Universidade de Aveiro
2022

Fábio Ribeiro Nunes

What a Nice Gesture: Supporting the Communication of People with Aphasia for the in-Bed Scenario

What a Nice Gesture: Suporte à Comunicação de Pessoas com Afasia no Cenário da Cama

Universidade de Aveiro
2022

Fábio Nunes

What a Nice Gesture: Supporting the Communication of People with Aphasia for the in-Bed Scenario

What a Nice Gesture: Suporte à Comunicação de Pessoas com Afasia no Cenário da Cama

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Doutor Samuel de Sousa Silva, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro, e da Doutora Ana Patrícia Oliveira Ferreira da Rocha, Investigadora do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

This work was supported by EU and national funds through the Portuguese Foundation for Science and Technology (FCT), in the context of the AAL APH-ALARM project (AAL/0006/2019).

o júri / the jury

presidente / president	**Professor Doutor Paulo Miguel de Jesus Dias** Professor Auxiliar, Universidade de Aveiro
vogais / examiners committee	**Professor Doutor Daniel Filipe Albuquerque** Professor Adjunto, Instituto Politécnico de Viseu - Escola Superior de Tecnologia
	Professor Doutor Samuel de Sousa Silva Professor Auxiliar em Regime Laboral, Universidade de Aveiro

agradecimentos / acknowledgements

Gostaria de agradecer à minha família pelo apoio incondicional que me deram ao longo deste percurso académico. Um grande agradecimento aos meus orientadores Samuel Silva e Ana Rocha por todas as palavras de confiança e apoio constante em todos os passos do desenvolvimento da dissertação. Agradeço também a todos os participantes das experiências realizadas. E um obrigado final a todos os que estiveram presentes e suavizaram os momentos de stress, tanto através de palavras como de gestos. Obrigado!

Palavras Chave

Comunicação Remota, Afasia, Acessibilidade, Gestos, Cenário da Cama, Ambientes Inteligentes

Resumo

A comunicação é uma parte essencial da vida e, quando afectada, como acontece com as pessoas com Afasia, tem um impacto severo na sua qualidade de vida. Abordagens de Comunicação Aumentativa e Alternativa (CAA) visam ajudar pessoas com dificuldades de comunicação. No entanto, nem todas as partes do dia são facilmente cobertas por estas soluções, tais como quando se está deitado na cama.

Nesta dissertação, o objetivo é desenvolver uma solução que permite comunicação bilateral remota entre uma pessoa com Afasia (PCA) e outras pessoas (por exemplo, o cuidador), quando o primeiro está deitado na cama, sozinho. Este trabalho foi realizado no âmbito do projeto APH-ALARM e evolui trabalhos anteriores dentro do mesmo projeto.

Para alcançar este objetivo, colaborámos com peritos no domínio para obter informação relevante sobre as necessidades e motivações de uma PCA e para avaliar a nossa proposta conceptual inicial. Estes resultados traduzem-se numa proposta do sistema refinada, podendo também ser considerados como um recurso em outros trabalhos nos quais ajudar pessoas com afasia seja um dos focos principais.

Como parte do sistema, implementámos um assistente que faz a mediação da comunicação com base em perguntas de Sim/Não com o objetivo de recolher informação antes de enviar uma mensagem ao cuidador. Estas perguntas são apresentadas audiovisualmente à PCA e respondidas através de gestos. Os gestos são reconhecidos utilizando um modelo que classifica gestos que podem ser executados em diferentes posturas e utilizando qualquer um dos braços. Considerando um conjunto de dados correspondente a 8 sujeitos, este modelo alcançou um F1-score de 99% e 93% para os casos dependentes e independentes do sujeito, respetivamente.

O cenário do quarto pode ser visto como o primeiro passo de uma visão ambiciosa de apoio em todas as divisões da casa, não se limitando apenas a afásicos, mostrando potencial para ajudar qualquer pessoa com dificuldades de fala.

Keywords Remote communication, Aphasia, Accessibility, Gestures, In-bed scenario, Smart environments.

Abstract

Communication is an essential part of life and, when affected, as it happens to people with Aphasia, it severely impacts their quality of life. Augmentative and Alternative Communication (AAC) approaches aim to aid people with their communication disabilities. However, not all parts of the day are easily covered by these solutions, e.g., when lying in bed.

In this dissertation, we aim at developing a solution that allows for two-way remote communication between a person with Aphasia (PWA) and other people (e.g., caregiver), while the former is lying in bed, alone. This work was carried out in the scope of the APH-ALARM project and builds on previous works within the same project.

To address this goal, we collaborated with domain experts to obtain relevant information about the needs and motivations of a PWA and to evaluate our initial concept proposal. These results translated into a refined proposal of the system, and can also be considered an asset in other works with aiding people with aphasia as the main focus.

As part of the system, we implemented an assistant that mediates communication relying on Yes/No questions with the purpose of gathering information before sending a message to the caregiver. These questions are presented audiovisually to the PWA and answered through gestures. The gestures are recognized using a model that classifies gestures that can be executed in different postures and using either arm. Considering a dataset corresponding to 8 subjects, this model achieved an F1-score of 99% and 93% for the subject dependent and independent cases, respectively.

The bedroom scenario can be seen as the first step of an ambitious vision of providing support in every division of a home, while also not being limited to only aphasics, showing potential to aid anyone with speech difficulties.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context and Motivation

The human being is known for being a social species, communicating to each other regularly in different ways. With communication being such an important aspect of a person's life, when one is unable to express information, feelings and needs, relationships are affected and mental health is impacted negatively out of loneliness and frustration [4, 5, 6]. Aphasia falls under the category of a communication impairment disability, affecting a variable range of skills regarding expression, comprehension, and/or verbal and written language [22]. Aphasia has been identified in around one-third of people that have suffered a stroke [7], with 180.000 Americans acquiring the disorder every year [8].

Given these real situations, Augmentative and Alternative Communication (AAC) solutions have been introduced as a gap-closer between the average person and the person with communication difficulties. Examples such as touchscreen applications based on pictograms and solutions integrating imaging or mechanical methods are typical proposed tools towards aiding communication, where technology had a big role in propelling the amount of currently existing solutions [9]. Nonetheless, when looking for solutions that tackle the bedroom scenario, no solution was considered adequate.

In this work, we consider the bedroom scenario and aim to develop a minimally intrusive system that allows a person with aphasia (PWA) to communicate with someone who is not present. This can be seen as a first step of an ambitious vision of providing support in every division of the house, while starting from a place of primacy. The scenario decision is also reinforced when taking into account the different obstacles one may face in the bedroom, such as something happening in the middle of the night, search of safety and emotional assurance to the family and themselves, search of a higher level of independency, etc. Using a smartphone is not always the preferred choice when one lacks fine motor skills, has one side of the body paralyzed or is digitally illiterate, and sometimes the smartphone is just not available due to matters of charging for example.

The work is developed in the scope of the AAL project APH-ALARM - Comprehensive safety solution for people with Aphasia (AAL/0006/2019)[1]. The project aims at providing solutions that contribute to an increased sense of safety and independence for generally older people with disabilities that stem from a stroke, such as aphasia, epilepsy and/or side-paralysis. An application for a smartphone in conjunction with hardware attached to the user

[1] http://www.aal-europe.eu/projects/aph-alarm/

bed is in development, aiding communication through the use of pictograms and alerts using activity detection. The use of sensors to support a person lying in bed communicating with a caregiver or family member is also under research.

In this context, this work looks to implement a proof-of-concept system for a person with aphasia to use while lying in bed, through use of sensors.

The current system builds on top and brings together previous work and research done in the context of the APH-ALARM project, making use of the smartwatch based gesture recognition prototype [1], and taking advantage of the smartwatch and smartphone applications built towards the gestures recordings and towards communicating with the caregiver.

1.2 Challenges

When pondering about assistive communication solutions for people with communication difficulties, one has to consider the challenges that come with the context presented in this work. In a world surrounded by technology, privacy can be a raised concern by the individual since some methods to aid communication make use of cameras. Allowing interaction with a system while avoiding the privacy issues that come with cameras is an identified challenge. Another identified need when aiding a person with a disability is the constant availability of the aiding tool, which means one has to take into account variables such as lighting possibly affecting this tool.

As mentioned in the previous section, a person with aphasia (PWA) has a communication impairment disability, which means that any communication with the PWA should remain as simple as possible, while staying relevant. With the possibility of using gestures as input, one has to consider the possiblity of a PWA not being able to memorize many gestures, and a lack of fine motor skills due to possible physical complications that might come with aphasia. Nonetheless, a wide range of communication is still ideal even with a small set of gestures available, and the communication itself requires relevancy.

In addition to what was already mentioned, the scenario of the bed also brings obstacles such as the gestures having to be feasible while both sitting and laying down, and taking into account the blankets possibly obstructing physical movement.

1.3 Objectives

The main goal of this work is to create a system that works as a remote communication facilitator between a person with aphasia that is in bed, and another person with the role of caregiver. Towards this, the understanding and use of the previous work towards gesture recognition in the context of the APH-ALARM project is recommended [1, 10], and applied most efficiently when taking into account the needs of a PWA.

The objectives for this work are:

- Perform a revision of the literature regarding Aphasia and the tools that aim to aid people with aphasia communicate, to understand and identify the challenges thoroughly;
- Master the gesture recognition research carried out in the context of the APH-ALARM project in order to shape the development of the system;
- Propose methods for communication by the PWA and understand what information to be conveyed by the PWA is relevant towards appropriate assistance;

- Implement a proof-of-concept with a scalable design.

Another important high-level objective is providing to both the PWA and the caregiver higher levels of reassurance and comfort in their life, while dealing with the obstacles presented when one has aphasia.

1.4 Publications

Related to the lines of work described in this dissertation, the attained advances contributed to evolve ongoing work and led to the participation as a co-author in two scientific publications: 1) a paper [11] entitled "Toward Supporting Communication for People with Aphasia: The In-Bed Scenario" that demonstrates an initial prototype developed in the context of the APH-ALARM project, resulting from the work of Afonso Guimarães [1], which was presented at the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '22); and 2) a poster entitled "Supporting Communication for People with Aphasia While Lying in Bed", which was presented at the European Week for Active and Healthy Ageing (EWAHA).

An article entitled "Gesture-Based Communication for People With Aphasia While in Bed", describing the conceptualization and first prototype for the proposed solution, is in the works to be submitted to an international conference.

1.5 Document Structure

This document is divided into five chapters, excluding this one:

- **Chapter 2** (Background and Related Work) - The literature review performed in order to understand the full scope of the project is presented here. The understanding of aphasia, augmentative and alternative communication tools, gesture recognition methods, and user-centered design are the four main topics. An analysis of related work, which informed and influenced the decisions towards tackling the challenges of this work is also included.

- **Chapter 3** (Concept Definition and Refinement) - In this chapter, we make use of Personas and Scenarios, which are part of the UCD methodology, and close the problem definition phase by having a discussion and evaluation section of an initial concept proposal and low fidelity sketches of the system. From this, we extract the requirements of the system, and make a final proposal of the system, which is the outcome of all the information gathered so far.

- **Chapter 4** (Gesture Interaction) - This chapter presents the steps taken towards gesture interaction for the communication support system, including a model for gesture recognition. An experiment carried out to evaluate the model is described, including details on data collection and the achieved results. This chapter also includes a description of the applications that enabled offline data acquisition.

- **Chapter 5** (Supporting Communication in the Bed-Scenario) - This chapter includes all the information regarding the implementation of the prototype. This includes a gesture input modality, an assistant, a smartphone application, output modalities and an interaction manager.

- **Chapter 6** (Conclusions) - A summary of the carried out work is presented in this chapter, as well as the final observations towards the implemented system, and a discussion of possible future work.

Chapter 2

Background and Related Work

This section will showcase all the research done to support the decisions made in this work. We start by describing the impact that Aphasia has in the life of a person, and identify the challenges that need to be tackled related to the usability of the system. We will dive into augmentative alternative communication (AAC) concepts, listing the modalities available and the scenarios they cover. Gesture recognition is the following topic and it is followed by a guided discussion of AAC and Gesture Recognition regarding the APH-ALARM project. Given the importance of having a user-friendly system, we will also dive into the User-Centered Design methodology as well.

2.1 Aphasia

Aphasia can be defined as as acquired language impairment that usually stems from a brain injury or a stroke. As presented in Figure 2.1, there are different subtypes of Aphasia that affect different aspects, such as comprehension and expression, both verbally and written.

The Boston classification system is regarded as the most used [6], and relates the subtype of aphasia to specific areas of the brain [12]. They are mainly categorized as fluent and non-fluent types, where fluency is considered the quality with which one is able to articulate himself. Some of the characteristics that represent fluency are the amount of words per minute, the effort committed to speak, phrase length used and intonation [13]. To understand in which way they affect people, they can be described the following way [6, 12]:

Fluent

- **Anomic** - This type of aphasia is considered the least severe. Most language skills are somewhat preserved, but there's difficulty in finding the right words. Speech can have some hesitations due to this naming problem. Some describe it as amnesia for words. Intact comprenhesion allows for the recognition and correction attempt of mistakes [14].
- **Conduction** - Those who suffer from conduction aphasia retain most of their comprehension skills and speech flow. The problems lay in the repetition and naming skills, including regular phonemic paraphasia (unintended substitution of sound for similar one). There's a disproportional difference in spontaneous speech, and speech that needs to be repeated, but just like in anomic aphasia, mistakes can be recognized and attempted to be fixed.

- **Wernicke** - The diagnosis of Wernicke implies that the person has fludity of speech, but is impaired on all skills related to comprehension, naming and repetion. This means that even though the speech can have a good rhythm, it will still be lacking meaning. There's no mistake awareness and the paraphasias are abundant, making it a severe type of fluent aphasia.
- **Transcortical sensory** - Comparable to Wernicke in most impairments except repetition, which is preserved. A different dysfunction appears called echolalia, which is defined in the dictionary as "the tendency to repeat mechanically words just spoken by another person" [15].

Non-Fluent

- **Broca's** - Broca's aphasia allows for some comprehension of language, mostly short sentences, but impairs all speech related skills, such as naming, repetition and fluency. Propicious to be accompanied by weakness of the right side of the body.
- **Transcortical motor** - Similar to Broca's aphasia, but with a relatively normal ability of repetition.
- **Global** - The most severe of the possible diagnosis, every language related skill is impaired, having to resort to stereotypical expressions or even facial expressions and gestures to communicate.
- **Mixed transcortica** - Less severe type of global aphasia, where repetition is kept, which leads to the presence of echolalia.

It is common for several of these Aphasia subtypes to have complications in common such as paraphasia, which is the interruption of the normal flow of speech with inadequate words, or even physical ones, such as paralysis of certain parts of the body [12].

A deeper medical research into aphasia exists where more intrical partitions are made about the brain areas and their impact on the language skills, such as the subcortical areas. These do not tend to differ enough from the described aphasia subtypes to change the adaption and rehabilitation methods [12].

Given the importance of communication, these conditions also tend to bring other issues with them on the mental health side. Speech impaired people will often describe despair and loneliness for their inability to have meaningful conversations. These feelings reach heights of existential dread, and affect not only themselves but the ones surrounding them [16, 17].

Looking at the variants that were described, we can identify different challenges that need to be tackled in order to have a communication system as helpful as possible: from the possibility of not knowing the meaning of certain words, to the inability of processing and remembering what was said, to the possible aggravation due to sentence complexity, these provide insightful motivation to the development of helping strategies.

2.2 Augmentative and Alternative Communication (AAC)

Augmentative and Alternative Communication (AAC) comprises the methods, strategies and tools that have the intent to help individuals with impairment in their communication

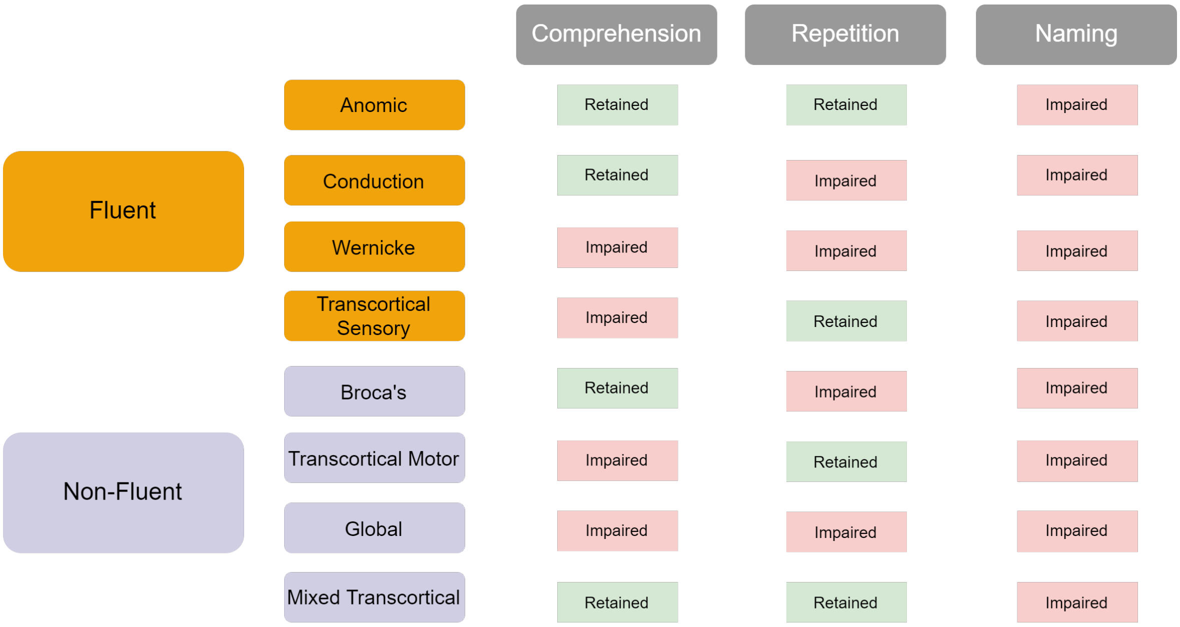

Figure 2.1: Aphasia types and how it affects a person regarding comprehension, repetition, and naming.

---

abilities. They are used in speech therapy as temporary communication assistance but can also be used as a full-fledged communication methods such as text-to-speak.

These strategies can be distinguished through their tech levels, which can be no-tech, low-tech and high-tech [9]. Sign language, facial expressions and other body gestures are part of the no-tech category. Low-tech implies some type of basic tool like books and boards where words and images can be found to help communicate. Electronic devices are the common form of high-tech examples, such as computers with specific software, where the continuous technology development has expanded the options that are available.

In relation to the sensing modalities of AAC, we can also distinguish them through the following methods [9]:

- **Imaging** - The tracking of the eye movement through the use of cameras describes the usual use of imaging methods. Eyegaze is a well known eye-driven device, being fully controlled by tracking the user's eye.

- **Mechanical and Electromechanical** - These methods require an input from the user either direct or indirect. Keyboards are an example of direct selection of the input, while scanning is part of the indirect methods [9].

- **Touch-Activated** - With the use of actions such as tapping and swiping, touchscreens are bearers of several AAC applications. This technology is already part of the common household, given its established development.

- **Breath-Activated** - With the aid of sensors, the user breathing can be a method of input. The speed, amplitude and phase of breathing can have the attribution of a meaning, where a system as morse code could translate it into a message [9].

- **Brain-Computer Interface** - Brain signals are being researched as a way to control devices, through different both intrusive and non-intrusive methods such as electroencephalograms and near-infrared spectroscopy [9].

2.2.1 Assistive Communication Solutions for Aphasia

When looking for assistive technology for people with communication difficulties, we can find solutions that resulted from the research in other projects. Early work includes PhotoTalk, a mobile device application that allows PWAs to capture and manage photos for supporting face-to-face communication [18]. Another proposed solution is TalkAbout, a context-aware, adaptive AAC system for PWAs, which is based on a tablet with touchscreen, a grid of word and phrases associated to pictures, and speech synthesis [19]. It additionally takes the context into account, by recognizing the current users through the tablet's cameras and obtaining the user's location.

More recently, other solutions have been explored, including CommBo, a web-based, speech-generating picture communication board [20], which offers customization and suggestions based on frequency and time of use and machine learning. Another proposed AAC tool is ECO (Easy Communication Application), a mobile application that facilitates communication for users with complex communication needs, by providing a way to create messages, which can be grouped into categories, using pictures and associated text, with the possibility of also using audio and video [21].

Three AAC mobile applications were proposed by the same authors, focusing on the specific scenario of ordering meals in restaurants [22]. For each application, a different strategy for supporting communication was adopted (e.g., automatic captioning of photos, image to text conversion, information retrieval based on the user's location). Regarding interaction, one of the applications provides multimodal input (text and voice), while another provides multimodal output (speech and images).

Regarding solutions for communication support based on gesture input, a Personal Gesture Communication Assistant was proposed [23]. This solution recognizes gestures using a camera and machine learning. The gestures and corresponding meanings are defined by the user, with each gesture being associated to a different word.

To the best of our knowledge, no solution has been found that relies on gestures to enable remote two-way communication between a PWA and another person, in the context of the in-bed scenario, besides some exploratory work by our group, where a first proposal of a communication support system was described [1, 24] and gesture recognition based on either wearables [2, 1] or a radar [3, 10] was explored.

2.3 Gesture Recognition

Hand gestures are used in a wide range of applications, providing means to interact with a device from a distance [25]. There are several approaches to the data acquisition of gestures for gesture recognition through sensors such as computer vision-based [26, 27], wearables [28], and radars [29].

Computer vision-based methods use cameras that apply different possible techniques to identify gestures. Isolating the relevant data from the non-relevant can be done through segmentation techniques, such as skin motion, shape, color and models of hands [26]. Frame-to-frame analysis of consecutive hand segmentation allows for the tracking of the hand. Models and classifiers can then be used to provide an informed guess of the gesture being executed. Using cameras for hand recognition is not an expensive option due to the banality of cameras in the present society, but it can be computationally heavy. Recognition of gestures in static images have achieved high reliability [30], whereas dynamic gestures provide a greater challenge and is still under constant research and scrutiny towards achieving the best accuracy values. Illumination and non-relevant background motion and objects are some of the complications recognition methods have to manage [26], hence the environment being an important variable. The idea of people being recorded also arouses privacy concerns.

Methods using wearables make use of physical principles to obtain information about the hand, using sensors such as accelerometers and gyroscopes to provide the data [28]. This data can be processed in the wearable itself, or in a processing unit. Wearables provide continuous oversight without being affected by the environment [29] and do not evoke relevant privacy concerns. The downside is the need to wear the device continuously, and remember to charge it regularly. Some of the work done in this area [31, 32, 33, 34] has showcased high gesture recognition accuracy results in user-dependent cases, whereas user-independent results have a higher range variability, with some works even deeming it unfeasible with their model.

Radars (short for Radio Detection And Ranging) such as Frequency-Modulated Continuous Wave (FMCW) based ones, are also widely used for hand gesture recognition [29]. As the name suggests, these makes use of a frequency modulated continuous wave that allows an estimation of the angle, velocity, and range of a target [29, 35]. Simply put, a radio signal is sent, scatters of the objects it finds, and receives some of the signal back, which is then processed [36]. A clear line of sight with the target is recommended, but similarly to wearables, privacy concerns are minimized since the values stem from physical principles. It does not suffer from illumination problems, but background motion can be an obstacle [29]. An interesting study references cases of Fusion of sensors to increase accuracy [37]. In this study, pressure sensors and radar were integrated together and increased accuracy results by up to 23.5%.

Having mentioned some data acquisition methods, the data has to be processed and an estimation of a gesture emerges as the expected output. Figure 2.2 shows a generic pipeline representing the process that starts with the data sent by the sensor and allows for recognition of the gesture. This generic representation is based on the pipelines used in the works about gesture recognition [28, 29, 26, 30, 1, 10] that were part of the research. Nonetheless, some sensors may require more work than others when pre-processing the data received.

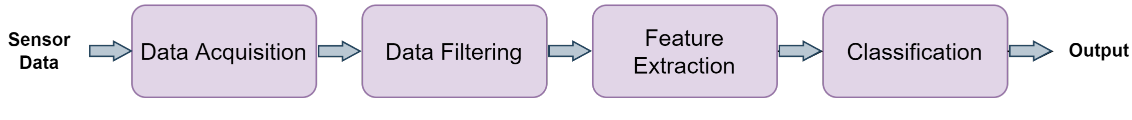

Figure 2.2: Generic pipeline of gesture recognition.

The sensor sends the data meant to reach the processing unit, and the first section of the pipeline guarantees that a connection is maintained with the sensor, where the data is extracted and parsed. Once the data is parsed, it can be pre-processed and filtered in order to extract the relevant features. With the aim of providing a gesture as output based on the relevant data, a model for gesture classification is commonly used. This model is usually trained using a machine learning algorithm together with a relevant data set.

In the following subsection, we will discuss the relevant options towards gesture recognition taking into account the goals of the project.

2.4 AAC and Gesture Recognition regarding APH-ALARM

As mentioned in the section 1.1, the APH-ALARM project aims to provide solutions that contribute to an increased sense of safety and independence of people that have disabilities that stem from stroke such as aphasia. Taking this into account, one must consider a solution that tackles obstacles such as communication difficulties and lack of mobility.

Listed previously in section 2.2.1, there are several solutions that aim to provide assistance for people with communication difficulties, mostly with the goal to aid communication between two people in the same space, but none with the intent to enable remote two-way communication between a PWA and another person, in the context of the in-bed scenario.

With this scope in mind, we have specific criteria that will guide the discussion on what is the best approach to aid communication taking into account our focus of bedroom scenario, a constant availability of the system, and minimizing privacy concerns and intrusiveness problems. We will consider intrusive as something that might disrupt the flow of one's action (e.g. electroencephalogram). An aspect also important to consider when aiding communication is the functionality and intuitive usage of the system.

When discussing sensing modalities of AAC, we can state that Breath-Activated methods are usually slow and limited, and Brain-Computer Interfaces are mostly not suitable for every day usage [9]. Both can also be considered too intrusive in a person daily life and have training requirements [9]. Mechanical and touch activated devices have the disadvantage of having to be carried around, and not being completely suitable for someone who has fine motor skill problems.

Imaging methods can be seen as one of the most suitable for gesture recognition at a distance, but can be considered by the users as too intrusive regarding privacy due to the need of constant recording, and they can also be affected by lighting. They also tend to come with high costs and are typically dependent on calibration [9]. A possible alternative would be thermal cameras, with research supporting its use [38, 39, 40].

With these observations in mind, wearables, radars and thermal cameras seem to stand out regarding our objectives. This analysis had already been performed in the scope of the APH-ALARM project, and two options deemed most adequate for the target user and scenario have already been explored (wearable [1] and radar [3]). Given that the use of gestures to aid communication is suitable for this work, seems sensible to understand and analyse these solutions.

2.4.1 Wearable

Watches have been part of the human life for decades, and smartwatches are already well known in the modern society, being used daily by many people. Since people are used to it, it minimizes the idea of a smartwatch being intrusive, and adding an extra functionality to it would not change that.

Using a smartwatch as a wearable, there is no dependency on a device that demands touch-based interaction, it requires minimal movement to use and their use is feasible while in bed. It does need to be worn and recharged, and it requires at least one side of the body to be functional. In counterpart, they raise less privacy concerns compared to imaging devices and they also provide data steadily. With this in mind, we take a look at a prototype that was developed by Afonso [1, 41] where gesture recognition was shown possible through use of a wearable.

An Oppo smartwatch was used, which has the Android Wear operative system, and the gyroscope, accelerometer and magnetometer are the sensors that provide the data necessary to identify dynamic arm gestures. The data is transferred through Bluetooth and processed in a Raspberry Pi 4.

The system has a simple architecture, where the data goes directly from the smartwatch to the processing unit, goes through a classification pipeline with the use of a MQTT broker, and the decision reaches both an android application and a webservice through the use of endpoints. The caregiver can make use of the android app to confirm the notification received or ask a question, and the webservice can make use of the only available output to the user, which is speakers.

Regarding the gestures and recordings, 5 gestures were recorded. A total of 10 participants performed each gesture continuously 10 times, for 5 seconds each time. At a sampling rate of 50 Hz, 2500 values would be acquired per 10 recordings of a gesture, per subject.

A model for gesture recognition was evaluated based on the dataset resulting from feature extraction over the acquired sensor data. For feature extraction, different types of sliding windows were considered: 1 or 2 seconds, and overlap of 0%, 50% or 96%. From an initial set of 84 extracted features, a subset was automatically selected, resulting in a dataset with 20 features. Several machine learning algorithms were explored: support vector machines (SVM), decision tree (DT), random forest (RF), and Gaussian Naïve Bayes (GNB).

For the trained gesture recognition model, the user dependent solution produced an overall accuracy close to 99%. The more practical user independent solution reached an overall accuracy and F1-Score over 92%, which is expected to be lower.

2.4.2 Radar

Looking at the work done with a radar by Santana [10, 42], this method minimizes privacy concerns since the detection method is through radio waves. It does demand to be constantly powered and it is recommended to have an open line of sight. The devices can be used without any constraints related to the time of the day, allowing for constant availability.

In this work, the radar has only been used for offline data recording with the aim of model evaluation. A total of 5 gestures were recorded 10 times by 4 subjects, which originated an original dataset of 200 images. Offline data augmentation was used to increase the size and variability of the dataset. It should also be mentioned that there were two types of solutions tested, a subject-dependent, and a subject-independent one.

In terms of feature extraction, we can mention that the data was transformed into images, and applied for recognition through use of the transfer learning method, in which the most successful pre-trained model used was MobileNetV2, when compared to the other two models used: NASNetMobile and DenseNet121. In terms of results when using the pre-trained model mentioned, the best mean accuracy result for the subject-dependent and independent cases was of over 95%, and 50% to 55%, respectively.

The solution that would ideally be used is the subject-independent one, which requires some improvements. To do this we can increase the size of the dataset, which was also considerably small. It would be beneficial to have both a bigger number of people provide data through recordings, and more data per person to increase the variability of the data.

An alternative is to use the subject-dependent solution, which involves acquiring data and training a model for each new subject that would use the system, which would provide higher accuracy values as seen above.

For integration with the communication support system, the existing solution needs to be adapted for the scenario of real time recording, which implies additional hardware and changes to the capturing software.

2.4.3 Wearable and Radar Overview

Analysing the values presented in both works, the subject-independent values are the most relevant and we can see they are the highest with the smartwatch, with an overall accuracy and F1-Score over 92%.

On an overview of both works, the radar stayed within the scope of researching about the feasibility of gesture recognition, while the wearable project went beyond that scope and implemented means of communication between the gesture input modality, which uses a gesture recognition model, and an Android smartphone application, with a pipeline of information flowing between them. Nonetheless, the communication is limited and the gestures explored were not thoroughly researched towards its appropriateness for a person with aphasia.

These observations will support the decisions made concerning the gesture input modality towards the development of this work in order to achieve the best results within the available time. Furthermore, when considering the complex context of this work and the obstacles presented to the target user, a User-Centered Design approach is deemed the most suitable towards the development of the system, and is described in the following section.

2.5 User-Centered Design

User-Centered Design (UCD) is the name given to a product development methodology in which the user is involved in the process since the very beginning, in order to improve the understanding of the requirements and assuring his needs are met. It is important to add that when the user is not immediately accessible or it is more challenging for him to provide feedback, this can be done by proxy through a carer or an expert of the subject.

This process can make use of the iterative framework, following the Agile philosophy, meaning it will have a life-cycle of continuous evaluation and re-definition of requirements.

As shown in Figure 2.3, we will take into account 4 phases, which are:

- **Problem Definition** - This is the beginning phase, in which you visualize the scope of the problem by empathising with the user and understanding their needs and goals. The creation of personas and scenarios help assure the extraction of the requirements, in order to start planning the product.

- **Requirements** - Once the scope is defined, the planning of the product begins. From the initial brainstorms, to top-level architectures, to diagrams of specific interactions, the ideas and preparations gain shape in this section, and the requirements of the product are defined.

- **Prototyping** - The ideas that came together are now put in practise, and prototypes are made through different techniques to represent and become real components of the

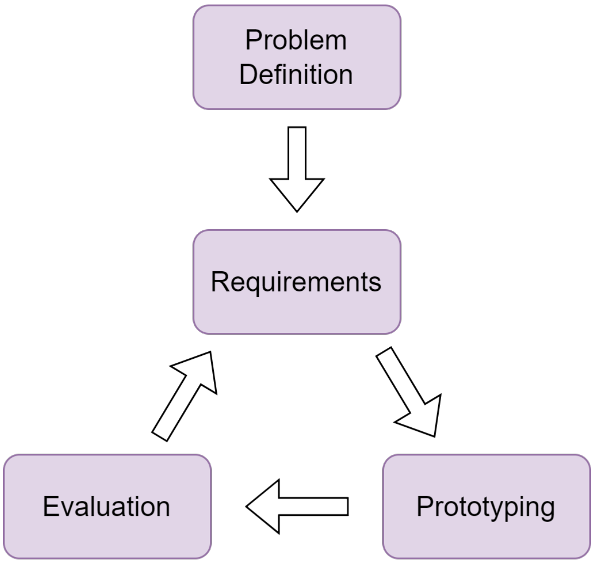

Figure 2.3: Generic process of UCD iterations

product. In the Internet of Things scenario, it can go from assembling the hardware, to developing the software itself.

- **Evaluation** - Once a prototype is ready, it should be tested and evaluated. The user should be given access to the product, and test himself, giving regular feedback that is considered ideal, since it comes from the user itself.

2.5.1 Problem Definition and Requirements

Taking the fact that understanding the user and their needs is crucial, an important part of the UCD is the creation of personas and scenarios. This is part of the problem definition process, and will help identify the target user and the situations in which the product is used. From these, we can retrieve the functional requirements of the product. The following are some guidelines used:

- **Personas** - Who are we making the product for, what is their motivation to use it, what limitations do they have, and how does it influence the final design? These are all questions looking to be answered with the development of personas. They represent the user, following the idea of having a user centered design, and benefit from having the designer empathy in the search of the problems. The motivation is an important asset which the scenarios will look to fulfill.

- **Scenarios** - The description of scenarios shall give us an idea of when and how the product is used. There should be detailed usage of the product, with specific interactions being showcased, proposing solutions that satisfies the persona's motivations.

- **Requirements** - The specifics of the product use are extracted from the scenarios, which help with the design and logic behind the interactions. What needs to be done, how it needs to be done, when it needs to be done, are all addressed and planned accordingly.

The personas and scenarios are developed in the problem definition phase, and the requirements are constantly under scrutiny and reconstruction on the requirements phase.

2.5.2 Prototyping

In the prototyping phase, prototypes of the product are made, which are a representation of the product, either as a whole, or as a part [43]. These allow for an early interaction and visualization of the product, while giving the developers insight as to the integration of the different components of the product.

The different types of prototype and their fidelity are suitable for different stages of the design and development and should be selected accordingly. The following are some techniques used to create prototypes, listed from smaller to higher fidelity, which is the resemblance with the finished product [44]:

- **Paper Sketch** - As the name indicates, this technique makes use of pen and paper to create an initial visualization of the UI and its navigation. Simple windows and layouts are part of this first approach.
- **Wireframe** - Considered a medium fidelity representation, tools such as powerpoint [45] can be used to create a schematic of the overall design. It should describe some functionality and represent the structure of the navigation [46].
- **Functional** - These representations already have a close resemblance of the final product, having real functionality and flow of navigation. It can also be described as an incomplete form of the desired product.

The lower fidelity prototypes, like a paper sketch, have the advantages of having low development costs and flexibility in terms of what can be tested. The counterpart is that this flexibility comes with limits in terms of the usefulness of what's being displayed and its navigation [44].

The higher fidelity prototypes come with the benefits of having high functionality and testability, which increases the value of the feedback given by the testing user. The disadvantages are the high time consumption in order to achieve this state and the higher costs.

Evaluation

On the next phase, the prototypes are tested and evaluated. These are either expert-based or user-based. The phase of the prototype also leads to having more and less suitable evaluation methods [47]. The following, are some of the most used evaluation methodologies that are relevant to the user centered design approach:

- **Focus Groups** - A group of participants is lead by a knowledgeable moderator who can have tasks or questions prepared to the participants. These shall initiate discussions with the intent of giving feedback and gathering thoughts about the prototype. It is one of the earliest feasible feedback gathering methods.

- **Heuristic Evaluation** - There are several lists of usability principles that can be used to inspect the interface and find the usability problems of the product. Nielsen's heuristics are probably the most known and used since they were refined in 1994 [48]. Ideally, a set of 3-4 experts examine the interface in light of the chosen heuristics, identifying flaws and inconsistencies that will help shape the interface into a better one. Some other heuristics that should be mentioned are Shneiderman's Eight Golden Rules of Interface Design [49], and Tognazzini's Principles of Interaction Design [50]. Can be done when the prototype is still at a low fidelity phase.

- **Think-aloud User Study** - Done in the functional phase by users, a participant narrates the actions done towards the completion of a task given to him. This can provide information that may not be remembered later, and give specific information about some of the steps [45]. Higher fidelity is required in order for the user to simulate the tasks.

- **Standardized Usability Questionnaires** - Being one of the least expensive methods, a standardized usability questionnaire provides data such as the perceived usability of the evaluated interface, and the user satisfaction towards it [51]. The System Usability Scale (SUS) and Post-Study System Usability Questionnaire (PSSUQ) are some of the most used, both considered universal. The mentioned systems use a Likert-type scale, which looks to know the agreement level of the user to a statement made about the interface [52]. Most suitable for a higher fidelity prototype.

During its several iterations, the UCD approach deems important the collection of user feedback and testing of different parts throughout it's life cycle. With every iteration, the goal is to have the end-user become increasingly satisfied, turning UCD into a highly chosen methodology to develop products that are targeted at a specific audience. This fits the context of this project hence it being the adopted methodology.

2.6 Summary

With all the knowledge gathered in this section, we now have a basis for the development of strategies and conceptualization of a system towards supporting people with aphasia communicating with their caregiver, relatives or friends, when they are not physically present.

The possible difficulties of a person with aphasia, such as difficulties in conveying a message, lack of fine motor skills, impaired comprehension of the language or even forgetting what was recently said, will guide the decisions made towards tackling the challenges they present to the PwA when wishing to communicate. There are also several studies reinforcing the idea that people with aphasia can use and benefit from high-tech aid [16, 53, 54].

The research about AAC methods and tools allowed us to understand what is the most appropriate technology to implement in our scenario. The options mentioned as part of the research towards assistive communication solutions for aphasia partially tackle some of the defined criteria and challenges, but to the best of our knowledge there is no system that fulfills all the criteria defined for the scenario.

Gesture recognition through non-intrusive methods at a distance is a desirable path for user input, one that has already been explored in the context of the APH-ALARM project, introducing the possibility of a quick-start in its implementation due to work carried out regarding wearables and radars. Nonetheless, a new dataset might have to be recorded according to the research on the appropriate gestures for a PwA, and both the wearable and radar solutions would need improvements to make it feasible for day-to-day use.

User Centered Design is an appropriate methodology to follow in the development of this project, given its focus on satisfying a specific audience. Taking into account the challenging aspect of receiving feedback from a PwA, asking experts on the subject to give insight and feedback on the system is an important course of action to consider.

As described in the UCD methodology, the next steps are to gain more understanding of the user and the scenarios of usage of the system, together with the extraction of the requirements, all of this supported by an evaluation of the concept done by discussing it with Speech and Language Therapists (SLTs), being this approached in the next section.

Chapter 3

Concept Definition and Refinement

In this chapter, we will support our human-centered development by describing in a more detailed way the people and some of the scenarios this project aims to help. The conceptualization of the system also benefited from a focus group done with Speech and Language Therapists (SLTs) that evaluated an early prototype of the system concept, and is already reflected on the personas and scenarios mentioned ahead. All of this helped in the extraction of the requirements, and informed a refined proposal of the system that aims at helping PWA.

3.1 Personas

Personas are descriptions of the target users. They help us identify what the user wants and needs from the system. All of the personas have some type of aphasia, which implies some level of communication impairment, or have a direct relationship with someone who has aphasia (e.g. caregiver).

In the context of the APH-ALARM project, personas have been researched, created and evaluated by professionals that work directly with people with aphasia. These reflect the characteristics of the target user, and were defined in the context of the work carried out by Cátia [55]. Therefore, the personas were reevaluated through an iterative process towards better representing the scenario considered in this dissertation, and the SLTs were part of this refinement process. This core role of expertise resulted in adjusted motivations to our particular scenario.

3.1.1 Persona 1: Judite

Name: Judite Rodrigues
Age: 38
Profession: Primary School Teacher
Aphasia type: Broca's

Judite Rodrigues is 38 years old, having been born in Lisbon, on January 17, 1984. She is single, has a daughter, Inês, and lives with Inês and her parents, both of them retired, in Lisbon.

As Judite never had siblings and always wanted to, being a mother of Inês and working with children was a dream come true. Judite worked as a primary school teacher and loved what she did. She is a very caring person, and in addition to the dedication to Inês and her students, she also took great care of her parents. She worked to provide them the best life possible, as they once did towards her.

Since Judite was 11 years old, she has been living with type 1 Diabetes, for which there is no cure. Regarding her condition, it is very important to keep the diabetes stable and controlled, and to achieve that, Judite needs daily insulin injections in addition to other preventive measures, including the use of medication, healthy eating and physical exercise.

In addition of taking advantage of mobile applications to help her organize her schedule and Inês', Judite also made use of her computer to prepare lessons, as she is very into finding dynamic and thematic approaches of teaching, for which she is well known in the teaching community, as she used to share some tips on social networks with great feedback.

Since Inês started attending primary school last year, Judite had been under more pressure as her work load and responsibilities increased a lot. In addition of having work related to school to do (correct tests and assignments, prepare lessons, etc) and also helping her mother (with everything related to the maintenance of the house, grocery shopping and meal preparation so as not to overwhelm her), she also helped Inês with her homework and school tasks, and since then, from the moment she got home, she never stopped.

Because of all that, Judite stopped exercising and taking care of her healthy eating habits as she used to. Judite had a stroke two months ago, which led her to be diagnosed with Broca's Aphasia. She now struggles to speak and in addition to that, suffers from paralysis on the right side of the body.

Since the brain injury episode, Judite struggles with helping at home like she used to. Seeing her parents taking care not only of Inês, but also of her and the house, brings sadness to Judite as she feels that she is not providing the life she wanted for them. She tries to help Inês with her school related tasks, but not as effectively as before, because she cannot always find the right words and even though she tries to get around it by explaining what she means, it is not the same thing.

Judite is not working, since the speech barrier can strongly compromise her teaching, and even though at home she can always try to explain slowly what she means when the words fail, at school as a teacher it is not feasible. Judite is now attending speech therapy and physiotherapy appointments weekly in order to work towards regaining her full communication and physical capabilities, but the doctors have already told her that it will take some time.

Motivation: Judite's greatest motivation is to make the situation as undemanding as possible and ease the burden she feels she puts on the family. Allowing her parents to do the most in their daily life by having an easy and quick way to contact them at any time of the day would help with her guilt.

3.1.2 Persona 2: Natália

Name: Natália Rodrigues
Age: 68
Profession: Retired
Role: Caregiver (mother)

Natália Rodrigues is 68 years old and was born on April 13, 1954. Natália is from Mozambique and when she was 13 years old she moved to Lisbon, where she lives until the present day with her husband, her granddaughter and her only daughter, Judite, the person diagnosed with Broca's Aphasia.

Before retirement, Natalia was a nanny in a nursery, but she decided to retire when she was 64 years old because, even though she loved her job and kids, it was time to rest and enjoy life. Since then, she was having a great life with her husband, as they were attending outdoor gymnastic classes and would meet with friends in the neighborhood almost every day, until everything changed when, two months ago, her daughter had a stroke.

Natália and her husband stopped attending dance classes and they only get to see their friends when they visit, as Natália almost never leaves the house because she has no time considering that Judite and Inês need her around almost all day.

Having Judite with Broca's Aphasia and suffering from paralysis on the ride side of her body breaks Natália heart, and even though Natália does not feel the naming deficit that Judite faces, as Judite is comfortable at home taking her time explaining what she intends to say, Natália knows that in order to go back to work, Judite needs to overcome that in some way. She also has been aware that Judite is sometimes apprehensive about leaving the house to do things she used to, afraid of being asked something she may struggle with when answering.

Motivation: Natália's greatest motivation is to see her daughter happy again, and in order to that, she needs a way to alleviate the challenges present in taking care of her, so her daughter regains some independence and feels less guilty. With Judite on the right track, Natália may think about going back to dance classes and get to see her friends more often.

3.2 Scenarios

To describe our vision of the solution being implemented, we adopted scenarios. Scenarios and are descriptions of situations the in which the system is used to fulfill their needs and motivations. They help us identify the usability obstacles, and plan to assure its accessibility. All the following scenarios are imagined in a bedroom context, where the person is lying in their bed.

3.2.1 Scenario 1: Judite is feeling dizzy

Judite is resting in bed in the afternoon when she feels dizzy when trying to get up. Her parents went to the supermarket for groceries and her daughter is at school.

Judite asks for help: She activates the communication support system using the corresponding pre-defined gesture. The system asks if she needs immediate assistance, to which Judite answers with the gesture defined for "Yes".

Caregiver receives help request: Given the answer above, the system sends a message to her mother (Natália) saying that Judite needs immediate help. After receiving this message on her smartphone, Natália confirms she has been notified and is on her way.

System feedback and context questions: The system informs Judite that help is on the way and asks her a few more questions to better understand the context, and additionally gives her feedback about the assistance status. Judite is asked if she is in pain, and she uses the gesture associated with the meaning "No" to answer the question. The system also asks if she fell, and Judite uses the same gesture to reply again "No". Another question is if she is feeling unwell, to which she answers "Yes" using the corresponding gesture. Meanwhile, all these answers are being given as context to Natália through her smartphone application, which helps her to better understand the situation.

3.2.2 Scenario 2: Judite is stuck in bed

Needing help getting up: Judite was lying in her bed, when she decided to change her posture. While trying to rotate her body to the side, the bed covers got stuck and Judite noticed she could no longer move her body properly. Having her right arm paralyzed, she decides to ask her mother, Natália, to help her out without waking up everyone in the house, by using the system.

System usage: Judite activates the system with the corresponding gesture, and is asked if she needs immediate help. Not considering it a situation worthy of panic, she answers no, but the next question asks if the caregiver should be asked to come, and she answers yes. More yes/no questions looking for details are asked, such as "Do you want water?", and then the message is sent.

Caregiver helps: A message is sent to her caregiver, and feedback from the system is given that it has been received, after the message is acknowledged. Soon after, she hears her mother climbing the stairs.

3.2.3 Scenario 3: Natália goes out and checks up on Judite through the smartphone

Natália sends message: Natália accepted her friends invitation to go out after dinner and is on the park with her friends. They have been talking outside for hours, and she knows her daughter should be already in bed. In order to decide if she extends her stay or goes home, Natália opens the app, and sends a message to Judite asking if everything is ok by typing that question.

Judite answers back: Judite was lying in bed and she notices the system activates and the question her mother sent is written on the display, and being said out loud through the speakers. She does the gesture that represents the yes answer, which is sent back to her mother.

Natália receives answer: Soon after her question is sent, Natália receives a notification, and is able to see that Judite answered her question affirmatively.

3.2.4 Scenario 4: Judite activates the system by mistake

System activation by mistake: Judite is sleeping during the night and is suddenly woken up by the sound of the speakers asking if she needs immediate help. She realized she must have moved while sleeping, since she is now lying with her stomach down and knows she is someone that moves a lot while sleeping.

System interaction: She does the gesture corresponding to no, and the system asks different questions such as "Want to inform the caregiver of something?". After all of them received a negative answer, the system proceeds to return to standby mode.

3.3 Concept Refinement and Evaluation

As part of the refinement of our initial concept, input from the user was a possibility considered ideal towards the development of the system. Unfortunately, the option of obtaining feedback directly from PWAs was dismissed since the complexity of the situation itself may render the explanation of the proposed approach and obtaining feedback very difficult, particularly without a tangible example that can be operated.

Therefore, we decided to have a discussion with Speech and Language Therapists (SLTs), which provided feedback by proxy. Furthermore, the SLTs can be considered an ideal source of feedback since their focus is also on communication and can offer a greater insight on the best approach and most appropriate strategies for supporting communication in their daily life, bringing a lot of value to this work which main focus is assisting communication.

While there has been some involvement from the start by the SLTs, the most important discussion was the one where we presented low fidelity sketches with some key concepts and questions crucial towards solidifying the core ideas and design of the system.

The adopted method to guide the discussion is described below, followed by the results obtained in the discussed concepts.

3.3.1 Method

Two focus group sessions were carried out with the participation of three SLTs in total with experience with PWAs and no prior knowledge about our proposal, where the first session involved two SLTs, and the second session the third SLT. In both of these the applied method was the same.

The focus groups were moderated by a human-computer interaction (HCI) researcher with the participation of two other Computer Scientists involved in the design and development.

After a general introduction to the proposal, the following four main topics were discussed: (a) the adequateness of the gestures for PWAs; (b) flow of communication supported by the assistant; (c) presentation of questions and feedback to PWAs; and (d) most notable motives and situations distressing PWAs in the bed scenario.

When discussing the gestures, three were presented to the SLTs as an initial set of gestures considered simple and non-demanding in terms of physical effort. These gestures are: knock, as in knocking on a table; clean, which is sliding the hand horizontally; and twist, a rotation of the wrist as if opening a door with a key. The knock and clean gestures can even be associated to the vertical and horizontal head movements that mean yes and no. Towards knowing the difficulty of the gesture being understood, remembered, executed and explained, we asked the SLTs to rate them following a 5-level Likert-like scale, which goes from 1 to 5 (very easy to very difficult). Suggestions were also encouraged towards different gestures that would fit the criteria.

Following the gestures, the concept of the assistant was described as an approach reliant on simple and hierarchically structured questions, in order to establish the intent of the request and the message to be sent. To aid this explanation, an initial diagram was showed, as seen in Figure 3.1, which was designed as first approach to what the interaction flow could be, in order to provoke discussion and a concept evaluation. The diagram was using questions already thought out as possibly helpful in the day-to-day life of a PWA for the same purpose.

The overall appropriateness of the approach was discussed and if the proposed sequence of questions was deemed feasible for a PWA to express a need. The output towards the user was discussed taking into consideration three different possibilities were available: text, graphics and speech. The needs, fears and motivations expressed by PWAs for the in-bed scenario was the final topic, leading to understand what questions are appropriate to include and in which order.

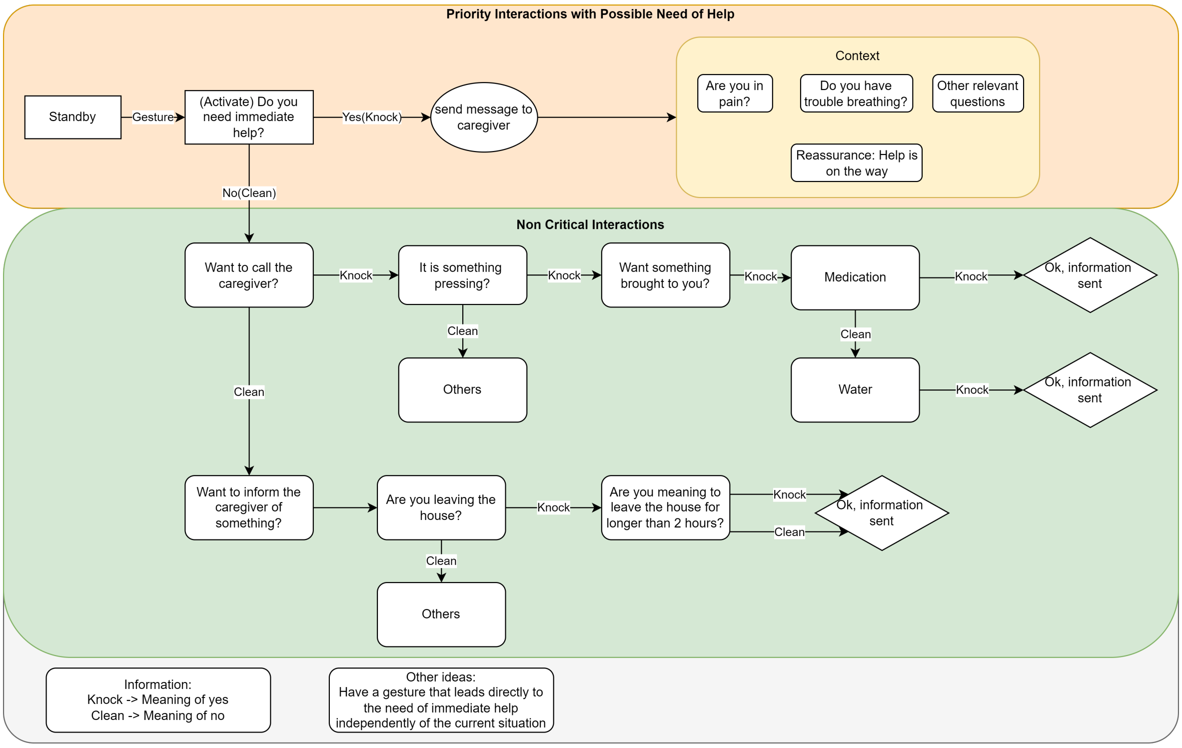

Figure 3.1: Low fidelity sketch of the interaction flow provided by the assistant, designed for evaluation and discussion purposes.

3.3.2 Results

There was a general agreement that remote communication enables a higher level of reassurance for the PWA as well as for their caregiver and family members. Actions such as constantly checking in with the PWA and yelling through the house can be avoided as well. Table 3.1 includes a summary of the key takeaways from the discussions with a brief rationale about them being discussed ahead.

Gestures — Overall, the knock gesture was considered the best gesture, followed by clean, with twist being deemed the most challenging. In Table 3.2, we can see the ratings given in each category, where the first column is a unanimous decision by both SLTs in the first discussion, and the second column is the rating given by a different SLT in the second one. Regarding other possible gestures, there was a suggestion to include the gesture of moving one's hand back and forth while calling someone. Concerning the number of gestures, the SLTs advocated utilizing no more than three gestures, but they felt that two gestures would be the best option for the majority of users.

Communication Mediated by the Assistant — The general approach taken through the assistant in gathering information and prioritizing demands, with the most pressing matters being the simplest to communicat was approved by the SLTs. The concept of knocking repeatedly leading to prompt assistance was also welcomed by the SLTs. Concerning the questions themselves, the use of clear, succint and brief questions was deemed a must towards the best understanding possible of what is being asked. It was also emphasized that the number of questions needed to gather information should be minimized.

Gestures	• *Knock* and *clean* gestures approved • *Twist* gesture might be too complicated • No more than three gestures
Assistant	• Concise and short questions are essential • Organization of questions by priority makes sense
Output	• Multimodality is a must • Complement questions with available answer possibilities • Associate the answers to the gestures
Needs and Motivations	• Providing independence and reassurance would be the main motivation for PWAs and caregivers/relatives • Even a PWA with good mobility can require help at night • Specific needs and their priority were identified

Table 3.1: Summary of notable conclusions from the concept discussion with experts.

Explaining		
Gesture	**First Session**	**Second Session**
Gesture 1 - Knock	1	1
Gesture 2 - Clean	1	1
Gesture 3 - Twist	2	1

Understanding		
Gesture	**First Session**	**Second Session**
Gesture 1 - Knock	2	3
Gesture 2 - Clean	2	3
Gesture 3 - Twist	2	3

Remembering		
Gesture	**First Session**	**Second Session**
Gesture 1 - Knock	1	3
Gesture 2 - Clean	2	3
Gesture 3 - Twist	3	3

Executing		
Gesture	**First Session**	**Second Session**
Gesture 1 - Knock	1	1
Gesture 2 - Clean	1	1
Gesture 3 - Twist	2	1

Table 3.2: Table with the rankings of 1 (very easy) to 5 (very hard) related to the ease of explaining, understanding, remembering and executing the gesture. The first session is the unanimous decision of two SLTs, and the second session correspond to another SLT.

Multimodal Output — When talking about the way the information should be displayed to the PWA, different ways of conveying the same message proved to be unanimous as a must have towards overcoming comprehension difficulties. The suggestions given of using text, images and speech were agreed as a good mesh of modalities to be used together when providing feedback to the user. The SLTs also suggested aiding the questions with visual aid by presenting the gestures corresponding to the answers in a way such as video.

Motives and Sources of Distress in Bed — The SLTs deemed that the system had potential in addressing a number of needs for the in-bed scenario, including: (1) getting up from bed and getting dressed; (2) physiological needs (e.g., going to the bathroom); (3) nutrition (i.e., eating); (4) medication; (5) problems concerning the bed (e.g., getting entangled in the bed covers). This order reflects both the commonality and perceived priority. The SLTs highlighted the idea that even a PWA with reasonable mobility often worries about getting out of bed and falling, especially at night when they are more likely to be alone and have less access to support.

These discussions with the SLTs were the last step in the problem definition phase, and from here on, the technical requirements and design of the system felt informed enough to start taking a practical form.

3.4 Requirements

Given the personas and the scenarios, and taking into consideration the results of the discussions with the SLTs, we can now extract the requirements, which will be divided in functional and non-functional requirements.

Non-functional requirements can be defined as the features or qualities that the users expects or needs the system to have [56], while functional requirements specifies what the system demands in order to do what is expected from it [57].

Non-Functional Requirements

- **Appropriate for the bedroom:** The system is expected to be used while the person is laying in bed, at any time of the day.
- **Two-way communication:** Enabling two-way communication between PWA and other people (e.g., caregiver, family member).
- **Multimodal Interaction:** Whenever feedback is provided to the user, it should be done through several methods, such as speech, graphics and text. The input from the user should be done with gestures, and should be allowed to do from a certain distance.
- **System as a Communication Facilitator:** The context of the communication should be relevant and helpful to the PWA needs.
- **Accessible and Simple:** The gestures need to be easy to execute and memorize, and obvious in their meaning. The communication made with the user also needs to be short and concise.

- **Privacy and non-intrusive:** The system aims to solve privacy concerns through alternative use to cameras, and raise no relevant concerns about intrusiveness in one's life.

Functional Requirements

- **Input through Gestures:** The system requires a gesture input modality for the PWA.
- **Gesture Recognition:** Recognizing the different pre-defined gestures, relying on sensors worn by the user and/or placed in the environment, and machine learning.
- **Messages Exchange:** Enabling the generation of simple message to be sent to the other person, aided by a virtual assistant.
- **Multimodal Feedback:** Providing multiple output modalities to the PWA (e.g., speech, text, and graphics), relying on speaker and a display.
- **Caregiver Application:** Providing an application that allows the secondary user (e.g., caregiver) to receive information from the PWA and send back a message.

3.5 System Proposal

Taking into account all the research and information gathered in the previous subsections, a system that aims to support a person with aphasia in the bedroom scenario can now be proposed. The system can be described as a facilitator of communication between the PWA and other people. It is meant to be constantly available, providing assurance any time the PWA is lying in bed. An overview of the proposed system can be seen in Figure 3.2 and an architecture design is proposed as seen in Figure 3.3.

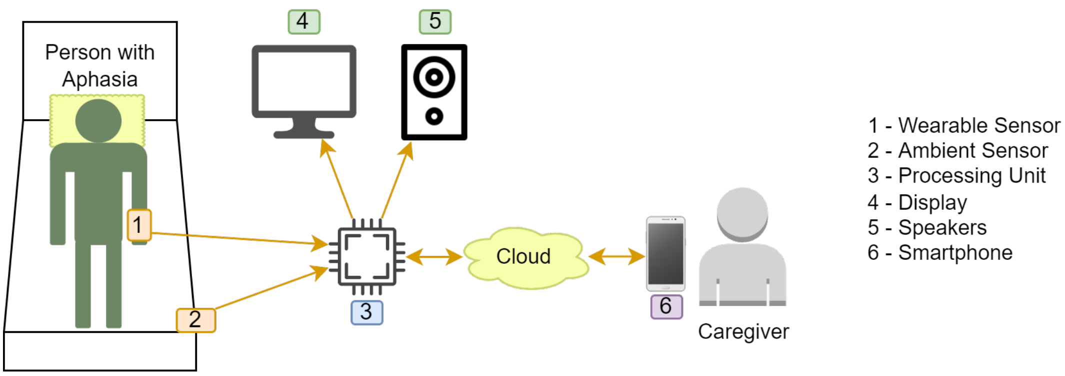

Figure 3.2: Overview of the setup of the proposed system for supporting communication in the in-bed scenario

Input - The former work carried out in the context of the APH-ALARM project [1] was chosen as a basis that can be built upon and progressed towards a system which input is gesture based. Having the PWA use simple arm gestures, which are recognized through the use of a wearable and a trained gesture recognition model, is considered a suitable approach for the input modality of the considered scenario (i.e., PWA lying in bed). There is also the possibility

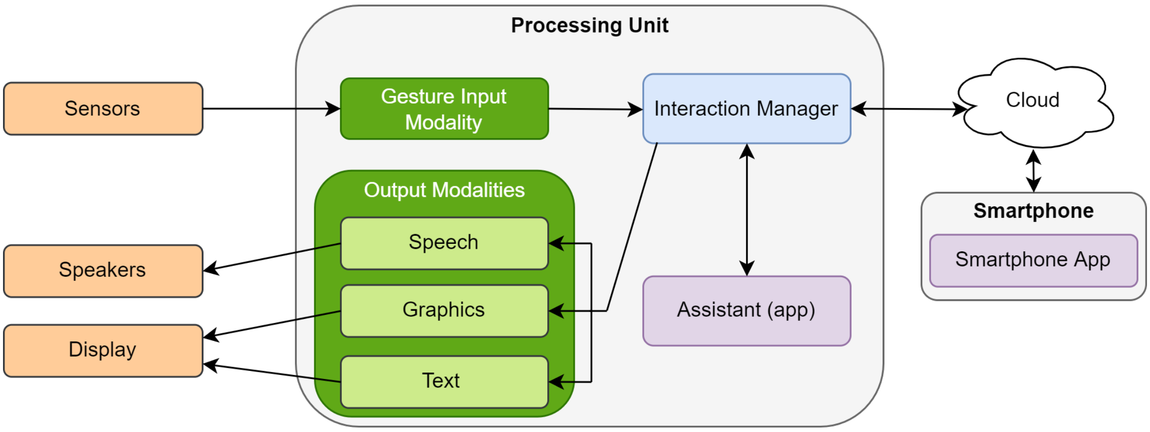

Figure 3.3: Interaction modalities, interaction manager, and applications of the system

of aiding the recognition model with ambient sensors, another concept researched in the APH-ALARM project [10]. Gestures can be performed without having to move around in bed to reach for a device and does not require physical interaction with a device (e.g., smartphone or tablet with touchscreen/buttons). Furthermore, even users that lack fine motor skills can use arm gestures. Taking into account the use of Yes/No questions (mentioned ahead), the gestures "Knock" and "Clean" were seen as the ones with the most potential, and will be the ones used in the system.

Assistant - An assistant would take care of providing local feedback to the PWA and gathering information about the reason for communication before sending any message to the caregiver. The design of the assistant should take into account that: (1) the priorities and motives of a PWA, (2) recommendation of concise questions and (3) yes/no answers are easier to give.

Output - When considering output, multimodality proved to be a key concept. With the former work starting from a point of only speakers being used for feedback, a display that presents information with graphics and text will complement the feedback given to the PWA. The speech output will also be able to ask the question itself out loud.

Interaction Manager - To manage the exchange of messages between the different interaction modalities and applications, an Interaction Manager (IM) should be implemented, guaranteeing the flow of information not only between the modalities on the processing unit, but also to the application on the caregiver smartphone. Proposed for the IM is following the AM4I architecture [58], which is aligned with the W3C multimodal architecture [59].

Caregiver Application - A smartphone application for the caregiver has also been previously developed, but is in need of adaptation to the current scenario. It makes use of confirmations and pre-defined questions to interact with the user, but it should now display the exact need of the PWA through the information gathered by the assistant. The message exchange also requires transformation from the previous use of a local webapp, to the new Interaction Manager method, that would receive the messages through the cloud.

3.6 Summary

In this chapter, we supported our human-centered development by increasing our knowledge and understanding of the target user. We began this approach by describing personas and scenarios that reflect the characteristics of the target user and help us identify the usability obstacles.

Still in the scope of deepening the knowledge about the user, a discussion with SLTs was carried out, where four main topics were discussed: interaction through gestures, flow of communication, means of conveying a message, and needs and motives of the PWA in the bed scenario. A low fidelity sketch of the interaction flow was also discussed.

The feedback gathered from the SLTs provided valuable results and encourages pursuing this approach, providing suggestions on how to further evolve the system, e.g., advancing the output modalities, providing a wide range of information, and refining the questions asked by the assistant. Finally, the requirements of the system were extracted, and a refined proposal was formed.

The system, which can be described as a facilitator of communication between the PWA and other people, is meant to be constantly available, providing assurance any time the PWA is lying in bed. In this proposal, the system includes a gesture input modality that receives data from the sensors. An assistant gathers information about the reason for communication, which is sent to a smartphone application through the cloud, once the purpose has been established. The PWA receives feedback from the output modalities of speech, graphics and text. All modalities and applications communicate among themselves through an interaction manager.

Gestures have been chosen as the preferred input for the system, with the work carried out with a wearable [1] in the scope of the APH-ALARM project used as basis to be built on top of. This requires an in-depth analysis of its functioning, and adaptations to be made accordingly. In order to explore different postures and make use of both arms, a new dataset is deemed important to be acquired, and consequently, a new gesture recognition model to be trained and evaluated. The following chapter describes these steps in detail.

Chapter 4

Gesture Interaction

As introduced in the system proposal (section 3.5), gestures were chosen as the method of input for the PWA in the system, being therefore a very important part of this work. In order to interact with gestures, the system needs to provides a gesture input modality.

In previous work in the context of the APH-ALARM project with a wearable [1, 41], a first version of this modality was implemented involving the following steps: (1) Data Acquisition; (2) Feature Extraction; (3) Gesture Classification; (4) Decision. However, that modality is limited to a single posture and the use of the right arm only.

With the intent of evolving the modality and improving the model used for classification, new acquisitions were carried out with different participants from the previous ones, involving several postures, both arms, and a set of gestures stemming from those used in both the wearable [1] and radar [10] works. These new acquisitions were carried out using the applications described in section 4.1. Using the newly obtained dataset, we carried out a new model evaluation, which is described in section 4.2.

4.1 Supporting Systematic Offline Data Collection

As mentioned before, the work that explored gesture recognition through the use of a wearable [1] brought forth two applications towards acquiring a dataset of gestures: one for the smartwatch and another for the smartphone.

The smartwatch application sends the data from the accelerometer and gyroscope sensors present in the smartwatch to the smartphone application when prompted by it. The smartphone application has control over the recording information, such as its duration and the gestures to be performed.

Given the decision of collecting new data with different postures and gestures, as well as using both arms, it was necessary to analyse both applications and improve accordingly, with the intention of making the data acquisition process simpler and less time consuming.

4.1.1 Recording Application - Wear OS Smartwatch

To send data from the Wear OS Smartwatch sensors to a smartphone, a smartwatch application had already been developed in Java [1]. The data come from the Accelerometer, Gyroscope, and Magnetometer sensors available in the wearable, and are transmitted through Bluetooth. The values are obtained at a rate of 50Hz, and sent every second to the connected device. The recording is triggered by use of a DataEventBuffer sent by the smartphone application, making use of the methods in the Wearable Data Layer API provided by the Google Play services. The triggering event comes with the information of the intended duration of the recording.

No changes were performed in the Smartwatch application, as it served the purpose of being activated on demand, with the data being sent back to the smartphone who is controlling the start and finish of the recording. Nonetheless, an in-depth look at the smartwatch application was needed to assure functionality even after any possible changes done to the smartphone application.

4.1.2 Recording Application - Android Smartphone

Considering that in the present work we wanted to investigate different postures and arms associated with the arm gestures, and explore a set of gestures that may differ from the ones considered in previous work with the smartwatch [1], it was necessary to implement changes through Android Studio to the smartphone application responsible for the recordings.

Definition of Gestures and Postures

The gestures to be used in the system were still under scrutiny at this phase, and a more dynamic app was required to allow for quick changes in the list of gestures and postures available for recording. The decision was to define them using a CSV (Comma-Separated Values) format as a resource file available to the application.

For the experiment described below, a total of 49 entries were defined,corresponding to: 4 postures with 7 gestures, 4 other postures with 3 gestures, and a mesh of 9 “no gestures” (movements and activities different from the defined arm gestures) that can be considered common in the bedroom scenario. The list of postures and gestures for each posture can be seen in Figure 4.1.

The arm gestures seen in this figure stem from the results presented in the section 3.3.2. Knock, Clean, and Twist have already been described in Section 3.3.1 and were considered the ones with the biggest potential. Additional gestures such as Come (moving arm back and forward as in calling someone), Circle (do a circle in the air with your arm), Wave (motion of waving to someone) and Raise Arm (as if wanting to ask a question) were added for exploratory reasons. Some postures have less gestures than others due to physical impossibilities of being performed in that specific posture.

The gestures identified in Figure 4.1 as belonging to the “Other” posture correspond to movements/activities different from the considered arm gestures, which are commonly carried out in the context of the bed. The inclusion of these other movements aimed at helping the system not mistakenly recognize them as one of the pre-defined gestures to be used for communication.

Each gesture was assigned a static recording duration (also in the CSV file), with 10 seconds being defined for the “Raise Arm” and 5 seconds for the remaining arm gestures, during which the gesture should be repeated by the subject wearing the smartwatch. For the “Other” movements/activities, a duration of 60 seconds was attributed. However, the application allows the recording to be stopped at any time, which should happen after the complete movement has been performed once. This guarantees that the whole movement is

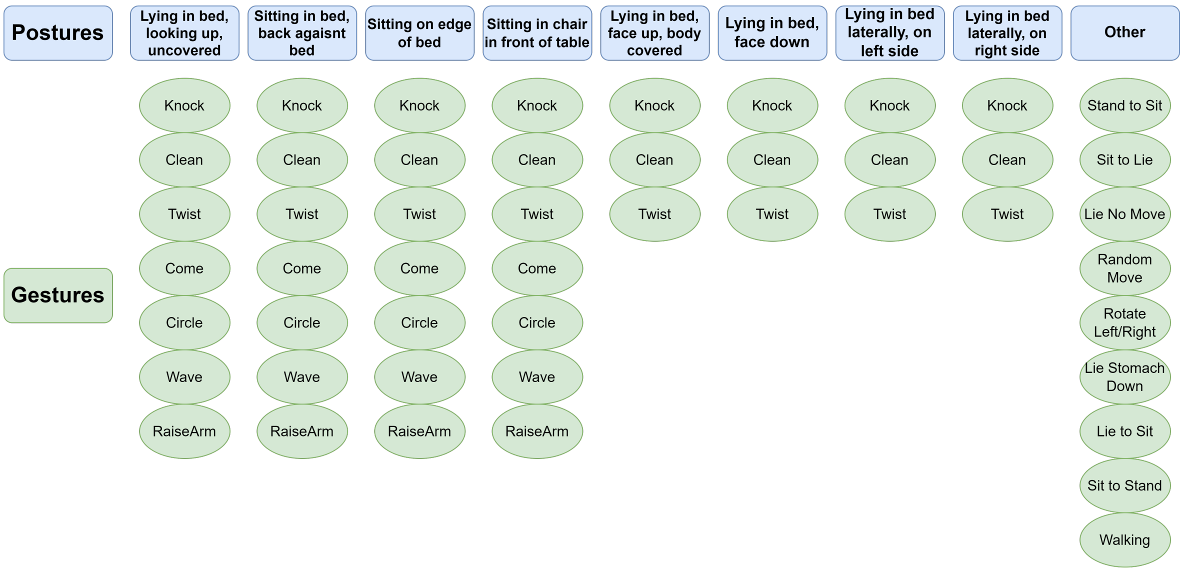

Figure 4.1: List of all the postures and arm gestures, as well as other movements/activities, considered for the acquisitions.

---

recorded, while not lasting more than needed, since the execution duration can vary from movement to movement, and from subject to subject.

Application UI and Recording Automation

One main concern was the time needed for a full recording session with a given subject. A new User Interface (UI), seen in Figure 4.2 next to the old UI, was designed in order to simplify and accelerate the managing side of the recordings. Before starting the acquisitions, the information corresponding to the person ID, age, gender, and dominant arm, needs to be filled. Since we wished to record gesture data for both arms, the UI also has the possibility of assigning which arm is going to be used in the recording.

Then, a randomization of the gestures is required through the use of the "Randomize" button. This random choice allows to avoid bias that can be introduced by having all subjects executing them in the same same order. Only gestures within each posture are shuffled, whereas the posture order remains the same. The current posture and gesture are displayed on top of the screen. This randomization also does not affect the "Other" movements/activities. The recordings can only be initiated after the gestures have been randomized.

When all the information has been filled, the gestures are randomized and the smartwatch application is open, the recording procedure for each posture/gesture or activity is simplified to the steps listed below.

1. Announce to the participant which posture and gesture, or activity, is meant to be carried out next.
2. Ask the participant to execute the gesture repeatedly, or the activity a single time, when the beginning of the recording is announced.
3. Press the "RECORD" button to start the acquisition, while announcing it to the participant.

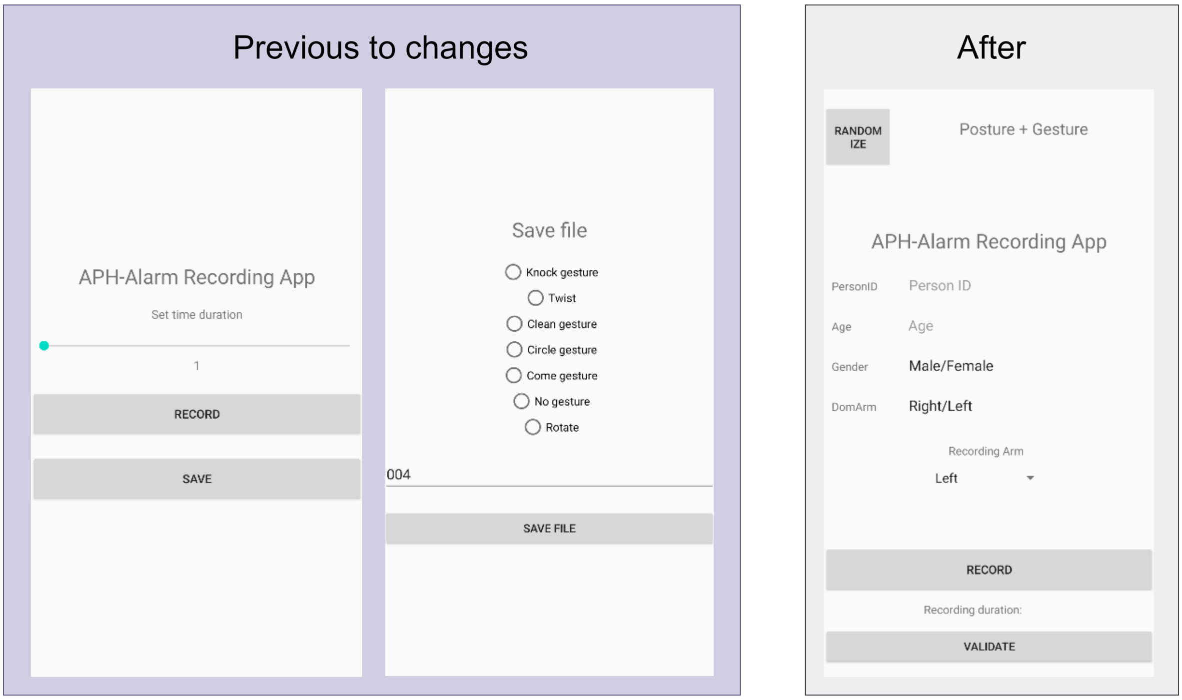

Figure 4.2: The image on the left shows the UI of the previously implemented smartphone application for sensor data recording, while the image on the right presents the UI of the new application. This new UI allows the personal information to be saved, and shows which posture and gesture are meant to be executed in place of the "Posture + Gesture" display after randomizing the gestures.

4. Observe the participant executing the gesture/activity to ensure it is being performed correctly overall, and wait for the data to be received in due time.

5. After the recording has ended, ensure that the duration is according to expected.

6. If the recording is correct, validate the recording by pressing the "VALIDATE" button. Otherwise, start the recording again by choosing the Record button instead.

At any point, if a recording is deemed unacceptable, simply pressing the "Stop Recording" button (the label associated with the Record button changes after clicking it) allows to redo the recording.

When comparing to the previous application, we can now state that once the information is filled and the recording setup is ready, one only needs to press the Record and Validate buttons, while announcing which gesture is meant to be performed. The previous experience of having to assign time individually to each gesture, transition to a different menu, identify the type of gesture to be saved while assigning an ID, and transition again back to the recording menu has now been simplified, ensuring that less time is spent performing the recordings, even with a higher number of variables such as new postures, new gestures, and both arms.

4.1.3 Recording Validation

As mentioned above, the "Validate" button was created to reduce the risk of saving recordings with unexpected results and increase the robustness of the recording process.

The only way for the recording for a given gesture to be considered complete, and the next one be started, is by clicking that button. It is locked by default, and only unlocks once the data has finished being sent by the smartwatch. This allows for the predefined duration to be enforced and minimizes the possibility to make the mistake of proceeding before the recording time is concluded. Above the validation button, the duration of the recording performed is shown to support the duration verification.

Overall, the changes implemented made the recording process of the experiment described in the next section very smooth. Furthermore, they allow the easy removal or addition of some postures and gestures, or even the definition of an entirely new set of postures/gestures, in future recordings by simply changing a text file.

4.2 Gesture Recognition Evaluation

As already explained above, we aimed at improving the classification of gestures performed by the previously implemented gesture modality [1], to allow the system to be used regardless of the current posture of the user and regardless of the arm they choose for wearing the smartwatch. To obtain a new model that enables this, we carried out a new experiment, where sensor data was acquired from several subjects considering different gestures and postures, and other movements/activities, performed with both arms. This data was then used to evaluate the model, for both the subject dependent and independent cases.

In this evaluation, we used a window of 2 seconds with an overlap of 50% for extracting 56 features from the collected data, and the random forest algorithm was used to train the model. These choices were supported on the results of a previous unpublished study carried out by our group, together with some trials done in this work.

Previously, an initial study was performed where a subset of 20 features was automatically selected from 84 features, based on the same dataset used for model evaluation [1, 41]. To avoid possible bias resulting from this approach, a new study was then performed using the same dataset. In that study, instead of automatic feature selection, manual feature selection was opted by considering the features computed over data from each sensor type only, as well as also from each sensor pair. The same window types and algorithms considered in [1, 41] were explored, namely window sizes of 1 s and 2 s, with overlap of 0%, 50%, and 96%, and the following algorithms: support vector machines (SVM), decision tree (DT), random forest (RF), and Gaussian Naïve Bayes (GNB).

The overall obtained results showed that the best performance, considering both the F1 score and false positive rate (FPR) metrics, are achieved when using a 2 s window, all 86 features, and random forest. Concerning the window overlap, although the FPR was statistically significantly better for 0%, it was decided to choose an overlap of 50%, since it allows to have a gesture output every second.

Similarly to the previous studies, the current model evaluation was performed using Python's "scikit-learn" package, with the default values being used for the random forest's hyperparameters.

4.2.1 Experimental Setup and Protocol

We carried out an experiment with the participation of 8 subjects, with an average age of approximately 42 years old, ranging from 17 to 72. Five participants were male and three female, and all of them have the right arm as dominant. All participants read and signed an informed consent. The recordings were performed in different bedrooms, generally the ones belonging to the participant. The same smartwatch, namely an OPPO Watch (41 mm), was used by all participants.

For each participant, all the postures to be adopted and the gestures meant to be performed were explained using the descriptions shown in the Tables 4.1 and Table 4.2, respectively. The activities/movements included in Table 4.3 were also explained. They were further informed that each gesture needed to be performed repeatedly until they were instructed to stop, while the other activities were to be performed only once. For the gestures, varying the speed and range of execution during the recording duration was incentivized.

The smartwatch was then attached to the wrist and the recording application mentioned in section 4.1.1 was opened. As a last step before starting the recordings, an ID was assigned to the participant and their personal information (age, gender, and dominant arm) was collected. After randomizing the gestures for each posture, the procedure for each gesture was the following (for both arms):

- The posture and the gesture meant to be performed is announced to the participant and the recording is started;
- Wait for the recording to be completed in the case of gestures, or manually end the recording in case of the other movements/activities;
- Ensure that the duration is the expected one and validate the recording if it is (otherwise, restart the recording).

When the recordings were finished for one arm, the wearable and the option in the application were changed for the other arm and the process was repeated.

4.2.2 Dataset Characterization

A total of 784 recordings resulted from the acquisitions sessions (98 recordings per subject). The data was acquired with a frame rate of 50 Hz. The 56 time-domain features described in Table 4.4 were then computed over that data using a sliding window of 2 seconds with an overlap of 50%. We used 56 features instead of the mentioned 84 due to trials we carried out with the different sensors. The magnetometer provided worse results, and we decided to use only the accelerometer and gyroscope.

This led to a dataset with the mentioned features, and, with the class corresponding to the different types of gestures (the other movements/activities were named as “Other”). That dataset was not balanced, since one of the gestures (Raise Arm) has twice as many examples of the other arm gestures, and the “Other” movements have a variable number of examples. Therefore, we balanced the dataset by randomly choosing N examples per class and subject, where N is the minimum number of examples when considering all class and subject combinations.

The resulting balanced dataset has a total of 2,048 examples, which corresponds to 256 examples per subject and also per gesture (including “Other”).

Posture	Description
Lying in bed, looking up, uncovered	Lying in bed on your back, facing up, without being covered by anything.
Sitting in bed, back against bed	Sitting in the bed without your back against the bed's headboard or the wall.
Sitting on edge of bed	Sitting on one of the edges of the bed/mattress, with the feet on the ground.
Sitting in chair in front of table	Sitting in a chair in front of a table, with the arms over the table.
Lying in bed, face up, body covered	Lying in bed on your back, facing up, with the body and arms covered with a bed sheet, blanket or similar.
Lying in bed, face down	Lying in bed facing downward, with the arms next to the face.
Lying in bed laterally, on left side	Lying in bed on your left side, with the arms next to the face.
Lying in bed laterally, on right side	Lying in bed on your right side, with the arms next to the face.

Table 4.1: Postures considered for the data acquisitions.

Gesture	Description
Knock	Knock with the hand on the mattress, close to the body.
Twist	Twist the wrist outwards and back.
Clean	Move the hand from left to right and vice-versa, with the arm in contact with the mattress.
Circle	Make a circle shape in the air, starting and ending closely at the same location.
Come (to me)	Move the forearm towards the arm and back.
Wave	Move the hand and arm from left to right and vice-versa, in the air.
Raise Arm	Raise the forearm until a 90° angle is formed with the body and then lower it back.

Table 4.2: Arm gestures considered for the data acquisitions. Table adapted from [2] and [3].

Activity/Movement	Description
Stand to Sit	Start standing up, then sit on the edge of the bed.
Sit to Lie	Start sitting on the edge of the bed, lie down in the bed on the back.
Lie No Move	Stay still while lying in bed.
Random Move	Random movement considered common while in bed (e.g., stretching, move the pillow, taking off the glasses).
Rotate Left/Right	Rotate the body to the left/right until lying in bed on that side.
Lie Stomach Down	Starting lying on the left/right side, rotate the body to lie in bed with the stomach down.
Lie to Sit	Starting from lying on the stomach, sit on the edge of the bed.
Sit To Stand	Starting from sitting on the edge of the bed, get up from the bed and stand.
Walking	Walk around in the room.

Table 4.3: Movements or activities considered for the data acquisitions.

Table 4.4: Features extracted from the raw sensor data, for each sliding window. All features were computed for each sensor (accelerometer and gyroscope) and each axis (x, y, and z), unless stated. Table adapted from [1].

Name	Description
Mean	Mean considering all samples
Median	Median considering all samples
Standard Deviation	Standard deviation considering all samples
Variance	Variance considering all samples
Range	Difference between maximum and minimum values of the signal
Skewness	Measure of asymmetry of the probability distribution of the signal about its mean
Kurtosis	Measure of the "tailedness" of the signal's probability distribution
Integral	Area under the curve
Correlation	Pearson's correlation coefficient, for each axis pair (xy, yz, and xz)
Sum of all squares	Sum of the squared value of all samples considering all three axes, for each sensor

4.2.3 Evaluation Approach

As mentioned above, the model was evaluated for two different solutions, subject dependent and subject independent, with the aim to investigate if it is possible to train a single model that can be used with new never-seen users, or if it is necessary to train a model for each new user.

For the subject dependent case, we applied the stratified 10-fold cross-validation approach to the data of each subject separately. This approach consists of dividing the considered dataset into 10 sub-samples with the same size, in a random way, but ensuring that the number of examples from each class are approximately the same for each sub-sample. Then, 10 iterations are performed. For each iteration, one of the sub-samples is used for testing, while the remaining 9 sub-samples are used for training. The test set is always different for each iteration.

For the subject independent case, we applied the leave-one-subject-out cross-validation (LOSO-CV) approach to the whole dataset. This approach consists of using the data from all subjects except one for training and from the remaining subject for testing. The number of iterations corresponds to the total number of subjects, where a different subject is used as the test set in each iteration. For each LOSO-CV iteration, we further used an adapted stratified 10-fold CV approach, where 10 iterations are further carried out. In each one of these inner iterations, 9 sub-samples from the training set are used to train the model, while 1 sub-sample from the test set is used for testing. The used samples are always different for each inner iteration.

Since our dataset is balanced, the following evaluation metrics were considered: overall accuracy (4.1), overall F1 score (4.2), and class F1 score (4.3). In our system, it is important to avoid false positives associated to the pre-defined gestures, i.e., avoid detecting a gesture when there is no gesture. Therefore, the false positive rate (FPR), when considering all arm gestures as the positive class and the other movements as the negative one, was also computed using (4.6).

$$\text{Overall accuracy (\%)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{4.1}$$

$$\text{Overall F1 score (\%)} = \frac{\sum_{i=1}^{C} \text{F1}(c_i)}{C} \tag{4.2}$$

$$\text{Class F1 score (\%)} = 2 \times \frac{\text{class precision} \times \text{class recall}}{\text{class precision} + \text{class recall}} \tag{4.3}$$

$$\text{Class precision (\%)} = \frac{TP}{TP + FP} \times 100 \tag{4.4}$$

$$\text{Class recall (\%)} = \frac{TP}{TP + FN} \times 100 \tag{4.5}$$

$$\text{Class FPR (\%)} = \frac{FP}{TN + FP} \times 100 \tag{4.6}$$

In (4.4), (4.5) and (4.6), TP, TN, FP and FN correspond to:

- True positives (TP): number of instances correctly classified as belonging to the considered class;
- True negatives (TN): number of instances correctly classified as belonging to a class other than the one considered;
- False positives (FP): number of instances incorrectly classified as belonging to the considered class;
- False negatives (FN): number of instances incorrectly classified as belonging to a class other than the one considered.

In (4.1), TP, TN, FP and FN correspond to the sum of the corresponding values considering all classes. In (4.2), C is the number of classes and $\text{F1}(c_i)$ is the F1 score for class c_i (computed using (4.3)).

4.2.4 Results with Complete Gesture Set

We began by verifying if a model can be built for recognizing the set of gestures listed in Table 4.2. As already explained above, this set resulted from a merge of two gestures sets investigated in two separate works [1, 10]. It is important to note that this set was considered before the gestures to be actually used by the system to support communication had been selected. Therefore, our main aim was to explore the most varied gestures possible that may be suitable for the bed scenario, and verify whether some of those gestures are easier to recognize than the others.

The accuracy, F1 score and FPR results obtained for the complete gesture set, considering all subjects and all gestures, are presented in Figure 4.3 for both subject dependent and independent cases. We can see that the results regarding accuracy, F1 score, and FPR are better for user-dependent (mean of 97%, 97%, and 9%, and median of 97%, 96%, and 0%,

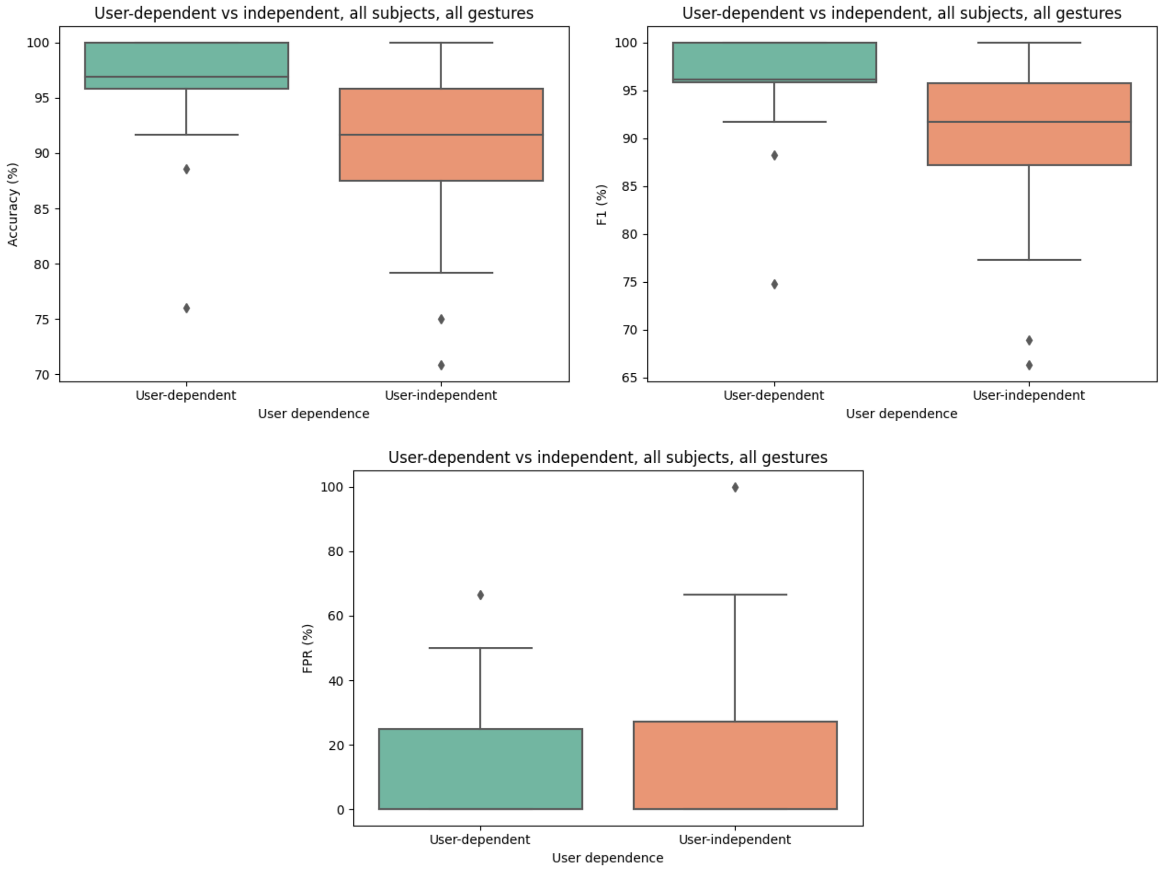

Figure 4.3: Accuracy, F1 score, and FPR values achieved when considering all subjects and all gestures, for both user dependent and independent cases.

---

respectively) than for user-independent (mean of 92%, 92%, and 13%, and median of 92%, 92%, and 0%, respectively).

The worse overall results for the independent case can be explained by the fact that the speed and range of gesture motion is naturally different from person to person, making it more difficult to classify the gestures of a never-seen subject than to perform classification for a given subject using a model trained with data from that same subject only. Nonetheless, a mean F1 score of 92% for the user independent case is still an adequate result.

Comparing with the previous work with wearables [1], the obtained F1 score results are not considerably different or are similar: mean of 97% vs 100% for subject-dependent, and 92% vs 91% for subject-independent. These results are encouraging, especially when considering that here we investigated a larger set of gestures, which were executed in several postures (instead of only one) and with both arms (instead of only the right arm).

To further analyse how the results vary among the different participants, the F1 score results per subject are presented in Figure 4.4. The corresponding mean and standard deviation (SD), as well as the median, values are included in Table 4.5.

Regarding the user dependent case, we can observe from the table that the mean F1 score ranges from 95% to 98%, which is not a considerably large range. The biggest difference can be seen in the user independent case, where the highest mean F1 score is 98% and the lowest is 84%, although only two participants in total presented a value under 90%.

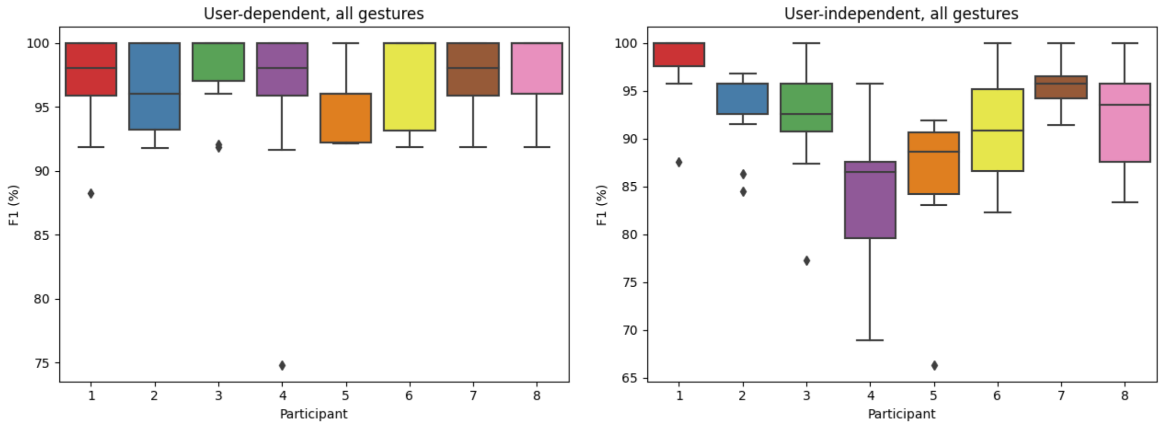

Figure 4.4: F1 score results for each participant, considering all gestures, for both user dependent and independent cases.

Participant		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	Mean	96.8	96.4	98.0	95.4	94.9	97.2	97.2	98.0
User-Dependent	**SD**	4.1	3.5	3.4	7.8	2.6	3.8	3.3	2.8
	Median	98.0	96.0	100	98.0	96.0	100	98.0	100
	Mean	98.0	93.5	92.3	84.2	86.1	90.6	95.6	92.3
User-Independent	**SD**	4.0	4.5	6.7	7.5	7.7	6.0	2.9	5.5
	Median	100	95.7	92.6	86.5	88.7	90.9	95.7	93.6

Table 4.5: Mean, standard deviation (SD), and median values for F1 score (%) for each participant, considering all gestures, for both user dependent and independent cases.

We also wanted to find out if there are gestures that are easier to recognize than others. Similar results shown for the individual subjects are presented in Figure 4.5 and Table 4.6 for each different gesture.

It is possible to see that the results are lower for the user independent case, for every gesture. Although the median is the same for all gestures except one (Circle), the mean is lower and SD is higher. Interesting to observe is that the Knock and Clean gestures rank as the best only after Wave, which pleasantly aligns with our proposal of using these first two as the input for the system.

To better understand which gestures are being confused with which, we also looked at the confusion matrix, considering all subjects, which is visually illustrated in Figure 4.6 for the user independent case, since it would be the solution to be ideally adopted in the system, avoiding training a model for each new user.

The matrix shows that the worst gesture, Circle, is mainly confused with the Come and Raise Arm gestures. Those gestures are also mostly confused with Circle and each other. As discussed above, the gestures with the best result are Wave and Knock, being only sometimes incorrectly classified as Come, and as Twist, Come, or Clean, respectively. When considering the “Other” class, the gesture that is most often confused with those other movements is

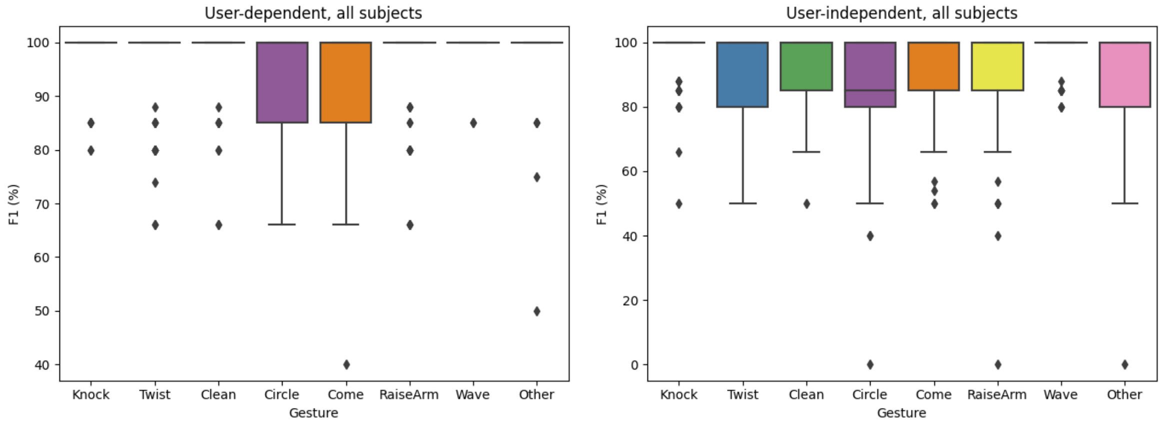

Figure 4.5: F1 score for each gesture, considering all subjects, for both user dependent and independent cases.

Gesture		**Knock**	**Twist**	**Clean**	**Circle**	**Come**	**RaiseArm**	**Wave**	**Other**
User-Dependent	**Mean**	98.6	96.3	97.8	94.9	91.3	96.7	99.6	98.5
	SD	4.6	8.3	6.7	8.5	12.0	8.2	2.2	6.7
	Median	100	100	100	100	100	100	100	100
User-Independent	**Mean**	96.6	90.5	93.5	86.1	90.6	89.5	97.7	88.0
	SD	8.5	13.8	10.5	17.3	13.0	17.1	5.7	17.9
	Median	100	100	100	85.7	100	100	100	100

Table 4.6: Mean, standard deviation (SD), and median values for F1 score (%) for each gesture, considering all subjects, for both user dependent and independent cases.

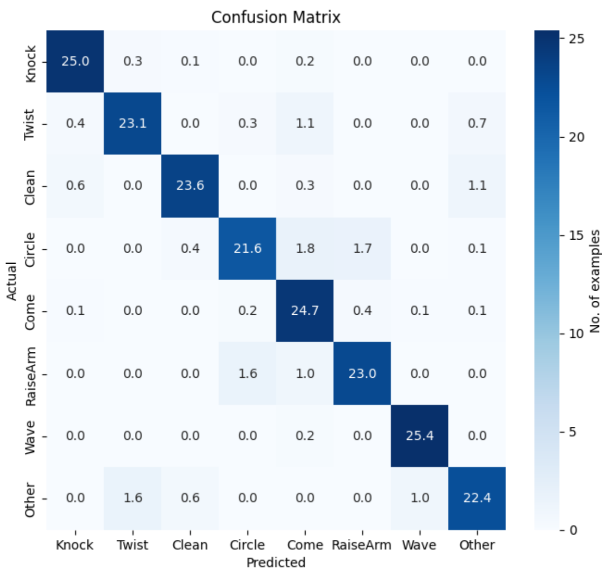

Figure 4.6: Confusion matrix for the user independent case. Each value corresponds to the sum of the mean value for all subjects.

Clean, followed by Twist. On the other hand, the other movements are sometimes mistaken for the Twist, Wave, or Clean gestures.

4.2.5 Results with Selected Gestures

As described in Section 3.5, and according to the feedback received from the SLTs concerning the use of Yes/No questions, “Knock” and “Clean” were the gestures chosen to interact with the system. For this reason, we carried out the same evaluation for only these two gestures, besides the other movements/activities. It is important to note that for this evaluation, the dataset was again balanced after selecting only the classes to be investigated, resulting in 1,296 examples (432 per class and 162 per subject). The same results presented for the complete gesture set are shown below in Figures 4.7 to 4.10 and Tables 4.7 to 4.8 for the selected gesture set.

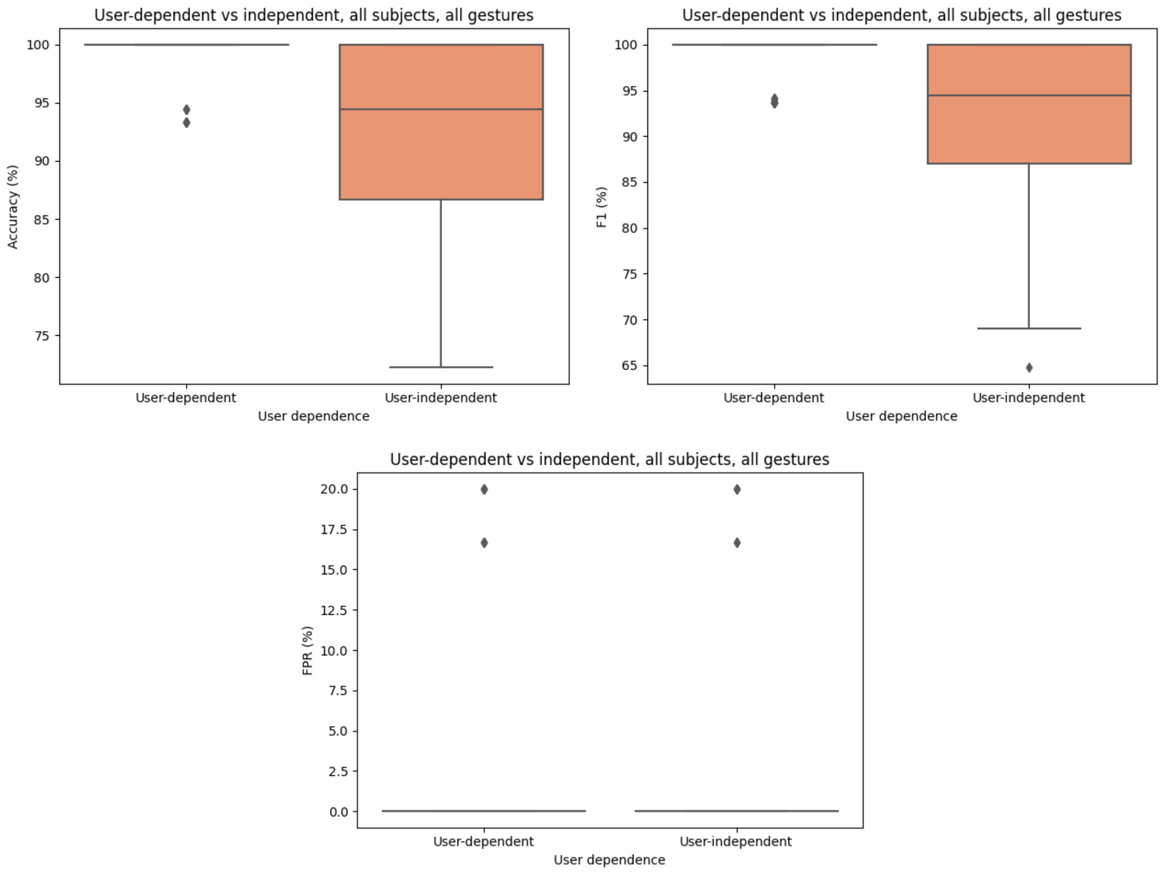

Figure 4.7: Accuracy, F1 score, and FPR values achieved when considering all subjects and the selected gestures, for both user dependent and independent cases.

For this smaller gesture set, the results regarding accuracy, F1 score, and FPR are also better for user-dependent (mean of 99%, 99%, and 3%, and median of 100%, 100%, and 0%) than for user-independent (mean of 93%, 93%, and 2%, and median of 94%, 94%, and 0%).

However, it is important to note that these results represent an improvement over the results achieved with the complete gesture set, especially for the user dependent situation.

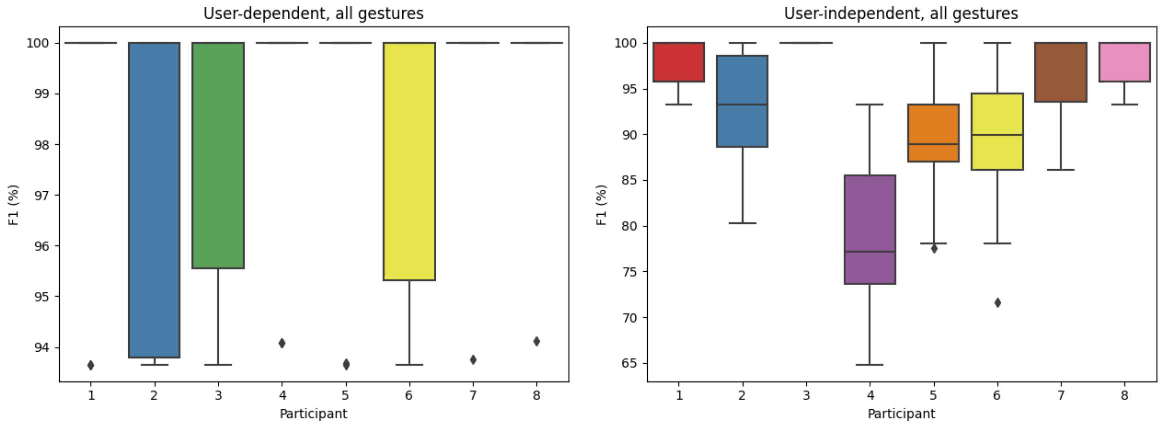

Figure 4.8: F1 score results for each participant, considering all selected gestures, for both user dependent and independent cases.

Participant		1	2	3	4	5	6	7	8
User-Dependent	**Mean**	98.7	97.5	98.1	98.8	98.7	98.1	99.4	99.4
	SD	2.7	3.2	3.0	2.5	2.7	3.0	2.0	1.9
	Median	100	100	100	100	100	100	100	100
User-Independent	**Mean**	98.1	92.5	100	78.9	89.4	89.1	96.7	98.1
	SD	3.1	6.9	0.0	8.9	7.7	9.2	4.8	3.1
	Median	100	93.3	100	77.1	88.9	90.0	100	100

Table 4.7: Mean, standard deviation (SD), and median values for F1 score (%) for each participant, considering all selected gestures, for both user dependent and independent cases.

It is also worth mentioning that, apart from some outliers, the achieved FPR results are considerably better when considering only two arm gestures (mean of 3% vs 9% for user-dependent, and mean of 2% vs 13% for user-independent). This was expected, since it should be easier to distinguish between different gestures when a smaller set is taken into account.

When analyzing the participants individually, it can be seen from Figure 4.8 and Table 4.7, that there are larger differences among subjects for the user independent case, similarly to what was observed and discussed for the most complete gesture set. Nevertheless, as referred for the general results, the F1 score values are better for most participants when considering only two gestures. The most notable exception is participant 4 for user-independent, where the mean and median values for the F1 score were of 79% vs 84%, and 77% vs 87%, respectively, for the selected vs the complete gesture set.

From Figure 4.9 and Table 4.8, we can see that, for the user dependent case, the Knock and Clean gestures obtained better F1 score results compared to those obtained with the complete gesture set: mean of 99.3% vs 98.6%, and 98.5% vs 97.8%, respectively. The results for "Other" decreased from 98.5% to 98.0%. On the other hand, for user-independent, the results were worse for Knock and Clean, while they were better for "Other": 93.2% vs 96.6%, 92.5% vs 93.5%, and 88% vs 92.8%, respectively.

From the confusion matrix shown in Figure 4.10, we can see that both Knock and Clean are more often confused with other movements than with each other. On the other hand,

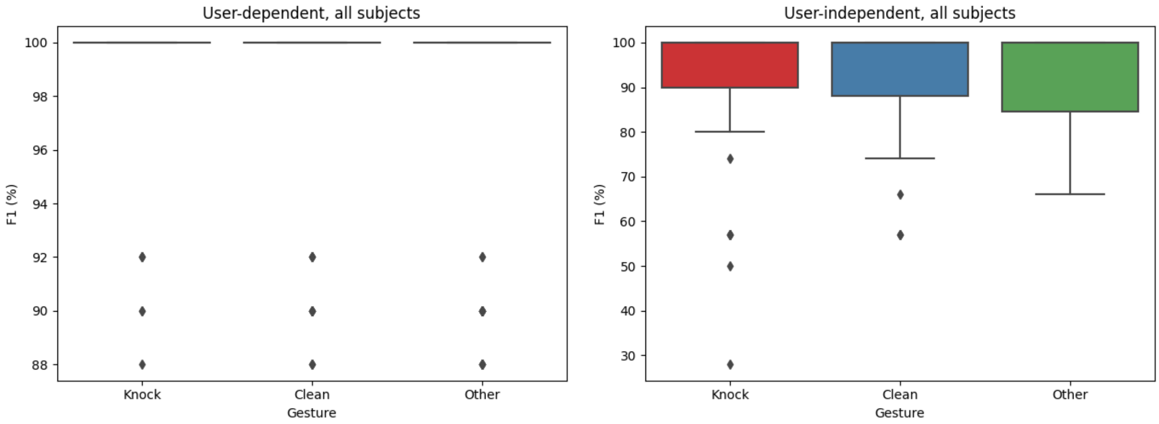

Figure 4.9: F1 score for each selected gesture, considering all subjects, for both user dependent and independent cases.

Gesture		Knock	Clean	Other
User-Dependent	**Mean**	99.3	98.5	98.0
	SD	2.3	3.4	4.0
	Median	100	100	100
User-Independent	**Mean**	93.2	92.5	92.8
	SD	13.6	10.6	8.9
	Median	100	100	100

Table 4.8: Mean, standard deviation (SD), and median values for F1 score (%) for each selected gesture, considering all subjects, for both user dependent and independent cases.

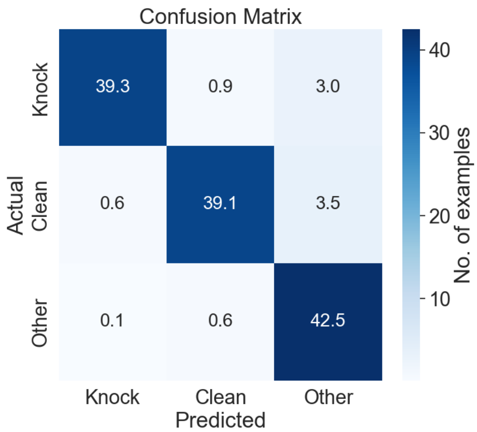

Figure 4.10: Confusion matrix for the user independent case, when considering only the selected gestures. Each value correspond to the sum of the mean value for all subjects.

---

the other movements are not frequently classified incorrectly as an arm gesture. Although it is important to correctly identify all considered gestures/movements, it is more important to minimize the number of false positives than to not recognize a pre-defined gesture. Furthermore, the results are still relatively good overall, being more balanced among the different classes than for the complete gesture set.

4.3 Summary

With the aim of training a model for the gesture input modality that is able to recognize a set of pre-defined arm gestures, in different bed-related postures and using either arm, we obtained a dataset corresponding to 8 subjects and to several postures and gestures performed with both arms.

The acquisitions for this new dataset relied on a previously implemented smartwatch application and a new version of a smartphone application, which was significantly improved with simplicity and acceleration of the recording process in mind, allowing more flexibility regarding the definition of postures, gestures, and used arm, as well as the verification and validation of each recording.

Based on the dataset, we evaluated a model for gesture recognition, and the results obtained for the dataset where we focus on the selected gestures ("Knock" and "Clean") were overall relatively good, with a mean F1 score of 93% for the user independent case, which is lower than for user-dependent (99%). Nonetheless, further performance improvement could be explored by increasing the dataset size both through having more participants providing data and having more data per participant. Techniques such as offline data augmentation and data filtering can also be used.

Regarding the dataset including all explored gestures, the results are slightly lower than for the selected gestures, with a mean F1 score of 92% and 97% for user-independent and dependent, respectively. Although not all gestures are expected to be used in the system proposal, these results are still relevant towards the possibility of developing solutions for other scenarios that can be explored in the future, such as gesture-based interaction with smart homes.

Based on the obtained results, a new model for gesture recognition was trained using the 8-subject dataset and integrated into the gesture input modality, which can now be used regardless of the posture in bed and the arm used to execute the gestures. The overall good results regarding the "Knock" and "Clean" gestures also reinforce the decision in the proposal of using these two to interact with the system. In the following chapter, we describe the implementation of the gesture modality, as well as the other modalities and applications of the system.

Chapter 5

Supporting Communication in the Bed Scenario

In this chapter, a detailed description of the implemented prototype will be given. We will start by looking at the general overview of the implemented prototype and the differences between the implemented and the initially proposed envisioned system. This will be followed by a description of every modality and application of the system, and how they communicate with each other.

5.1 Setup and Software Overview

The system overview shown in the conceptualization of the system is still relevant after the implementation, having suffered small adaptations to deal with the obstacles presented in the development. The implementation architecture can be seen in Figure 5.1 where the major difference was the extra service required to communicate with the phone application, as will be explained later. For the output modalities, an output manager was also added to work as intermediary between the IM and both the output modalities.

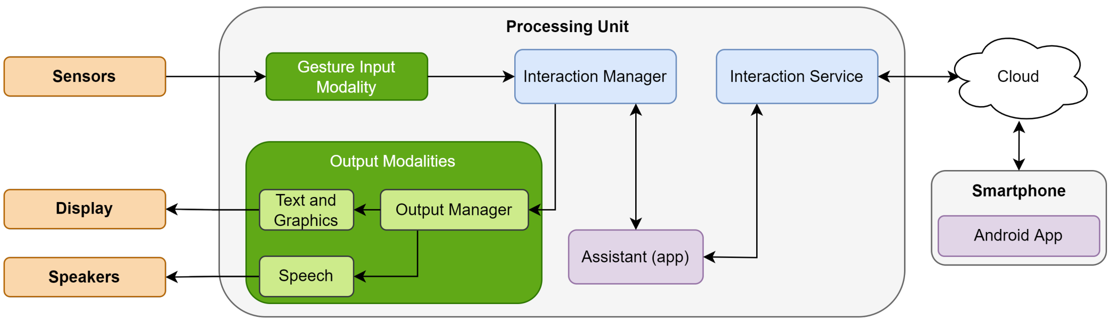

Figure 5.1: Implemented architecture with the interaction modalities, interaction manager, and applications of the system.

Regarding the used hardware, in this prototype, the sensors included wearable sensors only, namely a 3D accelerometer, gyroscope, and magnetometer embedded in a Wear OS smartwatch (Oppo Watch, in this case). The processing unit was a Raspberry Pi 4 Model B (8 GB RAM). Regarding the smartphone, a minimum version of Android 11 is required.

5.2 Gesture Input Modality

The system enables input by the PWA through the execution of specific arm gestures. This input is possible through a gesture input modality. This modality relies on the smartwatch, which is attached to the PWA's right wrist. A Wear OS application running on the device transmits the sensor data to the gesture input modality, running in the processing unit, over Bluetooth (Fig. 5.1), at a rate of 50 Hz.

The pipeline corresponding to the modality that results in a gesture decision stems from the wearable work [1] and can be seen in Fig 5.2. The data collection and feature extraction modules take care of receiving the sensor data and extract relevant features from this data. From these features, classification is then performed by a gesture recognition model trained with the dataset of the selected gestures described in section 4.2.5. Finally, a decision regarding the performed gesture is made based on three consecutive windows, and the meaning defined for that gesture is sent to the assistant through the Interaction Manager (IM), which will be detailed ahead in section 5.4.

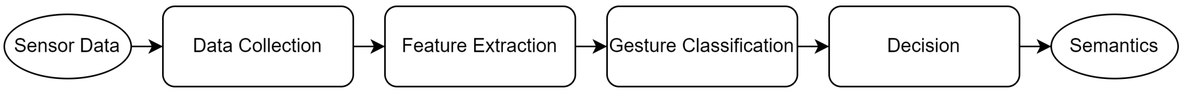

Figure 5.2: Different modules of the gesture input modality. Figure adapted from [1].

Abiding to the AM4I framework which is aligned with the W3C recommendations for a multimodal system, the messages sent by the modality make use of the IM, which replaced the webservice implemented in previous work [1]. Furthermore, as mentioned above, the gesture modality sends the semantics of the gesture performed, and not the gesture itself. For instance, in this prototype, the Knock gesture is associated to an affirmative response, and the Clean gesture to a negative response, with the modality sending only this semantic value to the assistant. This grants a lot of flexibility towards future expansion, where gestures can be replaced any time or be associated to other meanings, while maintaining the gestures decoupled from the actions in the assistant, which will be described in detail ahead.

5.3 Assistant

The assistant follows the concept of a state-machine that changes state (i.e., advances to the next question or action) depending on user input. It was implemented in Python and the interaction flow can be seen in Figure 5.3. The gestures corresponding to Yes/No are Knock and Clean.

Starting the interaction, activation begins by use of the Knock gesture, with the question "Need immediate help?" being sent to the output modalities (which present the question to the PWA). After activation, the states happen according to the gesture meanings, which are defined as "Affirmative" or "Negative" according to the Gesture Input modality.

If the assistant receives an affirmative answer to the first question, a message is sent immediately to the caregiver. Following the sending of this message, context of the help needed is asked for and once the caregiver confirms he saw the message, reassurance is given by displaying that "Help is on the way!". This set of interactions correspond to those of highest priority.

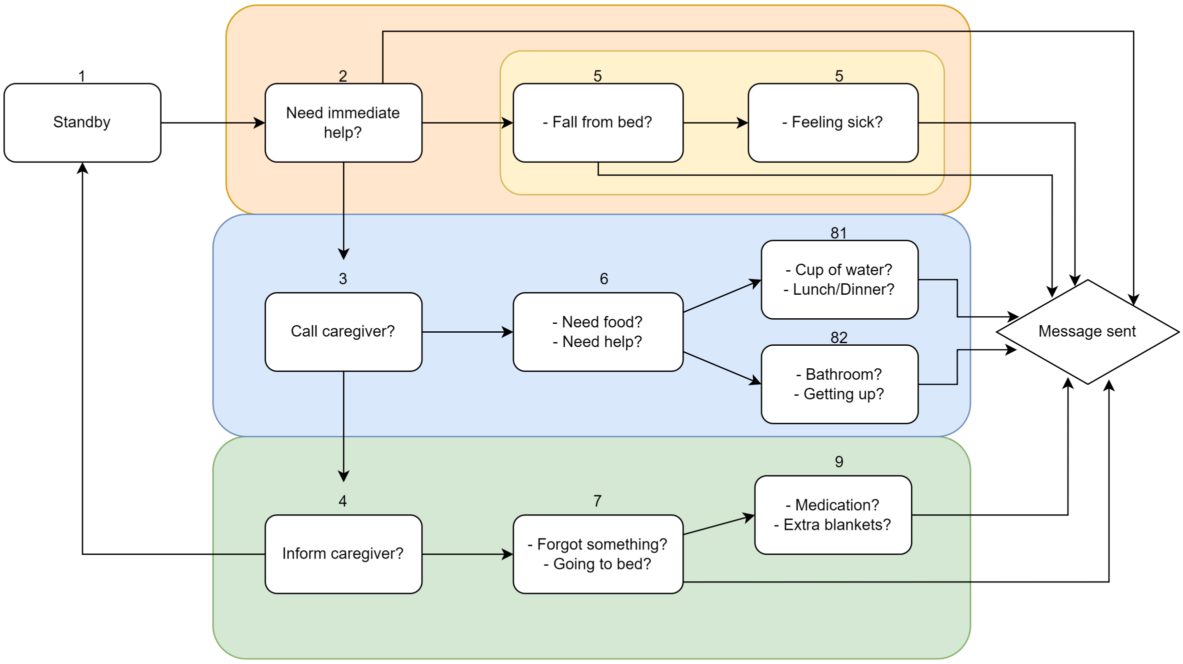

Figure 5.3: Diagram representing the interaction flow between the assistant and the PWA. The big three blocks, which are also colored, represent priority for what the PWA needs to convey.

If the answer to the first question (need of immediate help) is negative, then the assistant sends the question "Call caregiver?" (medium priority) to the PWA. The two main questions after are "Need food?" and "Need help?", with both offering a more detailed option if the answer is affirmative to any of them.

If the answer to the first medium priority question ("Call caregiver?") is negative, the assistant asks the question "Inform caregiver?" (lowest priority). If the answer is positive, the remaining sequence is similar to the medium priority block, but with a different set of pre-defined questions (see Figure 5.3).

If the answer to the question "Inform caregiver?" is negative, then the assistant returns to standby mode. When inside a more specific set of questions on a specific priority, if all questions receive "No" as an answer, then they return to the previous block. The assistant also goes to standby mode after sending a message.

Mentioned above but now more detailed, it is important to note that the assistant does not actually receive the specific gesture that was used; rather, it receives a message that contains the semantics, or the meaning of the interaction action that was performed by the user (affirmative/negative), decided in the gesture input modality. This is important to ensure a decoupling between the interaction and the application enabling alteration of the considered gestures without having to change the assistant. Additionally, and as a result of using a multimodal architecture, the assistant will function in the same way if another modality sends the same semantics, such as a touch modality that enables pressing buttons on a screen.

5.4 Communication Support Infrastructure

The information flow between the various modalities and applications is managed by the interaction manager (IM), which is an integral component of the AM4I framework [58]. Towards enabling the exchange of messages between the various components and the IM, an API in Python was implemented to allow the modalities and assistant to send and receive information to and from the IM using Life Cycle Events, containing the data formatted in accordance with the W3C EMMA (Extensible MultiModal Annotation) standards.

The important information that travels in the messages are: (1) Source - where it comes from; (2) Target - which modality or application is meant to be reached; (3) Data in JSON - where all the information necessary for the next step is saved.

Inside the JSON data, the content depends both from source and target module or application it is being sent to. The following is a more detailed description of the data according to the possible paths:

1. Gesture Input Modality to Assistant - The data contains the meaning of the gesture decided by the input modality, and the language in which the system will give feedback.

2. Assistant to Output Modalities - The output modalities receive the text meant to be displayed and said out loud to represent the current state, and the name of the image associated with the text.

3. Assistant to Android Application - The phone application is meant to receive a text representing the assistance requested by the PWA, and the timestamp of when it was associated.

4. Android Application to Assistant - A text is sent confirming that assistance is on the way.

The proposed design of the system required for every modality and application to communicate among them through the use of the IM. An alternative was implemented for the smartphone application, due to the fact that the IM uses HTTPS in its communication, and the android application raised security concerns and demanded the use of certificates in order to establish communication with the IM. This alternative comes in the form of an Interaction Service that communicates directly with the assistant, and makes use of Flask as framework to serve as a bridge between the IM and the android application.

5.5 Multimodal Output

When it came to the output, the need for multimodality was a clear conclusion from both the research done about Aphasia and the discussion and feedback given by the SLTs. The output modality is envisioned as a sophisticated system where the decoupled modalities would be resourceful enough to produce aphasia friendly content according to the message received.

Currently, a first version with redundant use of speech, graphics, and text was the approach taken, where speakers would allow for audio to be played, and a visual display unit would allow for written text and images to be shown to the PWA. An output manager guarantees that each modality receives the information needed in order to provide the feedback to the PWA by making the message available through a web app implemented in Python using Flask, allowing for a decoupling of both modalities from the system.

Speech - The speech output modality converts a message from text to speech, using the Python's package "pyttsx3" that enables offline text-to-speech. The resulting audio is played using speakers connected to the processing unit. The message received comes from the polling to the output manager, which translates into a speech modality that is decoupled from the rest of the system.

Graphics and Text - In order to allow for display, a web page was implemented, using Javascript and the front-end library "React", which can be accessed through a web browser. This web page constantly polls the output manager to receive the message intended to be displayed. An example of how the text and graphics information is presented to the PWA can be seen in Figure 5.4, where the shown images were manually defined in the graphics modality. Whenever the caregiver confirms that the message was seen, that information is displayed and highlighted with a green color to make it the most obvious possible, which can be seen in Figure 5.5.

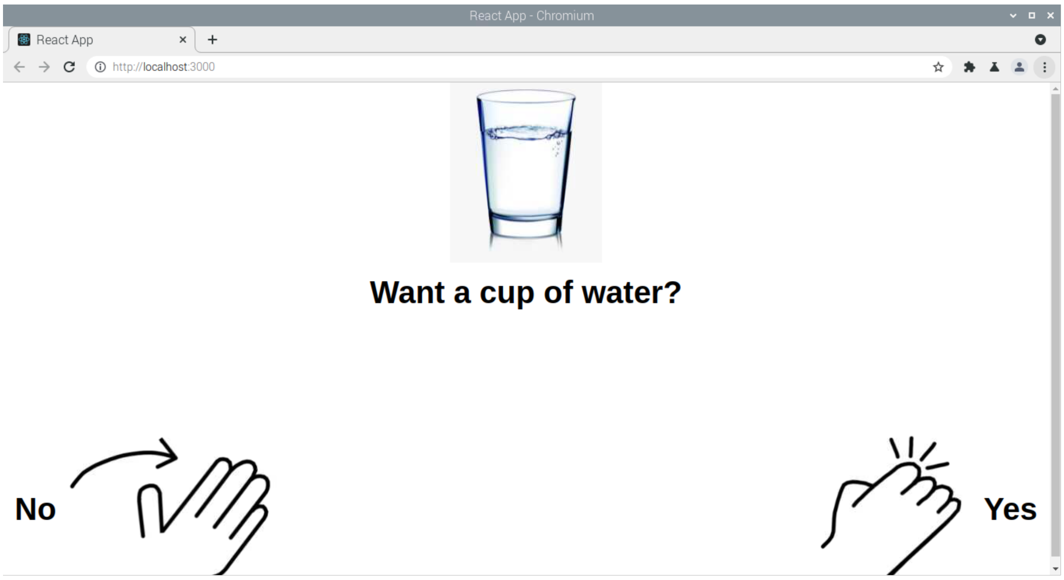

Figure 5.4: Screenshot of how the information is presented to the person.

5.6 Smartphone Application

In the context of the APH-ALARM project, an application for Android smartphones had already been developed [1] to be used, for example, by a caregiver or family member responsible for the PWA. What started as a linear process of a gesture having a direct meaning, the application now allows composed information to be exchanged between the smartphone application and the rest of the applications and modalities. This increases the versatility and the range of communication between PWA and caregiver.

The visual design was simplified, and the code required was adapted to the infrastructure that sends the message, and to the information received in itself.

The main features of the application can be described as:

1. Receiving the context of the assistance required by the PWA, with a timestamp included;

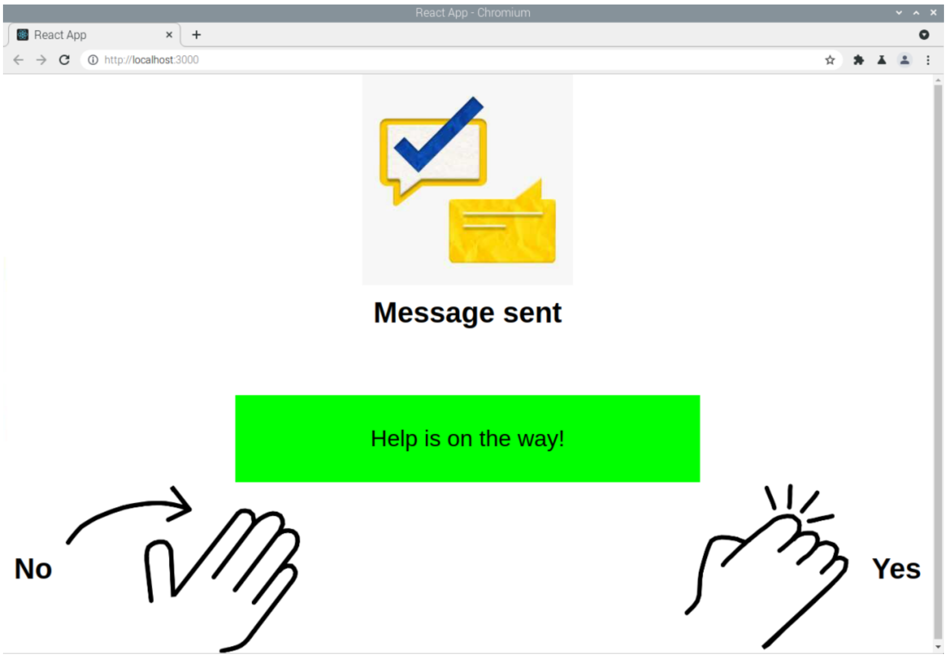

Figure 5.5: Screenshot of how the information is presented to the person after receiving confirmation by the caregiver application that the message was seen.

2. Confirming that the message was seen, and send back a message saying the caregiver is moving towards assisting the PWA.

The application is able to receive the message through the polling of the interaction service mentioned in the section 5.4. An HTTP GET request is sent periodically, which means the application requires stable connection to the Internet. The display can be seen in Figure 5.6, where the message is displayed with a timestamp, in which the confirmation button only becomes active after a message has been received.

5.7 Summary

In this chapter we started the description of the implementation by showing an overview of the system and mentioning the used hardware. In succession, the implementation of all the applications and modalities is explained.

The gesture input modality makes use of the new recognition model described in the previous chapter to provide output in the form of gesture meaning, replacing a direct gesture with its semantics. The gestures used, knock and clean, take into account being deemed the easiest to execute and understand their meaning.

The assistant makes use of all the information gathered about communicating with a PWA. It implements yes/no questions in their most concise form, questions which represent the most relevant and needed situations for a PWA to ask for assistance.

The interaction manager guarantees an organized flow of information between all the modalities and applications, abiding to the W3C multimodal architecture recommendations, and guaranteeing scalability for the system.

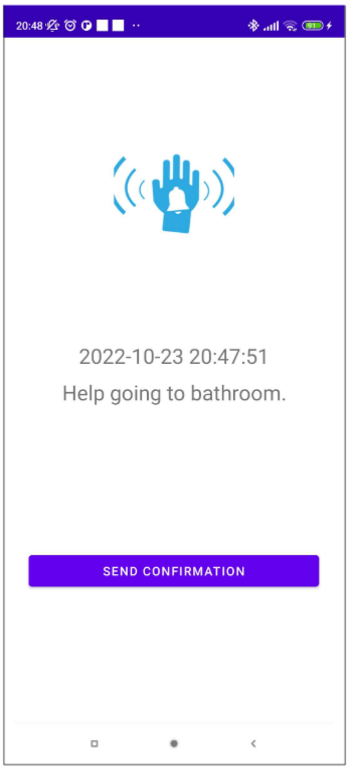

Figure 5.6: Display of the smartphone application, showcasing a request by the PWA to be assisted going to the bathroom.

With multimodality being a key concept, we apply redundant use of speech, graphics and text to provide feedback to the PWA. Due to being decoupled from each other, each modality has the potential to be evolved individually towards a more aphasia friendly design.

The smartphone application is able to receive a request with its context, displaying the exact type of assistance required by the PWA, and sending a confirmation back that the request is being addressed.

The system was implemented successfully, and with scalability taken into account, it shows a lot of potential for its continuation and improvement.

Chapter 6

Conclusions

6.1 Work Summary

The work developed in this dissertation had an initial approach of research and problem definition, followed by a conceptualization and its evaluation, a more final design of the system and finally its implementation.

The focus of this work was finding a solution to help someone with aphasia communicate remotely with their caregiver, in their bed, at any time of the day without raising privacy and intrusiveness concerns. This prompted a literature review on the concepts of aphasia, augmented and alternative communication and its tools and methods, gesture recognition methods that address the privacy and intrusiveness concerns, and an user centered design methodology of development. A guided discussion related to relevant previous work in the APH-ALARM project was also carried out.

Part of the UCD methodology, personas and scenarios relevant for the system usage were refined to the target context, particularly regarding their motivation, allowing us to deepen our understanding of the target users. Moreover, focus groups were carried out with SLTs, where an initial concept proposal and low fidelity sketches of the system were presented and evaluated, and a discussion about PWAs needs and motivations was had. From here on, the requirements of the system were extracted and a proposal of the system follows through, taking into account all the knowledge gathered.

Preceding the implementation of the system, both works in the context of the APH-ALARM project that explored gesture recognition through wearables and radars were dissected and analysed towards the possibility of integration with this work. The decision of building on top of the wearable work implied transforming the gesture recording application to be faster and simpler, and collecting a new dataset of gestures. A new model was trained and evaluated using this new dataset. Still part of the wearable work, the gesture input modality, the speech modality and the Android application were all scrutinized and adapted towards the new design and structure of the system.

From here on the focus was on the development of the assistant, the output modalities of text and graphics, the integration of the interaction manager that would guarantee the communication between the system's modalities and applications, and generally assuring everything was functioning together.

By the time the proof-of-concept was functional, the remaining effort went into testing the robustness of the prototype.

6.2 Main Results and Contributions

In this work, a vision of extending the potential of communication in the bed using gestures gained shape in several forms.

The implemented gesture input modality makes use of a new gesture recognition model to provide output, which has a meaning, and is not a gesture itself. This gesture recognition model stems from data acquired from an experiment with 8 participants, carried out in this work. The data acquisition process was evolved to be simpler and faster compared to the previous work. When evaluating the new model with the selected gestures that enables the execution of gestures with both arms in various postures, a mean F1 score of 99% and 93% was achieved, for all gestures, in the user dependent and independent cases, respectively.

The assistant developed for the system benefits from all the knowledge collected previously and follows the recommendations to gather the relevant information to provide appropriate assistance. Due to the use of semantics, it is able to work with any type of input that applies the same semantics.

The feedback to the PWA is presented through use of text, graphics and speech in order to provide redundant feedback, an unanimous recommendation as beneficial when trying to convey a message to a person with aphasia. This multimodal output makes use of an output manager that resembles a fission module, managing the decoupled output modalities, currently a first version of what can be a sophisticated approach of adaptive aphasia friendly feedback.

Furthermore, not just the output modalities, but all modalities and applications of the service are decoupled, and pulled together by the interaction manager, guaranteeing communication between them. The decoupling takes into account scalability of the system, and allows for independent changes to each application and modality without affecting the general functionality.

This communication between the modalities and applications allows for fully contextualized messages to be sent, reflected in the evolved smartphone application, that provides the caregiver with a contextualized request, and opens a path for a versatile and wide range of bilateral remote communication.

All of these resulted in a proof-of-concept of a system that allows for remote communication between a person with aphasia and their caregiver or relative. In addition, the research gathered through both literature and the discussions carried out with the SLTs when refining the conceptualization of the system is a major contribution. The results of the discussions are a great asset for any future study and work looking to provide means of communication assistance to a person with aphasia.

Overall, the system shows potential to provide higher levels of reassurance to the PWA and their caregiver and relatives while the PWA is laying in bed. It is also reasonable to consider it a first step in providing communication support in the whole house to any person with communication disabilities, and not only to a person with Aphasia.

6.3 Future Work

This proof-of-concept system was developed taking scalability into account, allowing for its evolution in several aspects. The following are observations and suggestions towards the technical implementation, the conceptualization, and future work with the intention of making the system practical in the life of a person with aphasia.

- One of the most important aspects to address in the future is the evaluation of the prototype. Asking SLTs and possible end users (i.e., PWAs) to perform practical use of the system with the intention of evaluating it is important.
- Taking a look at the input modality, the framework that supports the concept of multimodality, and the followed W3C recommendations that translate into a decoupling of the modalities, opens the possibility for both substitution of an input by another (e.g., button, voice command), or joining several together, admitting more than one source of input data (e.g. another sensor such as radar).
- An increased size of the dataset can benefit the performance of the gesture recognition model. Methods such as data augmentation to increase not only the dataset's size, but also its variety, and data filtering to remove noise/outliers, could also be beneficial.
- The implementation of the assistant can be improved by detaching the list of questions from the code, towards an improved management of the needs of each individual.
- When considering the output modalities, each modality is currently in its initial version with room for improvement in each. A more advanced text-to-speech service, such as the one provided by Microsoft with a less robotic voice could better ensure the understandability of the spoken message by the PWA. The information conveyed by the graphics and text modalities through the display can also benefit from an improved design, with the possibility to change the font size, image size, the used images, etc.
- Still concerning the output, the concept of an autonomous approach to transform common messages into aphasia friendly content is a bold future endeavour towards a more sophisticated output modality. The output manager is currently resembling a fission module, paving a path for this autonomous approach of the output modalities.
- The smartphone application allows the caregiver to confirm that they have seen the message and intend to provide assistance. The ideal use would imply providing the possibility of pre-defined options to use, or even have self-written responses to extend the range of communication with the PWA. Following this idea, the PWA should have the ability to answer the questions directly sent by the caregiver.
- When taking into consideration that the information gathered by the assistant may be sensitive to the user, and this information leaves the local network to reach the smartphone, implementation of security should be considered.

Due to the design of the system, which abides to the W3C multimodal architecture recommendations, new functionalities should suffer minimal resistance from the core of the system, given the decoupling and flexibility of the applications and modalities.

Furthermore, all these new paths for development made possible by the work carried out in this dissertation show potential for its continuation and improvement, with the choices made throughout the work assuring that new options were open, and not limited.

Bibliography

[1] Afonso M. Guimaraes. Bedfeeling - sensing technologies for assistive communication in bed scenarios. Master's thesis, University of Aveiro, Aveiro, Portugal, 2021.

[2] Afonso Guimarães, Ana Patrícia Rocha, Luís Santana, Ilídio C. Oliveira, José Maria Fernandes, Samuel Silva, and António Teixeira. Enhanced communication support for aphasia using gesture recognition: The bedroom scenario. In *2021 IEEE International Smart Cities Conference (ISC2)*, pages 1–4, 2021.

[3] Luís Santana, Ana Patrícia Rocha, Afonso Guimarães, Ilídio C. Oliveira, José Maria Fernandes, Samuel Silva, and António Teixeira. Radar-based gesture recognition towards supporting communication in aphasia: The bedroom scenario. In Takahiro Hara and Hirozumi Yamaguchi, editors, *Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 500–506, Cham, 2022. Springer International Publishing.

[4] Andrew D Palmer, Jason T Newsom, and Karen S Rook. How does difficulty communicating affect the social relationships of older adults? an exploration using data from a national survey. *Journal of communication disorders*, 62:131–146, 2016.

[5] Audrey L Holland. The value of "communication strategies" in the treatment of aphasia. *Aphasiology*, 35(7):984–994, 2021.

[6] Shannon M Sheppard and Rajani Sebastian. Diagnosing and managing post-stroke aphasia. *Expert review of neurotherapeutics*, 21(2):221–234, 2021.

[7] Heather L Flowers, Stacey A Skoretz, Frank L Silver, Elizabeth Rochon, Jiming Fang, Constance Flamand-Roze, and Rosemary Martino. Poststroke aphasia frequency, recovery, and outcomes: A systematic review and meta-analysis. *Archives of Physical Medicine and Rehabilitation*, 97(12):2188–2201.e8, 2016.

[8] How common is aphasia? *Author: National Aphasia Association, https://www.aphasia.org/aphasia-faqs/*. Accessed October 12, 2022 [Online].

[9] Yasmin Elsahar, Sijung Hu, Kaddour Bouazza-Marouf, David Kerr, and Annysa Mansor. Augmentative and alternative communication (aac) advances: A review of configurations for individuals with a speech disability. *Sensors*, 19(8):1911, 2019.

[10] Luís Fernando do Vale Santana. Exploring radar sensing for gesture recognition. Master's thesis, University of Aveiro, Aveiro, Portugal, 2021.

[11] Ana Patrícia Rocha, Afonso Guimarães, Ilídio C. Oliveira, Fábio Nunes, José Maria Fernandes, Miguel Oliveira e Silva, Samuel Silva, and António Teixeira. Toward supporting communication for people with aphasia: The in-bed scenario. In *Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[12] José Fonseca. *Afasia e comunicação após lesão cerebral. Definição, classificação e reabilitação.* 05 2018.

[13] João Lobo Antunes. A herança de Franz Joseph Gall. o cérebro ao serviço do comportamento humano, 2001.

[14] Andrew Kertesz. Anomia. In Michael J. Aminoff and Robert B. Daroff, editors, *Encyclopedia of the Neurological Sciences*, pages 193–198. Academic Press, New York, 2003.

[15] Echolalia. *Collins Dictionary.* Accessed February 28, 2022 [Online].

[16] Monica Blom Johansson, Marianne Carlsson, and Karin Sonnander. Communication difficulties and the use of communication strategies: from the perspective of individuals with aphasia. *International journal of language & communication disorders*, 47(2):144–155, 2012.

[17] Li Huang, Szu-Han Kay Chen, Shutian Xu, Yongli Wang, Xing Jin, Ping Wan, Jikang Sun, Jiming Tao, Sicong Zhang, Guohui Zhang, et al. Augmentative and alternative communication intervention for in-patient individuals with post-stroke aphasia: Study protocol of a parallel-group, pragmatic randomized controlled trial. 2021.

[18] Meghan Allen, Joanna McGrenere, and Barbara Purves. The design and field evaluation of phototalk: A digital image communication application for people. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '07, page 187–194, New York, NY, USA, 2007. Association for Computing Machinery.

[19] Shaun K. Kane, Barbara Linam-Church, Kyle Althoff, and Denise McCall. What we talk about: Designing a context-aware communication tool for people with aphasia. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, page 49–56, New York, NY, USA, 2012. Association for Computing Machinery.

[20] Kaveet Laxmidas, Cory Avra, Christopher Wilcoxen, Michael Wallace, Reed Spivey, Samantha Ray, Seth Polsley, Puneet Kohli, Julie Thompson, and Tracy Hammond. Commbo: Modernizing augmentative and alternative communication. *International Journal of Human-Computer Studies*, 145:102519, 2021.

[21] Daniel Guasch, Israel Martín-Escalona, José A Macías, Virginia Francisco, Raquel Hervás, Lourdes Moreno, and Susana Bautista. Design and evaluation of eco: an augmentative and alternative communication tool. *Universal Access in the Information Society*, pages 1–23, 2021.

[22] Mmachi God'sglory Obiorah, Anne Marie Marie Piper, and Michael Horn. Designing aacs for people with aphasia dining in restaurants. In *Proceedings of the 2021 CHI*

Conference on Human Factors in Computing Systems, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.

[23] Rúbia E. O. Schultz Ascari, Luciano Silva, and Roberto Pereira. Personalized interactive gesture recognition assistive technology. In *Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems*, IHC '19, New York, NY, USA, 2019. Association for Computing Machinery.

[24] Ana Patrícia Rocha, Afonso Guimarães, Ilídio C. Oliveira, Fábio Nunes, José Maria Fernandes, Miguel Oliveira e Silva, Samuel Silva, and António Teixeira. Toward supporting communication for people with aphasia: The in-bed scenario. In *Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '22 Adjunct)*, MobileHCI '22 Adjunct, New York, NY, USA, 2022. Association for Computing Machinery. (accepted for publication).

[25] Mais Yasen and Shaidah Jusoh. A systematic review on hand gesture recognition techniques, challenges and applications. *PeerJ Computer Science*, 5:e218, 2019.

[26] Siddharth S Rautaray and Anupam Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial intelligence review*, 43(1):1–54, 2015.

[27] Munir Oudah, Ali Al-Naji, and Javaan Chahl. Hand gesture recognition based on computer vision: a review of techniques. *journal of Imaging*, 6(8):73, 2020.

[28] Narayanan C Krishnan, Colin Juillard, Dirk Colbry, and Sethuraman Panchanathan. Recognition of hand movements using wearable accelerometers. *Journal of Ambient Intelligence and Smart Environments*, 1(2):143–155, 2009.

[29] Shahzad Ahmed, Karam Dad Kallu, Sarfaraz Ahmed, and Sung Ho Cho. Hand gestures recognition using radar sensors for human-computer-interaction: A review. *Remote Sensing*, 13(3):527, 2021.

[30] Gongfa Li, Hao Wu, Guozhang Jiang, Shuang Xu, and Honghai Liu. Dynamic gesture recognition in the internet of things. *IEEE Access*, 7:23713–23724, 2018.

[31] Zhiyuan Lu, Xiang Chen, Qiang Li, Xu Zhang, and Ping Zhou. A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices. *IEEE transactions on human-machine systems*, 44(2):293–299, 2014.

[32] Pyeong-Gook Jung, Gukchan Lim, Seonghyok Kim, and Kyoungchul Kong. A wearable gesture recognition device for detecting muscular activities based on air-pressure sensors. *IEEE Transactions on Industrial Informatics*, 11(2):485–494, 2015.

[33] Yang Zhang and Chris Harrison. Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 167–173, 2015.

[34] Jian Wu and Roozbeh Jafari. Orientation independent activity/gesture recognition using wearable motion sensors. *IEEE Internet of Things Journal*, 6(2):1427–1437, 2018.

[35] SANOAL MACHADO and SANTIAGO MANCHENO. Automotive fmcw radar development and verification methods. Master's thesis, 2018.

[36] Simon Kingsley and Shaun Quegan. *Understanding radar systems*, volume 2. SciTech Publishing, 1999.

[37] Xiangpeng Liang, Haobo Li, Weipeng Wang, Yuchi Liu, Rami Ghannam, Francesco Fioranelli, and Hadi Heidari. Fusion of wearable and contactless sensors for intelligent gesture recognition. *Advanced Intelligent Systems*, 1(7):1900088, 2019.

[38] Jörg Appenrodt, Ayoub Al-Hamadi, and Bernd Michaelis. Data gathering for gesture recognition systems based on single color-, stereo color-and thermal cameras. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 3(1):37–50, 2010.

[39] Francesco Pittaluga, Aleksandar Zivkovic, and Sanjeev J. Koppal. Sensor-level privacy for thermal cameras. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, 2016.

[40] Francesco Pittaluga, Aleksandar Zivkovic, and Sanjeev J Koppal. Sensor-level privacy for thermal cameras. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2016.

[41] Afonso Guimarães, Ana Patrícia Rocha, Luís Santana, Ilídio C. Oliveira, José Maria Fernandes, Samuel Silva, and António Teixeira. Enhanced communication support for aphasia using gesture recognition: The bedroom scenario. In *2021 IEEE International Smart Cities Conference (ISC2)*, pages 1–4, 2021.

[42] Luís Santana, Ana Patrícia Rocha, Afonso Guimarães, Ilídio C. Oliveira, José Maria Fernandes, Samuel Silva, and António Teixeira. Radar-based gesture recognition towards supporting communication in aphasia: The bedroom scenario. In Takahiro Hara and Hirozumi Yamaguchi, editors, *Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 500–506, Cham, 2022. Springer International Publishing.

[43] Lasse Skovgaard Jensen, Ali Gürcan Özkil, Niels Henrik Mortensen, et al. Prototypes in engineering design: Definitions and strategies. In *Ds 84: proceedings of the design 2016 14th international design conference*, pages 821–830, 2016.

[44] Samuel Marie-Rose, Giovanny Arbelaez, and Mauricio Camargo. A referential framework for the choice of prototypes in the design phase of a mobile application. In *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pages 1–7. IEEE, 2019.

[45] Matthew Schmidt, Yvonne Earnshaw, Andrew A Tawfik, and Isa Jahnke. Methods of user centered design and evaluation for learning designers. *Learner and user experience research*, 2020.

[46] Larry Burns. *Human-centered Design*. Technics Publications, 2018.

[47] Catherine Burns. *Human-centred design*, pages 207–227. 05 2018.

[48] Nielsen usability heuristics. *Author: Jakob Nielsen, nngroup, https://www.nngroup.com/articles/ten-usability-heuristics*. Accessed April 5, 2022 [Online].

[49] Shneiderman eight golden rules. *University of Maryland, http://www.cs.umd.edu/ ben/goldenrules.html*. Accessed August 9, 2022 [Online].

[50] Bruce Tognazzini. Principles, techniques, and ethics of stage magic and their application to human interface design. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, pages 355–362, 1993.

[51] Ahlem Assila, Houcine Ezzedine, et al. Standardized usability questionnaires: Features and quality focus. *Electronic Journal of Computer Science and Information Technology*, 6(1), 2016.

[52] Hamed Taherdoost. What is the best response scale for survey and questionnaire design; review of different lengths of rating scale/attitude scale/likert scale. *Hamed Taherdoost*, pages 1–10, 2019.

[53] W Mieke E van de Sandt-Koenderman, Jiska Wiegers, Sandra M Wielaert, Hugo J Duivenvoorden, and Gerard M Ribbers. A computerised communication aid in severe aphasia: An exploratory study. *Disability and rehabilitation*, 29(22):1701–1709, 2007.

[54] Melanie Fried-Oken, David R Beukelman, and Karen Hux. Current and future aac research considerations for adults with acquired cognitive and communication impairments. *Assistive Technology*, 24(1):56–66, 2012.

[55] Cátia Azevedo. "afluentia: Communication support for fluent aphasia. Master's thesis, University of Aveiro, 2022.

[56] Aleksander Jarzebowicz and Paweł Weichbroth. A systematic literature review on implementing non-functional requirements in agile software development: Issues and facilitating practices. In *International Conference on Lean and Agile Software Development*, pages 91–110. Springer, 2021.

[57] Martin Glinz. On non-functional requirements. In *15th IEEE international requirements engineering conference (RE 2007)*, pages 21–26. IEEE, 2007.

[58] Nuno Almeida, António Teixeira, Samuel Silva, and Maksym Ketsmur. The AM4I architecture and framework for multimodal interaction and its application to smart environments. *Sensors*, 19(11), 2019.

[59] Deborah A Dahl. The W3C multimodal architecture and interfaces standard. *Journal on Multimodal User Interfaces*, 7(3):171–182, 2013.