyzes the received 2D coordinates to recognize the touch-free gesture then executes the corresponding command. The touch-free gestures that should be recognized by the system were built using the Visual Gesture Builder (VGB) provided by the Kinect SDK 2.0. Besides creating the gestures, the VGB was used to train the system to recognize the gestures using the machine learning algorithms that it integrates. At the end, the VGB was also used to test the recognition of each gesture by the system. Therefore, for each gesture some information was provided to the VGB through the VGB Gesture Wizard such as if the gesture relies on the joints that are in the lower part of body (below the hips) or on the hand states (open, closed or lasso), to choose a training joint mask that represents best the gesture, if the gesture is a body side gesture or if the gesture is symmetrical, etc. Providing the previous information helps reducing the size of the data, reduce the training time and improve the accuracy of the recognition. For each gesture the Kinect V2 were used to record different clips that contain the gesture throughout the Kinect Studio software. A part of those clips (two thirds) was used to train the system to recognize the gesture, while the other part (a third) of the clips was used to test the acquired recognition. The training clips were imported to the VGB. For each training clip, the gestures were manually labeled: the Boolean value of the gesture is set up to true in the frames in which the gesture is present, this value is left empty in the frames in which the gesture is not present and, in some cases, this value is set up to false in the frames in which the gesture is not present, but a situation in which it can be viewed similar to the gesture by the system is present (the false-positive cases). After labeling all the training clips, the training process was launched and after the training process was finished the testing clips were imported to the VGB, so as to test the obtained recognition performance. This procedure was repeated for each of the touch-free gestures that should be recognized by the system.

6.4. Emotion Module Implementation

As mentioned previously, the emotion module is composed of two components: data acquisition and data processing. It has two modes of working: Online and Offline.

For the offline mode of working, the data acquisition component tracks the face of the user through a Kinect from which it receives the 3D coordinates of the user's face. Then it sends those coordinates to the data transformation component where they are transformed into 2D coordinates using classical viewport transformations and affine transformations. Subsequently, the obtained 2D coordinates are sent to the core system. The Face Tracking Basics API provided by the Microsoft Kinect for Windows SDK 2.0 was used to track and analyze the user's face expression in real time. It creates a 3D mask of the user's face, which allows the extraction of the most common and basic features of the user's face (e.g. position of the eyes, nose and mouth, status of the lips, jaw and eyebrows; etc.). Despite the fact that the Face Tracking Basics API generates over 1000 reference points from the user's face, the module uses only 51 of those reference points and maps them to 19 Action Units (AU) which are associated to basic emotions. These were chosen based on the associations described in the Facial Action Coding System (FACS). In addition, 19 Support Vector Regression (SVR) networks that integrate a part of the extracted reference points were used for computing intensities of the AU.

For the online mode of working, the data acquisition component tracks the face of the user through a camera from which it receives a video. Then it sends the video to the data transformation component in which video frames are extracted from the video. Afterwards, the extracted video frames are sent to the Microsoft Azure Face API in which they are analyzed and after that, the cloud sends back a JSON file that contains the result of the analyze. The system core deduces the emotion of the user from the recognition confidence rates received within the JSON file.

6.5. Some Features and Aspects of Implementation

Avatar: The avatar enhances the communication between the user and the system. It animates the interactions and captures the attention of the user. In addition to the inputs of the traditional input devices, the avatar can process speech and touch-based inputs. It generates phonetic and visual outputs. The avatar is multilingual and uses the current language of the interface to communicate with the user. Some aspects of the avatar can be customizable by the user such as its name, clothes, output types and virtual character (e.g. human-like, animal, object, etc.). The avatar also has some adaptive capabilities such as the adaptation of its clothes in function of the time or weather and the adaptation of its behavior in function of the user status. The avatar answers the queries of the user by extracting information from the system (locally) or by searching for information over the internet (if a connection is available) and informs the user about notifications, reminders and appointments. Moreover, the avatar assists the user to navigate through the interface if the user asks for help or if the system detects that the user needs help. The avatar can be easily switched on or off, by the user. The avatar is developed using HTML5, CSS3 and JavaScript and it is connected to the different modules of the interface.

Portability: The interface is packed into a compressed folder that is accompanied by a PHP file. Once the PHP file is executed, the interface's installation process is launched. During the installation process some information should be provided, some is optional while other is mandatory, such as the language that should be used during the installation process, the path of the installation directory, the users authentication method, the different cloud services that are available, the path of some resources (e.g. JSON files) and the different modules of the system as well as its functionalities. The only requirement for the installation process is the ability of the system to run PHP files.

Security & Privacy: To insure the security of the transmitted data, the interface uses JSON web token (JWT) to transmit data between the system and the interface, or between the different modules of the interface. The JWTs are encrypted and/or digitally signed using a secret key or a pair of public and private keys (depending on the option that was selected during the installation process of the interface). For privacy purposes, the user can access a page that displays what modules are activated and what data does each module collect. From the same page, the user can easily switch on or off any module, clear the collected data and set up different parameters.

Communication with 3rd party applications: The interface can export the preferences of the user to 3rd party applications, after obtaining the accord of the user, through a JSON file such as in the case of a smart scheduler, the interface exports toward the scheduler, the time-frames during which, the user prefers to exercise.

Fusion and Fission of Data: The interface integrates a fusion engine that composes and correlates data that comes from the different input channels. The interface also integrates a fission engine that separates the output data in function of their type.

Simultaneous Contradictory Commands: In case that two simultaneous contradictory commands are recognized, the interface integrates methods that analyze these commands and chooses which one should be executed. The decision is relying on a rule-based approach that takes into consideration the degree of confidence of each command recognition, the conditions of the environment, the health status of the user, the presence of one or more factors that may represent an advantage or disadvantage for an input modality and some other information.

Chapter 7

Experiments and Evaluation

This chapter presents a general overview about the implementation of the multimodal interface.

The previous two chapters presented the design and the implementation of the multimodal interface. The different modules of the interface have been tested in several stages to improve the final solution. The results of the tests that took place in earlier stages of the development are illustrated in [70, 71, 72 and 73]. The modules were tested in the laboratory, but also at users' residences. At the end, the whole interface was tested as a single solution.

The evaluation process envisaged to ensure the good functionalities for the different features and modules of the interface as integrated in a single solution, but also as a stand-alone solution for each feature and module. The evaluation process took into consideration the main users of the interface, which are the elderly people.

7.1. Graphical User Interface Evaluation

The Graphical User Interface module has been evaluated by a group of sixteen people aged between fifty and sixty-five, in the laboratory, and by a group of ten people aged between fifty-five and sixty-five, at their residence. During the evaluation each user used a computer with touch-enabled screen and a tablet (Apple iPads with 9.7-inch screen size or Samsung Galaxy Tab 4). The design of the module was very appreciated by the users, both on the computer and on the tablet.

The users appreciated, especially, the used icons which were found by the user as a very useful way that allows to rapidly identify the information location as well as the colors of the health icons which were found by the users as a nice and easy way to rapidly identify their health status. In addition, the big buttons, the graphical representations of the health data and the display mode of the health records history were very much appreciated by the users.

Besides the previous group of people, six people aged between forty and fifty evaluated the module using a mobile phone (Apple iPhone 7 or Samsung Galaxy S7). They appreciated the design of the module and the way it works on the mobile phones.

Although users used some touch commands, the purpose of those evaluations was to evaluate the GUI and not the touch commands.

7.2. Voice Module Evaluation

The voice module of the interface has been evaluated by a group of twenty-four people aged between fifty and sixty-five in the laboratory and by a group of twelve people aged between fifty-five and sixty-five in their residence.

All the users tested the voice module using the Romanian language, then half of the users tested the module again using the English language while the other half tested the module using the French language. For the Romanian language the users tested the module in the online mode of working, while they tested the module in both modes of working (online and offline) for the French and English languages. During the tests, two microphones were used (a Plantronics Voyager 5200 UC and a Conexant HD ISST microphone).

Therefore, the tests have been performed on a set of 7200 interactions between the user and the system formed of twenty-five different interactions for each user, as illustrated in Table7.1. In those tests, each user repeated twice the interaction for the Romanian language, each time using a different microphone, while each user repeated the interaction six times for the French or English languages: three times using each microphone, once in the offline mode of working and twice in the online mode of working (once using the Microsoft speech-to-text service as ASR solution and once using the Google speech-to-text service as ASR solution).

The evaluation procedure used the Levenshtein distances to calculate the differences between the results of the speech recognition and the original text phrase. A part of the results for the ASR tests that were performed in the laboratory are illustrated in Table 7.2, while a part of the results for the for the ASR tests that were performed at users' residence are illustrated in Table 7.3.

Table 7.1. Number of Interactions for the voice module tests

Language	Number of Users (A)	Number of Interactions/User (B)	Number of Repetitions in different settings (C)	Number of Interactions (A*B*C)
English	18	25	6	2700
French	18	25	6	2700
Romanian	36	25	2	1800
TOTAL				7200

The ASR test results illustrate that the Microsoft speech-to-text service and the Google speech-to-text service have similar performance for French and English with a small advantage for the Microsoft speech-to-text service. As expected, the results of the tests performed in the laboratory are better than those performed in the users' residence due to the noise presence in the environment at users' residence. The results also illustrate that the use of the Plantronics Voyager 5200 UC microphone improved in general the recognition of the users' speech commands. The improvement degree varied between the cases, it was significantly in the offline mode of working, it was important for the tests that took place at the users' residence and it was less important for the tests that took place in the laboratory. Despite the variation between the obtained results, all the results obtained during the ASR tests are satisfying.

To insure a correct test for the NLU implementation and to reduce the errors caused by a poor ASR recognition, the NLU tests were performed only with the utterance recorded in the laboratory using the Plantronics Voyager 5200 UC microphone. For the online mode of working, the utterance that was recognized by the ASR was sent to the LUIS platform for French and English or to the wit.ai platform for Romanian. On each platform, the utterance was processed and compared with the original text query. The results of the NLU tests for the online mode of working are very satisfying. For English, regarding the intent recognition, the percentage of the successful recognition was 82% and regarding the entity recognition, the percentage of the successful recognition was 75% (regardless of the used ASR solution). For French those percentages were 76% and 70% respectively (regardless of the used ASR solution), while for Romanian those percentages were 73% and 68% respectively. For the offline mode of working, the utterance that was recognized by the ASR were sent to the RASA NLU tool, in which the utterance was processed and compared with the original text query. The results of the NLU tests for the offline mode of working are satisfying, even if they are lower than the ones obtained for the online mode of working. For English, regarding the intent recognition, the percentage of the successful recognition was 76% and regarding the entity recognition, the percentage of the successful recognition was 69%, while for French those percentages were 65% and 57% respectively. In both modes of working, the tests illustrated that the intent interpretation, as well as the entity extraction, are related in a proportional way to the accuracy of the ASR.

Table 7.2. Part of the ASR tests results performed at the laboratory - A using Microsoft speech-to-text service and B using the Google speech-to-text service as ASR solution

Users' Command	Average Recognition Percentage (%)					
	Plantronics Mic.			Conexant Mic.		
	Online Mode		Offline Mode	Online Mode		Offline Mode
	A	B		A	B	
English						
How will the weather be this weekend?	99.84	99.51	90.32	98.15	98.21	79.88
Show my calendar	99.42	99.34	90.23	99.08	98.79	79.68
Display my calendar	99.18	99.00	89.97	98.95	98.34	78.14
What is my health status?	97.96	97.52	86.71	96.25	95.86	74.23
Display my blood pressure	99.00	98.85	87.79	98.85	98.42	75.51
How much have I walked today	97.24	97.35	85.98	95.67	95.51	73.28
How will the weather be today?	98.91	98.79	89.16	96.28	95.99	77.86
French						
Afficher mon calendrier	97.82	96.58	84.21	94.52	93.86	72.34
Quel est mon état de santé	95.28	94.72	81.65	91.96	91.08	70.16
Afficher ma tension artérielle	94.58	94.05	79.74	90.07	89.61	68.34
Combien j’ai marché aujourd'hui	93.31	92.81	77.16	88.96	88.52	66.85
Comment sera le temps aujourd'hui	95.44	94.96	80.34	92.23	91.82	69.44
Romanian						
Afișează-mi calendarul	N/A	96.52	N/A	N/A	92.78	N/A
Care este starea mea de sănătate	N/A	95.34	N/A	N/A	91.69	N/A
Arată-mi tensiunea arterială	N/A	91.57	N/A	N/A	86.95	N/A
Cât am mers astăzi	N/A	94.86	N/A	N/A	89.18	N/A
Cum va fi vremea astăzi	N/A	95.49	N/A	N/A	90.52	N/A

Table 7.3. Part of the ASR tests results performed in the users' residence - A using Microsoft speech-to-text service and B using the Google speech-to-text service as ASR solution

Users' Command	Average Recognition Percentage (%)					
	Plantronics Mic.			Conexant Mic.		
	Online Mode		Offline Mode	Online Mode		Offline Mode
	A	B		A	B	
English						
Display my calendar	96.11	94.60	77.23	93.08	92.46	56.25
What is my health status?	94.21	92.54	74.83	90.28	88.75	54.17
Display my blood pressure	93.47	91.96	76.49	89.73	90.05	60.32
How much have I walked today	92.24	91.23	76.21	88.61	87.17	55.46
How will the weather be today?	94.99	94.78	78.01	90.44	89.65	59.44
French						
Afficher mon calendrier	96.73	93.19	72.71	87.06	87.61	53.91
Quel est mon état de santé	92.87	91.88	68.95	85.27	83.92	51.26
Afficher ma tension artérielle	91.29	90.63	68.68	83.78	83.04	50.74
Combien j’ai marché aujourd'hui	90.53	88.94	67.94	82.03	81.73	50.08
Comment sera le temps aujourd'hui	92.07	90.87	72.26	84.94	83.89	51.82
Romanian						
Afișează-mi calendarul	N/A	92.22	N/A	N/A	85.92	N/A
Care este starea mea de sănătate	N/A	91.92	N/A	N/A	82.45	N/A
Arată-mi tensiunea arterială	N/A	88.96	N/A	N/A	79.16	N/A
Cât am mers astăzi	N/A	90.64	N/A	N/A	80.09	N/A
Cum va fi vremea astăzi	N/A	92.54	N/A	N/A	80.67	N/A

After extracting the intent and the entities from the recognized utterance, the message generated by the NLU solution is sent DM component which is the Azure Bot Service for the online mode of working or the Bot Builder tool for the offline mode of working. The DM component matches first the received user intent, then it matches the received entities, therefore a wrong detection of the intent in the NLU will lead to wrong interpretation of the user’s utterance, while an incomplete extraction of entities will lead to an incomplete answer/action from the system or it will make the system ask for the entities that are missing. Since the differences in the results of the ASR in the online mode of working are insignificant between the Microsoft speech-to-text service and the Google speech-to-text service, the tests of the DM component took into consideration for the online mode of working only the interactions that were performed using the Microsoft speech-to-text service as ASR solution for French and English and the Google speech-to-text service as ASR solution for Romanian. The tests of the DM component compared the reaction of the system with the reaction of the system that should take place as a response to the speech command of the user. The results of the tests are satisfying as illustrated in Table 7.4.

Table 7.4. DM tests results

Case	Average Percentage (%)							
	Plantronics Mic.				Conexant Mic.			
	Online Mode		Offline Mode		Online Mode		Offline Mode	
	Lab	Dom	Lab	Dom	Lab	Dom	Lab	Dom
English								
First case	73	68	67	60	71	65	61	51
Second case	18	20	19	21	19	22	20	22
Third case	5	7	8	11	6	8	12	18
Fourth case	4	5	6	8	4	5	7	9
French								
First case	69	67	61	55	68	65	55	46
Second case	19	20	20	21	18	18	19	21
Third case	7	8	12	15	9	11	17	22
Fourth case	5	5	7	9	5	6	9	11
Romanian								
First case	68	64	N/A	N/A	61	52	N/A	N/A
Second case	18	19	N/A	N/A	19	20	N/A	N/A
Third case	9	11	N/A	N/A	14	17	N/A	N/A
Fourth case	5	6	N/A	N/A	6	11	N/A	N/A

For the TTS component, after each session with the system, each user had filled a questionnaire regarding the degree of understanding of the phonetic output and its degree of clarity. The tests took place with the default setup of the TTS component (voice gender, the speed and the pitch of the voice). For each question, the user answered with a score from 0 to 10 in which 10 represents full understanding or excellent clearness. A part of the results for Romanian TTS tests are illustrated in Table 7.5. While a part of the results for French and English TTS tests, that were performed in the laboratory are illustrated in Table 7.6, and those performed at users' residence are illustrated in Table 7.7.

Table 7.5. Part of the Romanian TTS tests results

Phonetic Output (Romanian)	Average Score			
	Understood		Clearness	
	Lab	Dom	Lab	Dom
Ce măsură de sănătate doriți să luați?	8.50	7.50	8.00	7.25
Acestea sunt variațiile greutății voastre în ultima lună.	7.50	7.00	7.25	6.75
Acestea este calendarul săptămânal. Timpul vostru se umple.	7.75	7.25	7.75	7.25

Table 7.6. Part of the TTS tests results performed at the laboratory - A using Microsoft text-to-speech service and B using the Google text-to-speech service

Phonetic Output	Average Score					
	Understood			Clearness		
	Online Mode		Offline Mode	Online Mode		Offline Mode
	A	B		A	B	
English						
What measurement do you want to take?	10.00	10.00	9.75	9.50	9.50	9.50
Here are your weight variations during the past month.	9.75	10.00	9.50	9.25	9.50	9.25
Here is your weekly calendar. Your time is getting filled	10.00	9.75	9.75	9.50	9.25	9.25
French						
Quelle mesure voulez-vous prendre?	9.75	9.50	9.25	9.25	9.25	8.75
Voici les variations de votre poids au cours du dernier mois.	9.50	9.25	9.00	9.00	9.00	8.50
Votre calendrier hebdomadaire est affiché. Votre temps se rempli.	9.50	9.50	9.00	9.00	9.00	8.50

Table 7.7. Part of the TTS tests results performed at the users' residence - A using Microsoft text-to-speech service and B using the Google text-to-speech service

Phonetic Output	Average Score					
	Understood			Clearness		
	Online Mode		Offline Mode	Online Mode		Offline Mode
	A	B		A	B	
English						
What health measurement do you want to take?	10.00	10.00	9.50	9.50	9.50	9.25
Here are your weight variations during the past month.	9.50	9.75	9.25	9.25	9.25	9.00
Here is your weekly calendar. Your time is getting filled	10.00	10.00	9.75	9.25	9.50	9.00
French						
Quelle mesure voulez-vous prendre?	9.50	9.25	8.75	9.00	9.00	8.25
Voici les variations de votre poids au cours du dernier mois.	9.25	9.00	8.50	8.75	8.50	8.00
Votre calendrier hebdomadaire est affiché. Votre temps se rempli.	9.25	9.25	8.75	8.75	8.50	8.00

The TTS tests results illustrate that the Microsoft text-to-speech service and the Google text-to-speech service have similar performance for French and English with an insignificant advantage for the Microsoft text-to-speech service. As expected, the results of the tests performed at the laboratory are better than those performed in the users' residence due to the conditions of the environment.

The results of the TTS tests are satisfying for both modes of working. Regarding the tests that took place in the laboratory, for the online mode of working, the average understanding score of the spoken text was 9.85 from 10 for English, 9.50 for French and 8.05 for Romanian, while the clearness score of the spoken text was 9.45, 9.05 and 7.75 respectively. For the offline mode of working, the average understanding score of the spoken text was 9.70 for English and 9.05 for French, while the clearness score of the spoken text was 9.25 and 8.75 respectively. Regarding the tests that took place in the users' residence, for the online mode of working, the average understanding score of the spoken text was 9.65 for English, 9.25 for French and 7.65 for Romanian, while the clearness score of the spoken text was 9.25, 8.75 and 7.15 respectively. For the offline mode of working, the average understanding score of the spoken text was 9.50 English and 8.65 for French while the clearness score of the spoken text was 9.05 and 8.10 respectively.

Finally, the results of the tests for each component of the voice module are satisfying. The module was very welcomed by the users, especially the possibility to have multiple voice commands for the same order and the possibility to give the information needed for a command through a single speech interaction or through multiple speech interactions. The ability to adjust the gender, the speed and the pitch of the voice was also very welcomed by the users.

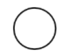
7.3. Gesture Module Evaluation

The gesture module of the interface has been evaluated by a group of twenty-two people aged between fifty and sixty-five in the laboratory and by a group of twelve people aged between fifty-five and sixty-five in their residence.

For the touch-based gestures: the users navigated between the different pages of the interface and gave commands to the system using touch interactions. The test used the same devices that were used for the GUI evaluation (a computer with touch-enabled screen and a tablet: an Apple iPads with 9.7-inch screen size or a Samsung Galaxy Tab 4). The users appreciated very much the possibility to interact with the system through touch-based commands. The users considered that the single touch commands are very easy to use. The system always reacted as the user was expecting during those interactions and the satisfaction of the users regarding the single touch commands is almost 100%. However, the users considered that the multi-touch commands are complicated. In many cases, that kind of commands got the user lost, since in the majority of cases, the user initiated a multi-touch command while in fact the user was intending to use a single touch command. Despite the high detection rate of the multi-touch commands, the multi-touch option is turned off in the default settings of the gesture module.

For the touch-free gestures: the users executed simple commands using their hands. During the tests six touch-free gestures were evaluated: toward left (L), toward right (R), toward up (U), toward down (D), circle (C) and push (P). A Microsoft Kinect V2 sensor was used to track the movement of the hand and each user repeated each gesture sixteen times. Consequently, a total of 2970 interactions have been performed from which 1980 interactions were performed in the laboratory and 990 interactions were performed at the users' residence; 495 interactions were performed for each gesture, respectively 330 and 165 interactions. The results are listed in Table 7.8.

Table 7.8. Touch-Free gesture recognition tests results

Gesture Name	Toward Left	Toward Right	Toward Up	Toward Down	Circle	Push
Gesture Form						
Laboratory						
True Detection	324	323	325	316	312	305
False Detection	6	7	5	14	18	25
Detection Percentage (%)	98.18	97.88	98.48	95.76	94.55	92.42
Users' Residence						
True Detection	152	150	154	141	134	131
False Detection	13	15	11	24	31	34
Detection Percentage (%)	92.12	90.91	93.33	85.45	81.21	79.39

The results of the touch-free gestures tests are satisfying. The results of the tests performed at the laboratory are better than those performed in the users' residence due to the light conditions in the environment at users' residence which affected the tracking of the hand movements. The confusion matrix for the tests that took place in the laboratory is illustrated in Table 7.9. The average gesture detection percentage in the laboratory was 96.21%. The confusion matrix for the tests that took place at the users' residence is illustrated in Table 7.10. The average gesture detection percentage at the users' residence was 87.07%. The combined average gesture detection percentage was 93.16%.

Table 7.9. The confusion matrix for the laboratory tests

	L	R	U	D	C	P
L	324	0	2	3	0	1
R	0	323	3	2	0	2
U	2	3	325	0	0	0
D	3	2	0	316	0	9
C	5	3	2	6	312	2
P	4	3	0	18	0	305

Table 7.10. The confusion matrix for the users' residence tests

	L	R	U	D	C	P
L	152	2	4	6	0	1
R	1	150	5	7	0	2
U	4	6	154	0	1	0
D	5	5	0	141	0	14
C	9	3	4	13	134	2
P	5	4	0	25	0	131

As it is illustrated in both matrices, there are no serious confusion cases. The most visible confusion cases are between the toward down and the push commands. Another visible confusion case is when a circle command is not performed right enough by the user, the command is sometimes confused with the toward left or toward down commands.

The users appreciated very much the possibility to interact with the system through touch-free commands especially when they don't have a device with-touch enabled screen near them. The users considered that the touch-free commands are easy to use. However, in the cases in which both types of gesture commands can be used, 93.94% of users preferred to use the touch-based commands while only 6.06% preferred to use the touch-free commands.

Beside the previous group of people, eight people aged between forty and fifty evaluated the module on eight mobile phones: four Apple iPhone 7 and four Samsung Galaxy S7. They appreciated the possibility to interact with the system through touch-based commands (single-touch and multi-touch).

7.4. Emotion Module Evaluation

The emotion module of the interface has been evaluated by a group of twenty people aged between fifty and sixty-five in the laboratory and by a group of ten people aged between fifty-five and sixty-five in their residence. A Microsoft Kinect V2 sensor was used to track the facial expression of the user in the offline mode of working, while a Logitech Socialize HD 1080 camera was used to track the facial expression of the user in the online mode of working. During the tests four emotions were evaluated: neutral, happiness, sadness and anger. Each user mimicked each emotion ten times for each mode of working. Consequently, a total of 2400 interactions have been performed. For each mode of working 1200 interactions from which 800 interactions were performed in the laboratory and 400 interactions were performed at the users' residence; 600 interactions were performed for each emotion, respectively 300, 200 and 100 interactions. The results are listed in Table 7.11.

Table 7.11. Emotion recognition tests results

Emotion	Online Mode				Offline Mode			
	Neutral	Happy	Sad	Anger	Neutral	Happy	Sad	Anger
Laboratory								
True Detection	188	197	184	182	181	188	176	174
False Detection	12	3	16	18	19	12	24	26
Detection Percentage (%)	94	99	92	91	91	94	88	87
Users' Residence								
True Detection	90	95	87	81	86	94	78	73
False Detection	10	5	13	19	14	6	22	27
Detection Percentage (%)	90	95	87	81	86	94	78	73

The results of the emotion recognition tests are satisfying. The results of the tests performed in the laboratory are better than those performed in the users' residence due to the light conditions in the environment at users' residence which affected the quality of the face recognition. The confusion matrix for the tests that took place in the laboratory is illustrated in Table 7.12, while the confusion matrix for the tests that took place at the users' residence is illustrated in Table 7.13. The average emotion detection percentage in the laboratory was 93.88% for the online mode of working and 89.88% for the offline mode of working. Those average percentages at the users' residence were 88.25% and 82.75% respectively. The combined average emotion detection percentages were 92.00% and 87.50% respectively for the online and offline modes of working.

Table 7.12. The confusion matrix for the laboratory tests (A: online mode, B: offline mode)

A	Neutral	Happy	Sad	Anger
Neutral	188	2	9	1
Happy	1	197	0	2
Sad	10	0	184	6
Anger	3	5	10	182

B	Neutral	Happy	Sad	Anger
Neutral	181	3	14	2
Happy	4	188	0	8
Sad	16	0	176	8
Anger	5	5	16	174

Table 7.13. The confusion matrix for the users' residence tests (A: online mode, B: offline mode)

A	Neutral	Happy	Sad	Anger
Neutral	90	2	6	2
Happy	2	95	0	3
Sad	8	0	87	5
Anger	4	6	9	81

B	Neutral	Happy	Sad	Anger
Neutral	86	4	8	2
Happy	2	94	0	4
Sad	11	1	78	10
Anger	7	7	13	73

As illustrated in the four previous matrices, there are no serious confusion cases. The most confusion cases are between the neutral and sad emotions as well as between anger and sad emotions. The users found the capability of the system to recognize their emotions as an interesting feature.

7.5. Features Evaluation

The personalization feature of the interface has been evaluated by a group of thirty people aged between fifty and sixty-five in the laboratory. Before the beginning of each test, a professional explained the different personalization capabilities that the interface has for each user. The professional assisted the user during the test in which the user adjusted different options multiple times. Each time the user changed one or multiple option(s), the user was checking the interface to evaluate the changes that took place. At the end of each session the user has filled a questionnaire regarding the importance of each personalization option that the user had tested. For each question, the user answered with a score from 0 to 5 in which 5 represents an indispensable option and 0 represents a useless option. The results of the personalization feature tests are satisfying. The users appreciated the majority of the options that were tested very much. The average score for the personalization options that were tested is 3.88, which is a good score. Moreover, if the two options that were considered by the users useless are eliminated, the average score increases to 4.22, which is a great score. The users were satisfied with the customization results of all the tested options.

The automatic adaptation feature of the interface has been evaluated by a group of twelve people aged between fifty-five and sixty-five in the laboratory. A professional assisted the user during the tests and before the start of each test the professional explained to the user the features that the user will test during the session. During the tests, each user evaluated the adaptation of the interface according to his/her own profile, activity, emotional status and to the environment conditions. Each user tested each of the previous adaptation cases five times. Consequently, a total of 240 adaptation cases have been evaluated, 60 evaluations for each adaptation case. At the end of each session the user filled a questionnaire regarding the importance of the tested adaptation case. For each question, the user answered with a score from 0 to 5 in which 5 represents a must have adaptive capability and 0 represents a useless adaptive capability. The users considered that the adaptation according to the user profile and according to the environment conditions (especially for the output type), as a must to have adaptive capabilities in the interface. They considered that the adaptation according to the user activity is an extremely important adaptive capability. Finally, they found that the adaptation according to the user emotional status is a welcomed adaptive capability. The users appreciated the majority of the adaptive capabilities that were tested very much. The average score for the adaptive capabilities that were tested by the users is 4.50, which is a great score. In addition to the previous tests, the interface has been tested in the laboratory on several platforms and on several devices that have different screen sizes: computers with and without touch-enabled screens and smartphones. The interface was fully responsive and compatible across all the tested devices and platforms. The adaptation according to the configuration of the system has been tested in the laboratory by changing the configuration of the system that integrates the interface, but also by integrating the interface into two other AAL systems that have similar functionalities: the IONIS system presented in [74] and the Mobile@Old system presented in [75]. The functionalities that were disabled in the initial system or those that were not integrated into the new systems, disappeared automatically from the interface together with all the related information and features. The results of the different automatic adaptation features tests are satisfying.

7.6. Whole Interface Evaluation

The interface has been evaluated by a group of ten people aged between fifty-five and sixty-five in the laboratory. Those people have never interacted with the system before and they did not participate in the tests of any feature or module that are implicated in this evaluation. In this evaluation, the different modules and features of the interface were integrated and tested together as a single solution.

At the beginning of the evaluation of each module or feature, a professional explained to the users the functionalities of the module or feature and clarified any uncertain situation for the user. The professional was always present during the evaluation of the different modules.

Each user evaluated the graphical user interface and the touch-based gestures on a computer with touch-enabled screen and on tablet (Apple iPads with 9.7-inch screen size or Samsung Galaxy Tab 4).

After that, each user evaluated on a big screen the touch-free gestures using a Microsoft Kinect V2 sensor. For this evaluation, the users gave different touch-free gesture commands to the system and repeated during a physical exercise session the movement of the trainer avatar.

After the evaluation of the touch-free gestures, each user tested the voice module by using a Plantronics Voyager 5200 UC microphone. All the users tested the voice module for the Romanian language. Then, for both modes of working of the voice module, half of the users tested the English language while the other half tested the French language. To reduce the time of the test, the online mode of working was tested just by using the Microsoft speech-to-text service as the ASR solution for the French and English languages, the Google speech-to-text service as the ASR solution for the Romanian language, the Microsoft text-to-speech service as the TTS solution for the French and English languages and the Google text-to-speech service as the TTS solution for the Romanian language.

After the evaluation of the voice module, each user tested the two modes of working for the emotion module by using a Logitech Socialize HD 1080 camera for the online mode of working and a Microsoft Kinect V2 sensor for the offline mode of working.

After the evaluation of the emotion module, each user evaluated the different adaptive capabilities of the interface. The evaluation procedure is similar to the procedure followed during the automatic adaptation evaluation session which was already described earlier.

After the evaluation of the adaptive capabilities, the users have interacted with the system through the Pepper robot as illustrated in Figure 7.1. They used the speech and touch commands to interact with Pepper. It was noticeable that the users were hesitating to interact with Pepper and the majority of the users that participated in this test disliked the idea to have a robot at their residence.

Then, the users were left to interact with the system through all the available modalities and in the way that they want. The users combined different input modalities to achieve a specific task. At the end, each user filled a questionnaire regarding the results, which are satisfying and similar to the results obtained during the evaluation sessions of each module and feature. Having the whole interface tested as a single solution assured the integrity of the interface and allowed the extraction of some useful information such as the preferred ways of interaction for the users, as illustrated in Figure 7.2 and the users' acceptance of the interface depending on the generated output types, as illustrated in Figure 7.3.

Figure 7.1. The interface displayed on the Pepper robot

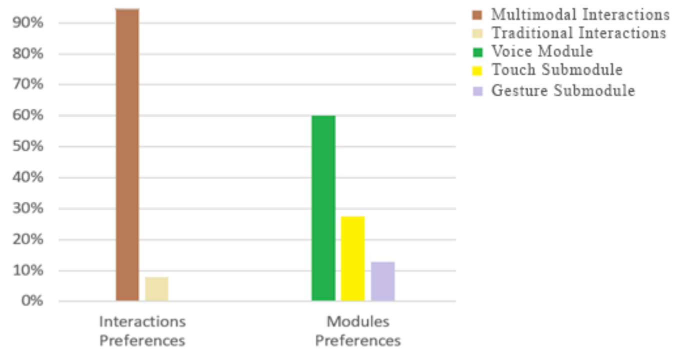


Figure 7.2. Preferences of the users regarding the available interaction options

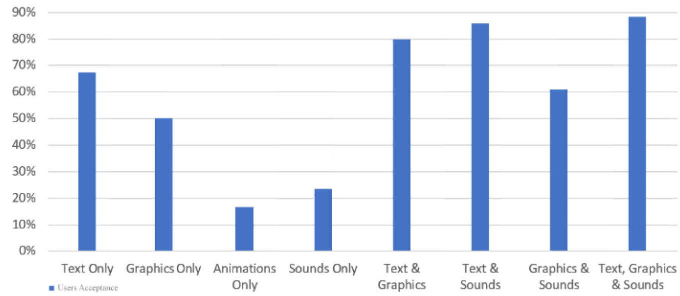


Figure 7.3. Users' acceptance of the interface depending on the generated output types

Chapter 8

Conclusions

This chapter concludes the work presented in the thesis.

The work in this thesis has resulted in the creation of a multimodal interface that can be integrated mainly to active and assisted living systems, but also to other kind of systems. The multimodal interface integrates a graphical user interface, a voice module, an emotion module and a gesture module that supports touch-based and touch-free gestures. The multimodal interface is multilingual, it works independently of any device or any operating system. Lastly, it integrates different automatic adaptive capabilities and customizable features.

In addition, the thesis provided an overview around the ambient intelligence field, the active and assisted living sector, and the CAMI project. This thesis also provided a detailed state of the art around the human-machine interaction field and the active and assisted living sector.

In what follows, a summary of our contributions is listed:

- Firstly, the creation of a multimodal interface for an active and assisted living system, that integrates four modules and that works across the different devices and platforms. Usually the active and assisted living systems integrates a multimodal interface that integrates two modules or a traditional interface.
- Creating the possibility for the user to switch the mode of working for the speech and emotion modules between the Online and Offline. Usually those modules have a single mode of working in different systems.
- Creating the possibility for the user to choose what speech recognition and speech synthesis solutions he/she wants to use for the online mode of working for the speech module. Usually a default choice is only available on the other systems.
- Enabling the speech interactions with the active and assisted living for the Romanian language. No other active and assisted living system supported the speech interactions for the Romanian language.
- Creating a wide range of adaptive features for the interface (automatic adaptive capabilities and customizable features). Usually, the systems integrate limited adaptive capabilities.
- Creating an emotion recognition module that makes recommendations regarding the physical exercises that a user should practice based on the emotional status of the user during each previous practice.
- Creating a portable interface that can be mainly integrated to active and assisted living systems, but also to other kind of systems. The multimodal interfaces are usually not portable.
- Creating the possibility for the interface to automatically communicate the preferences of the user to any 3rd party application that can be customized according to the user preferences.
- Testing extensively the different modules and features of the interface as stand-alone solution, but also integrated together into a single solution with an important number of users.
- Lastly, creating a set of rules that deal with the contradictory commands that are received at the same time. The set of rules also contains rules to ensure that any actions that are contradictory with any parameter that was set up by the caregiver is not executed despite the wide range of the automatic adaptive capabilities and the customizable features.

Many tasks are envisaged for the near future:

- At the beginning, the current implementation for the offline mode of working of the emotion module and of the touch-free gesture submodule depends on the presence of the Microsoft Kinect V2 sensor, the first envisaged task is to modify the implementation of the previous module and submodule such that the dependency on the Microsoft Sensor V2 is removed.
- Another planned task that targets to improve the implementation of the emotion module is to add to the capability to recognize the emotion from the voice of the user. The implementation will extract, in addition to the features of the user's facial expression, features from the user's voice to improve the emotion detection, but also to rely only on the voice of the user to detect the user's emotion in the cases in which the face of the user is not trackable by the system.
- Another envisaged task is composed of three stages and it targets the offline mode of working of the speech module: on the first stage, the work targets to improve the performance of the automatic speech recognition and the text-to-speech components; in the second stage, the work targets to extend the number of the recognized commands (to match those recognized in the online mode of working); while in the third stage, the work targets to add the support of the Romanian language (and maybe other languages).
- Another planned task is to enhance the implementation of the fusion engine by taking into account the contextual information that can be extracted from the environment.
- Another planned task is to enhance the implementation of the fusion engine by taking into account the contextual information that can be extracted from the environment, as well as to benefit more from the emotion module.
- Last but not least, an envisaged task is to automatically identify the users through their voices and/or faces in the cases of the active and assisted living systems or any systems that are used by multiple users, once the user is identified, the interface will have access to the information that is related to the identified user and it will adapt itself according to the preferences of the identified user.

List of Publications

The work that led to the thesis was published in nine research papers from which seven in conference proceedings that have been already indexed WoS and one in conference proceeding that is currently under WoS indexing:

1. A. Sorici, **I.A. Awada**, A. Kunnappilly, I. Mocanu, O. Cramariuc, L. Malicki, C. Seceleanu and A.M. Florea, "CAMI - An Integrated Architecture Solution for Improving Quality of Life of the Elderly", in: Ahmed M., Begum S. & Raad W. (Eds.) Internet of Things Technologies for HealthCare, HealthyIoT 2016, Vasteras, Sweden, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 187, Springer, Cham, DOI: 10.1007/978-3-319-51234-1_23, ISSN: 1867-8211, 2016, WOS: 000428954100023.

2. **I.A. Awada**, A. Codreanu, I. Mocanu, A.M. Florea and M. Apostu, "An Adaptive Multimodal Interface to Improve Elderly People's Rehabiliation Exercises", in: Proceedings of the 13th eLearning and Software for Education Conference (eLSE), Bucharest, Romania, Vol. 2, pp. 41-47, DOI: 10.12753/2066-026X-17-092, 2017.

3. **I.A. Awada**, I. Mocanu, A.M. Florea and B. Cramariuc, "Multimodal Interface for Elderly People", in: Proceedings of the 21st International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania, pp. 536-541, DOI: 10.1109/CSCS.2017.82, IEEE, 2017, WOS: 000449004400075.

4. **I.A. Awada**, I. Mocanu, S. Jecan, L. Rusu, A.M. Florea, O. Cramariuc and B. Cramariuc, "Mobile@Old - An Assistive Platform for Maintaining a Healthy Lifestyle for Elderly People", in: Proceedings of the 6th International Conference on E-Health and Bioengineering (EHB), Sinaia, Romania, pp. 591-594, DOI: 10.1109/EHB.2017.7995493, ISSN: 2575-5137, IEEE, 2017, WOS: 000445457500148.

5. A. Kunnappilly, A. Sorici, **I.A. Awada**, I. Mocanu, C. Seceleanu and A.M. Florea, "A Novel Integrated Architecture for Ambient Assisted Living Systems", in: Proceedings of the 41st Computer Society International Conference on Computers, Software and Applications (COMPSAC), Turin, Italy, pp. 465-472, DOI: 10.1109/COMPSAC.2017.28, ISSN: 0730-3157, IEEE, 2017, WOS: 000424861400061.

6. **I.A. Awada**, I. Mocanu, L. Rusu, R. Arba, A.M. Florea and B. Cramariuc, "Enhancing the Physical Activity of Older Adults Based on User Profiles", in: Proceedings of the 16th RoEduNet Conference: Networking in Education and Research (RoEduNet),

Targu Mures, Romania, pp. 120-125, DOI: 10.1109/ROEDUNET.2017.8123749, ISSN: 2068-1038, IEEE, 2017, WOS: 000425040000021.

7. **I.A. Awada**, O. Cramariuc, I. Mocanu, C. Seceleanu, A. Kunnappilly and A.M. Florea, "An End-User Perspective on the CAMI Ambient and Assisted Living Project", in: Proceedings of the 12th Annual International Technology, Education and Development Conference (INTED), Valencia, Spain, pp. 6776-6785, DOI: 10.21125/INTED.2018.1596, ISSN: 2340-1079, 2018, WOS: 000448704001118.

8. **I.A. Awada**, I. Mocanu and A.M. Florea, "Exploiting Multimodal Interfaces in eLearning Systems", in: Proceedings of the 14th eLearning & Software for Education Conference (eLSE), Bucharest, Romania, Vol. 2, pp. 174-181, DOI: 10.12753/2066-026X-18-094, 2018, WOS: 000467466800023.

9. **I.A. Awada**, I. Mocanu, D.I. Nastac, D. Benta and S. Radu, "Adaptive User Interface for Healthcare Application for People with Dementia", in the: Proceedings of the 17th RoEduNet Conference: Networking in Education and Research, Cluj-Napoca, Romania, pp. 96-100, DOI: 10.1109/ROEDUNET.2018.8514150, ISSN: 2247-5443, IEEE, 2018. *(Currently under WoS indexing)*

An additional research work was accepted to be published within the next few months as a book chapter in the "Recent Advances in Intelligent Assistive Technologies: Paradigms and Applications" series:

- **I.A. Awada**, I. Mocanu, A. Sorici, A.M. Florea, "An Integrated System for Improved Assisted Living of Elderly People", in: Recent Advances in Intelligent Assistive Technologies: Paradigms and Applications, H.N. Costin, B. Schuller & A.M. Florea (Eds.), Springer. (*to be published in 2019*)

In addition, the work that led to this thesis was presented in different scientific events:

- A presentation at the TiCIA workshop that took place in Cluj-Napoca, Romania, between 5th and 7th September 2015.
- One poster at the 2016 Ambient Assisted Living Forum that took place in St. Gallen, Switzerland, between 26th and 28th September 2016.
- Two posters at the 2017 Ambient Assisted Living Forum that took place in Coimbra, Portugal, between 2nd and 4th October 2017.
- One Poster at the Doctoral Day that was organized by UPB in the framework of the Semi-Centenary of the Faculty of Automatic Control and Computers that took place in Bucharest, Romania, between 11th and 14th October 2017.
- Two posters at the 2018 Ambient Assisted Living Forum that took place in Bilbao, Spain, between 24th and 26th September 2018.

During the work that led to this thesis, I have participated in the following research projects:

- CAMI: Artificial Intelligent Ecosystem for Self-Management and Sustainable Quality of Life in AAL - Active and Assistive Living project, AAL-2014-1-087 (No. 30/2015), 1.06.2015 – 30.11.2018.
- IONIS: NITICSplus - Indoor and Outdoor Solution for Dementia Challenges - Active and Assistive Living project - AAL2017-AAL-2016-074-IONIS-2 (No. 53/2.10.2017), 1.10.2017-31.03.2020.
- Mobile@Old: An Assistant for Elderly People based on Mobility Patterns - National project, PN-II-PT-PCCA-2013-4-22410 (No. 315/1.7.2014), 1.07.2014-30.09.2017.
- SPARC: Performance Services Assistance to Customers through Robotic Platforms - National project - PN-III-P2-2.1-BG-2016-0425 (No. 56BG/1.10.2016), 1.10.2016-30.09.2018.

Bibliography

1. United Nations World Population Ageing 2017 Report, New York: Department of Economic and Social Affairs, Population Division, 2017, available on the link: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf

2. J. Bohn, V. Coroamă, M. Langheinrich, F. Mattern and M. Rohs, "Social, Economic, and Ethical Implications of Ambient Intelligence and Ubiquitous Computing", in: Ambient Intelligence, Springer, Berlin, Germany, pp. 5-29, DOI: 10.1007/3-540-27139-2_2, 2005.

3. A.S. Crandall and D.J. Cook, "Current State of the Art of Smart Environments and Labs from an Ambient Assisted Living Point of View", in: F. Florez-Revuelta & A.A Chaaraoui (Eds.) Active and Assisted Living: Technologies and Applications, pp. 11-28, DOI: 10.1049/pbhe006e_ch2, August 2016.

4. M. Mohri, F. Pereira and M. Riley, "Speech Recognition with Weighted Finite-State Transducer", in: Computer Speech and Language Journal, Vol. 16, Iss. 1, pp. 69-88, DOI: 10.1006/csla.2001.0184, 2002.

5. L. Deng, J. Li, J.T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams, Y. Gong and A. Acero, "Recent Advances in Deep Learning for Speech Research at Microsoft", in: Proceedings of the 38th International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada, pp. 8604-8608, DOI: 10.1109/ICASSP.2013.6639345, IEEE, 2013.

6. Li D. and X. Li, "Machine Learning Paradigms for Speech Recognition: An Overview", in: Transactions on Audio, Speech and Language Processing Journal, Vol. 21, Iss. 5, pp. 1060-1089, DOI: 10.1109/TASL.2013.2244083, IEEE, 2013.

7. S. Toshniwal, T.N. Sainath, R.J. Weiss, B. Li, P. Moreno, E. Weinstein and K. Rao, "Multilingual Speech Recognition with a Single End-to-End Model", in: Proceedings of the 43rd International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada, pp. 4904 - 4908, DOI: 10.1109/ICASSP.2018.8461972, IEEE, 2018.

8. D.W. Griffin and J.S. Lim, "Signal Estimation from Modified Short-Time Fourier Transform", in: Proceedings of the 8th International Conference on Acoustics, Speech and Signal Processing (ICASSP), Boston, United States of America, pp. 804+807, DOI: 10.1109/ICASSP.1983.1172092, 1983.

9. A.W. Black, "Multilingual Speech Synthesis" in: T. Schultz & K. Kirchhoff (Eds.) Multilingual Speech Processing, pp. 207-231, DOI: 10.1016/B978-012088501-5/50010-X, ISBN: 978-0-12-088501-5, 2006.

10. H. Zen, K. Tokuda and A.W. Black, "Statistical Parametric Speech Synthesis", in: Speech Communication Journal, Vol. 51, Iss. 11, pp. 1039-1064, DOI: 10.1016/j.specom.2009.04.004, 2009.

11. P.F. Brown, P.V. DeSouza, R.L. Mercer, V.J. Della Pietra and J.C. Lai "Class-Based n-Gram Models of Natural Language", in: Computational Linguistics Journal, Vol. 18, Iss. 4, pp. 467-479, Cambridge, United States of America, 1992.

12. S.F. Chen and J. Goodman, "An Empirical Study of Smoothing Techniques for Language Modeling", in: Proceedings of the 34th Annual Meeting on Association for Computational Linguistics, Santa Cruz, United States of America, pp. 310-318, DOI: 10.3115/981863.981904, 1996.

13. B. Roark, M. Saraclar and M. Collins, "Discriminative n-Gram Language Modeling", in: Computer Speech & Language Journal, Vol. 21, Iss. 2, pp. 373-392, DOI: 10.1016/j.csl.2006.06.006, 2007.

14. J. Peckham, "A New Generation of Spoken Dialogue Systems: Results and Lessons from the SUNDIAL Project", in: Proceedings of the 3rd European Conference on Speech, Communication and Technology (Eurospeech), Berlin, Germany, pp. 33-40, 1993.

15. L. Lamel, S. Rosset, J.L. Gauvain and S. Bennacef, "The LIMSI ARISE System for Train Travel Information", in: Proceedings of the 24th International Conference on Acoustics, Speech and Signal Processing (ICASSP), Washington, United States of America, pp. 501-504, IEEE, 1999.

16. J. Chu-Carroll, "Form-based Reasoning for Mixed-Initiative Dialogue Management in Information-Query Systems", in: in: Proceedings of the 6th European Conference on Speech, Communication and Technology (Eurospeech), Budapest, Hungary, pp. 1519-1522, 1999.

17. S. Seneff and J. Polifroni, "Dialogue Management in the Mercury Fight Reservation System", in: Proceedings of the 2000 ANLP/NAACL Workshop on Conversational Systems, Stroudsburg, United States of America, pp. 11-16, 2000.

18. V. Zue, S. Seneff, J. Glass, J. Polifroni, C. Pao, T.J. Hazen and L. Hetherington, "Jupiter: A Telephone-Based Conversational Interface for Weather Information", in: Transactions on Speech and Audio Processing Journal, Vol. 8, pp. 85-96, IEEE, 2000.

19. O. Lemon and X. Liu, "Dude: A Dialogue and Understanding Development Environment, Mapping Business Process Models to Information State Update Dialogue Systems", in: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations, Stroudsburg, United States of America, pp. 99-102, 2006.

20. S. Varges, G. Riccardi and S. Quarteroni, “Persistent Information State in a Data-Centric Architecture”, in: Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue, Stroudsburg, United States of America, pp.68-71, 2008.

21. A. Mertens, N. Jochems, C.M. Schlick, D. Dünnebacke and J.H. Dornberg, “Design Pattern TRABING: Touchscreen-Based Input Technique or People Affected by Intention Tremor”, in: Proceedings of the 2nd ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS), Berlin, Germany, pp. 267-272, DOI: 10.1145/1822018.1822060, ACM, 2010.

22. A.M. Piper, R. Campbell and J.D. Hollan, “Exploring the Accessibility and Appeal of Surface Computing for Older Adult Health Care Support”, in: Proceedings of the 28th Conference on Human Factors in Computing Systems (CHI), Atlanta, United States of America, pp. 907-916, DOI: 10.1145/1753326.1753461, ACM, 2010.

23. C. Wacharamanotham, J. Hurtmanns, A. Mertens, M. Kronenbuerger, C. Schlick and J. Borchers, “Evaluating Swabbing: A Touchscreen Input Method for Elderly Users with Tremor”, in: Proceedings of the 29th Conference on Human Factors in Computing Systems (CHI), Vancouver, Canada, pp. 623-626, DOI: 10.1145/1978942.1979031, ACM, 2011.

24. M. Kobayashi, A. Hiyama, T. Miura, C. Asakawa, M. Hirose and T. Ifukube, “Elderly User Evaluation of Mobile Touchscreen Interactions”, in: P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque & M. Winckler (Eds.) Human-Computer Interaction - INTERACT 2011, Lecture Notes in Computer Science, Vol. 6946, Springer, Berlin, Heidelberg, pp. 83-99, DOI: 10.1007/978-3-642-23774-4_9, 2011.

25. L. Findlater, J.E. Froehlich, K. Fattal, J.O. Wobbrock and T. Dastyar, “Age-Related Differences in Performance with Touchscreens Compared to Traditional Mouse Input”, in: Proceedings of the 31th Conference on Human Factors in Computing Systems (CHI), Paris, France, pp. 343-346, DOI: 10.1145/2470654.2470703, ACM, 2013.

26. C. Stößel, H. Wandke and L. Blessing, “Gestural Interfaces for Elderly Users: Help or Hindrance?”, in: S. Kopp & I. Wachsmuth (Eds.), Gesture in Embodied Communication and Human-Computer Interaction - GW 2009, Lecture Notes in Computer Science, Vol. 5934. Springer, Berlin, Heidelberg, pp. 269-280, DOI: 10.1007/978-3-642-12553-9_24, 2010.

27. T. Apted, J. Kay and A. Quigley, “Tabletop sharing of digital photographs for the elderly”, in: Proceedings of the 24th Conference on Human Factors in Computing Systems (CHI), Quebec, Canada, pp. 781-790, DOI: 10.1145/1124772.1124887, ACM, 2006.

28. D. Facal, C. Buiza, M.F. González, J. Soldatos, T. Petsatodis, F. Talantzis, E. Urdaneta, V. Martínez and J.J. Yanguas, “Cognitive Games for Healthy Elderly People in a Multitouch

Screen", Proceedings of the 2009 International Congress on Digital Homes, Robotics and Telecare for All (IDRT4ALL), Barcelona, Spain, pp. 91-97, 2009.

29. N. Hollinworth and F. Hwang, "Investigating Familiar Interactions to Help Older Adults Learn Computer Applications More Easily", in: Proceedings of the 25th British Computer Society Conference on Human Computer Interaction (BCS-HCI), Swinton, United Kingdom, pp. 473-478, 2011.

30. A. Kendin, "Some Relationships Between Body Motion and Speech" in: A. Seigman & B. Pope (Eds.) Studies in Dyadic Communication, pp. 177-216, 1972.

31. A. Kendin, "Gesture and Speech: Two Aspects of the Process of Utterance", In: M.R. Key (Eds.) Nonverbal Communication and Language, The Hague: Mouton, pp. 207-227, 1980.

32. J. Bressem and S. Ladewing, "Rethinking Gesture Phases: Articulatory Features of Gestural Movement?", in: Journal of the International Association for Semiotic Studies (Semiotica), Iss.184, pp. 53-91, DOI: 10.1515/semi.2011.022, 2011.

33. M. Kipp, "Gesture Generation by Imitation: From Human Behavior to Computer Character Animation", Dissertation.com, Boca Raton, United Stated of America, ISBN: 1-58112-255-1, 2004.

34. U. Gut and J.T. Milde, "Annotation and Analysis of Conversational Gestures in the TASX Environment", in: Künstliche Intelligenz Journal (KI), Vol. 17, Iss. 4, pp. 34-49, 2003.

35. M.E. Sargin, Y. Yemez, E. Erzin and A.M. Tekalp, "Analysis of Head Gesture and Prosody Patterns for Prosody-Driven Head-Gesture Animation", in: Transactions on Pattern Analysis and Machine Intelligence Journal, Vol. 30, Iss. 8, pp. 1330-1345, DOI: 10.1109/TPAMI.2007.70797, IEEE, 2008.

36. G. Galanakis, P. Katsifarakis, X. Zabulis and I. Adami, "Recognition of Simple Head Gestures Based on Head Pose Estimation Analysis", in: Proceedings of the 4th International Conference on Ambient Computing, Applications, Services and Technologies (AMBIENT), Rome, Italy, pp. 88-96, ISBN: 978-1-61208-356-8, 2014.

37. C. Lelardeux, D. Panzoli, J. Alvarez and P. Lagarrigue, "Serious Game, Simulateur, Serious Play: État de l'Art pour la Formation en Santé", in: Proceedings of the 1st SeGaMed Conference, Nice, France, pp. 27-38, 2012.

38. I. Di Loreto, B. Lange, A. Seilles, S. Andary and W. Dyce, "Game Design for All: The Example of Hammer and Planks", in: Proceedings of the 4th International Conference on Serious Games Development and Applications (SGDA), Trondheim, Norway, pp. 70-75, DOI: doi.org/10.1007/978-3-642-40790-1_7, 2013.

39. G.R. Vineetha, C. Sreeji and J. Lentin, "Face Expression Detection Using Microsoft Kinect with the Help of Artificial Neural Network", in: Online Proceedings on Trends in

Innovative Computing 2012 - Intelligent Systems Design, pp. 176-180, ISSN: 2150-7996, 2012.

40. A.E. Youssef, S.F. Aly, A.S. Ibrahim and A.L. Abbott, “Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD Therapeutic Aid”, in: International Journal of Modeling and Optimization, Vol. 3, Iss. 2, pp. 112-115, 2013.
41. P. Lemaire, L. Chen, M. Ardabilian and M. Daoudi, “Fully Automatic 3D Facial Expression Recognition using Differential Mean Curvature Maps and Histograms of Oriented Gradients”, in: Proceedings of the 10th International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, pp. 1-7, DOI: 10.1109/FG.2013.6553821, IEEE, 2013.
42. F. Loi, J.G. Vaidya and S. Paradiso, Recognition of Emotion from Body Language Among Patients with Unipolar Depression”, in: Psychiatry Research Journal, Vol. 209, Iss. 1, pp. 40-49, DOI: 10.1016/j.psychres.2013.03.001, 2013.
43. L. Kiforenko and D. Kraft, “Emotion Recognition through Body Language using RGB-D Sensor”, in: Proceedings of the 11th International Conference on Computer Vision Theory and Applications (VISAPP), Vienna, Austria, pp. 398-405, DOI: 10.5220/0005783403980405, 2016
44. S. Yacoub, S. Simske, X. Lin and J. Burns, “Recognition of emotions in interactive voice response systems”, in: Proceedings of the 8th European conference on speech communication and technology (EUROSPEECH), Geneva, Switzerland, pp. 729-732, September 2003.
45. L. McCauley and S. D’Mello, “MIKI: A Speech Enabled Intelligent Kiosk”, in: J. Gratch, M. Young, R. Aylett, D. Ballin & P. Olivier (Eds.) Intelligent Virtual Agents (IVA), Lecture Notes in Computer Science, Vol. 4133, Springer, pp. 132-144, 2006.
46. M. Johnston and S. Bangalore, “MATCHKiosk: A Multimodal Interactive City Guide”, in: Proceedings of the 42nd Annual Meeting of the. Association for Computational Linguistics (ACL) Interactive Poster and Demonstration Session, Barcelona, Spain, pp. 222-225, DOI: 10.3115/1219044.1219077, 2004.
47. K. Saroha, S. Sharma and G. Bhatia, “Human Computer Interaction: An intellectual approach”, in: International Journal of Computer Science and Management Studies (IJCSMS), Vol. 11, pp. 147-154, 2011.
48. R. Raisamo, “A Multimodal User Interface for Public Information Kiosks”, in: Proceedings of the 2nd Perceptual User Interfaces (PUI) Workshop, San Francisco, United States of America, pp. 7-12, 1998.

49. R.A. Bolt, “Put-that-there: Voice and gesture at the graphics interface”, in: ACM SIGGRAPH Computer Graphics Journal, Vol. 14, Iss. 3, pp. 262-270, ACM, DOI: 10.1145/965105.80750, 1980.

50. P. Biswas and P. Langdon, “A new input system for disabled users involving eye gaze tracker and scanning interface”, in: Journal of Assistive Technologies, Vol. 5, Iss. 2, pp. 58-66, DOI: 10.1108/17549451111149269, June 2011.

51. G. Lamprinakos, E. Kosmatos, D. Kaklamani and I. Venieris, “An Integrated Architecture for Remote Healthcare Monitoring”, in: Proceedings of the 14th Panhellenic Conference on Informatics, Tripoli, Greece, pp. 12-15, DOI: 10.1109/PCI.2010.20, September 2010.

52. M.R. Tazari, F. Furfari, J.P.L. Ramos and E. Ferro, “The persona service platform for AAL spaces”, in: Handbook of Ambient Intelligence and Smart Environments, Springer, pp. 1171–1199, DOI: 10.1007/978-0-387-93808-0_43, 2010.

53. Active and Assisted Living Program: NITICS, retrieved from the project page on the Official AAL Website | URL: http://www.aal-europe.eu/projects/nitics/, accessed on: 2018-08-21.

54. Retrieved from the SOCIABLE project page on the JoinUp platform | URL: https://joinup.ec.europa.eu/document/sociable-project-sociable, accessed on: 2018-08-21.

55. C. Azevedo, C. Chesta, J. Coelho, D. Dimola, C. Duarte, M. Manca, J. Nordvik, F. Paterno, A. Sanders and C. Santoro, “Towards a Platform for Persuading Older Adults to Adopt Healthy Behaviors”, in: R. Orji, M. Reisinger, M. Busch, A. Dijkstra, M. Kaptein & E. Mattheiss (Eds.): Proceedings of the 2nd International Workshop on Personalization in Persuasive Technology, pp. 50-56, Amsterdam, Netherlands, 2017.

56. CareWell Official Website, retrieved from the URL: www.carewell-project.eu, accessed on: 2018-08-21.

57. Active and Assisted Living Program: healthy@work, retrieved from the project page on the Official AAL Website | URL: http://www.aal-europe.eu/projects/healthywork/, accessed on: 2018-08-21.

58. R. Planinc, M. Hödlmoser and M. Kampel, “Enhancing the Wellbeing at the Workplace”, in: Proceedings of the 7th International Conference on eHealth, Telemedicine, and Social Medicine (eTelemed), pp. 213-216, Lisbon, Portugal, February 2015.

59. V. Giannoglou, K. Smagas, E. Valari and E. Stylianidis, “Elders-Up! An Adaptive System for Enabling Knowledge Transfer from Senior Adults to Small Companies”, in: Proceedings of the 22nd International Conference on Virtual System and Multimedia

(VSMM), Kuala Lumpur, Malaysia, pp. 17-23, DOI: 10.1109/VSMM.2016.7863163, October 2016.

60. Active and Assisted Living Program: CAMI (AAL-2014-1-087) Official Website, retrieved from the URL: www.camiproject.eu, accessed on: 2018-08-22.

61. I.A. Awada, I. Mocanu, A.M. Florea and B. Cramariuc, "Multimodal Interface for Elderly People", in: Proceedings of the 21st International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania, pp. 536-541, DOI: 10.1109/CSCS.2017.82, IEEE, 2017.

62. C. Dood, R. Athauda and M.T.P. Adam, "Designing User Interfaces for the Elderly: A systematic Literature Review", in: Proceedings of the 2017 Australasian Conference on Information Systems (ACIS), Hobart, Australia, December 2017.

63. B. Dumas, D. Lalanne and S. Oviatt, "Multimodal Interfaces: A Survey of Principles, Models and Frameworks", in: D. Lalanne & J. Kohlas (Eds.) Human Machine Interaction - Lecture Notes in Computer Science, Springer, Berlin, Vol. 5440, pp. 3-26, DOI: 10.1007/978-3-642-00437-7_1, 2009.

64. D.A. Norman, in: "The Design of Everyday Things", Basic Books, New York, United States of America, ISBN: 10:0465067107, 2002.

65. H.C. Lukaski, "Commentary: Body mass index persists as a sensible beginning to comprehensive risk assessment", in: International Journal of Epidemiology, Vol. 43, Iss. 3, pp. 669-671, DOI: 10.1093/ije/dyu059, 2014.

66. R.M. Carey and P.K. Whelton, "Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Synopsis of the 2017 American College of Cardiology/American Heart Association Hypertension Guideline", in: Annals of Internal Medicine Journal, Vol. 168, Iss. 5, pp. 351-358, DOI: 10.7326/M17-3203, March 2018.

67. J. Hart, "Normal resting pulse rate ranges", in: Journal of Nursing Education and Practice, Vol. 5, Iss. 8, pp.95-98, DOI: 10.5430/jnep.v5n8p95, 2015.

68. C. Tudor-Locke, C.L. Craig, Y. Aoyagi, R.C. Bell, K.A. Croteau, I. De Bourdeaudhuij, B. Ewald, A.W. Gardner, Y. Hatano, L.D. Lutes, S.M. Matsudo, F.A. Ramirez-Marrero, L.Q. Rogers, D.A. Rowe, M.D. Schmidt, M.A. Tully and S.N. Blair, "How many steps/day are enough? For older adults and special populations", in: International Journal of Behavioral Nutrition and Physical Activity, Vol. 8, Iss. 80, Doi:10.1186/1479-5868-8-80, 2011.

69. M. Hirshkowitz, K. Whiton, S.M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, E.S. Katz, L. Kheirandish-Gozal, D.N. Neubauer, A.E. O'Donnell, M. Ohayon, J. Peever, R. Rawding, R.C. Sachdeva, B. Setters, M.V. Vitiello, J.C. Ware, and P.J. Adams Hillard, "National Sleep Foundation's sleep time duration recommendations: methodology

and results summary”, in: Sleep Health: Journal of the National Sleep Foundation, Vol. 1, Iss. 1, pp. 40-43, DOI: 10.1016/j.sleh.2014.12.010, 2015.

70. I.A. Awada, O. Cramariuc, I. Mocanu, C. Seceleanu, A. Kunnappilly and A.M. Florea, “An End-User Perspective on the CAMI Ambient and Assisted Living Project”, in: Proceedings of the 12th Annual International Technology, Education and Development Conference (INTED), Valencia, Spain, pp. 6776-6785, DOI: 10.21125/INTED.2018.1596, ISSN: 2340-1079, 2018.

71. I.A. Awada, I. Mocanu and A.M. Florea, “Exploiting Multimodal Interfaces in eLearning Systems”, in: Proceedings of the 14th eLearning & Software for Education Conference (eLSE), Bucharest, Romania, Vol. 2, pp. 174-181, DOI: 10.12753/2066-026X-18-094, 2018.

72. I.A. Awada, I. Mocanu, S. Jecan, L. Rusu, A.M. Florea, O. Cramariuc and B. Cramariuc, “Mobile@Old - An Assistive Platform for Maintaining a Healthy Lifestyle for Elderly People”, in: Proceedings of the 6th International Conference on E-Health and Bioengineering (EHB), Sinaia, Romania, pp. 591-594, DOI: 10.1109/EHB.2017.7995493, ISSN: 2575-5137, IEEE, 2017.

73. I.A. Awada, I. Mocanu, L. Rusu, R. Arba, A.M. Florea and B. Cramariuc, “Enhancing the Physical Activity of Older Adults Based on User Profiles”, in: Proceedings of the 16th RoEduNet Conference: Networking in Education and Research (RoEduNet), Targu Mures, Romania, pp. 120-125, DOI: 10.1109/ROEDUNET.2017.8123749, ISSN: 2068-1038, IEEE, 2017.

74. I.A. Awada, I. Mocanu, D.I. Nastac, D. Benta and S. Radu, “Adaptive User Interface for Healthcare Application for People with Dementia”, in the: Proceedings of the 17th RoEduNet Conference: Networking in Education and Research, Cluj-Napoca, Romania, pp. 96-100, DOI: 10.1109/ROEDUNET.2018.8514150, ISSN: 2247-5443, IEEE, 2018.

75. L. Rusu, S. Jecan and D. Sitar, “Mobile@Old an ADL Solution”, in: W. Abramowicz (Eds.) Business Information Systems Workshops (BIS) - Lecture Notes in Business Information Processing, Vol. 228, pp. 173-184, DOI: 10.1007/978-3-319-26762-3_16, Springer, Cham.