Design and Implementation of a Novel EEG-based Brain-Computer Interface to Improve "Perception - Understanding - Action" in Humanoid Robotics

Politecnico di Bari

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program
SSD: ING-INF/01– ELECTRONICS

Final Dissertation

Design and Implementation of a Novel EEG-based Brain-Computer Interface to Improve "Perception - Understanding - Action" in Humanoid Robotics

by
MEZZINA GIOVANNI

Supervisor:

Prof. De Venuto Daniela

Coordinator of Ph.D. Program:
Prof. Mario Carpentieri

Course n°35, 01/11/2019-31/10/2022

Table of Contents

Abstract

The recent global increase in average life expectancy has led to an inevitable increase in the elderly population (22% in 2050), with an (unrelated) reduction in the young population. The increase in life expectancy leads to negative consequences due to the onset of diseases related to aging causing physical and social disabilities of the elderly. This condition leads to the need for hospitalizations (with a consequent increase in the costs at the expense of the sanitary system) or 24/7 care in dedicated facilities that often suffer from dedicated specialized personnel shortage. Home care through dedicated caregivers is also following the same trend. The context of Ambient Assisted Living (AAL) proposes a solution to the problem by equipping domestic or assistive infrastructures with intelligent sensor networks scattered throughout the environment, and robotic platforms (i.e., AAL robots). Moreover, the employment of the latter recorded a drastic increase (>20%) in the last two years due to the pandemic emergency.

In this context, this Ph.D. thesis presents a novel robot-empowered AAL infrastructure capable of creating a pervasive environment, which is fully-aware of the psychophysical status of the patient and that adapts to different degrees of disability and emotional involvement of the patient. Specifically, the features introduced by this thesis in the context of an AAL infrastructure are: (i) the introduction of a Brain-Computer Interface (BCI) for the selection of services, able to allow patients with severe disabilities to formalize requests; (ii) the incorporation of a cognitive status recognition system based on electroencephalographic (EEG) signals able to provide monitoring parameters and support for remote diagnosis; (iii) the inclusion in the infrastructure of an emotion recognition system capable of triggering an adaptive behavioral system by the robot component of the architecture.

Concerning the BCI-based selection system, two neural interfaces are proposed. A BCI exploits movement-related cortical potentials (MRP), while the other one is based on event-related potentials (ERP) and in particular on an ERP component: the P300. The former BCI is a first-of-a-kind MRP-based neural interface that allows a quick selection of services by using nested binary choices and a low computational complexity processing algorithm. Specifically, the BCI is based on a symbolization method capable of translating the EEG signal into a binary string, speeding up the subsequent inferential system based on a Support Vector Machine (SVM).

Differently, the second BCI (i.e., the P300-based one) permits a greater number of choices in a single session (i.e., 12) and employs a Dense Neural Network (NN) as an inferential system. Specifically, the thesis proposes a framework for the selection of the appropriate "user-tailored" NN topology based on a dedicated tuning of the hyperparameters.

The cognitive status extraction system analyzes in parallel the EEG signals provided by the P300-based BCI. In fact, this system exploits the biomarking properties of the P300 component for the identification of cognitive impairment. For this purpose, an algorithm

for the spatio-temporal reconstruction of ERPs, named t-RIDE, is introduced. It provides a quantitative parameter useful for the remote support to diagnosis or monitoring of the patient. The EEG-based emotional state recognition system employs a user-specific selection framework. The proposed selection method analyzes, through the cascade of two grid-search-based routines, a sequence of feature extraction algorithms used in the emotion recognition field, assessing their accuracy-versus-complexity ratio. The implemented system is designed to discriminate up to 8 different emotions by using a circumplex model based on three parameters for the emotion definition: arousal, valence, and dominance.

BCIs and cognitive/emotion recognition systems are trained/calibrated offline on a dedicated computer, while the online inference is realized through three different Android apps running in background on Pepper tablet.

Since the services offered by BCIs integrated into the AAL infrastructure also include picking and delivery of various goods, as part of the thesis work the Pepper robot has been equipped with object handling capabilities, which is natively missing.

In this context, a set of routines for object manipulation based on the built-in RGB cameras and the 3D sensor has been realized. The implemented routines exploit the Pepper sensing systems to (i) identify the object to be grabbed, (ii) plan the arms movement sequence, and (iii) grab the object for final recognition. The routines, programmed via Python to run in the background on the robot's operating system, are designed to operate fully automatically, without the need for an internet connection (ensuring intrinsic protection for sensitive data).

To test innovative features introduced by this thesis, in-vivo measurement sessions have been carried out on 13 volunteers from the Politecnico di Bari and g.tec medical engineering GmbH (Austria) company. Overall, *MRP-based BCI* demonstrated to be able to discriminate two choices with an average accuracy of 84.07 % with a mean Information Translate Rate (ITR) of 11.17 commands/minute. However, this BCI can achieve a mean ITR of 22.33 comm./minute (i.e., 1 out of 2^{22} available choices/ min) by simply reducing the interstimulus interval from 2 s to 1 s. In the same context, the P300-based BCI demonstrated to be able in achieving a choice recognition accuracy of 96.66 % after 21s of stimulation. Considering 12 possible choices, the *P300-based BCI* showed a maximum ITR of 16 comm./min. The *emotion recognition system* outcomes show that in an 8-emotion discrimination problem it can achieve a multiclass accuracy of ~76 % (mean value), or considering a model that involves only 4 emotions, ~ 80 %.

Experimental tests on the *cognitive status extraction system* demonstrated that all the subjects involved in the tests, belonged to the healthy subjects class, having the ERP reconstructed characteristics compatible with this class.

Finally, concerning the object manipulation capabilities, the proposed routine tested in real-life scenarios, returned a grabbing accuracy of ~87% for different shelf heights, demonstrating the employability of improved social robotics for daily-life assistance and ambulatorial contexts.

Introduction

Globally, the population over 65 is increasing dramatically in the face of a reduction in births. This analysis of aging and health emerged in the World Health Organization (WHO) report and sees an increase in the 65+ population to 22% of the global population by 2050. In this context, many countries, including Italy, already present a population of 65+ above 25%.

These numbers appear critical considering that with the increase in life expectancy, chronic diseases also increase at the same time, causing physical and social disabilities to the elderly.

The onset of these disabling diseases leads to the need for continuous assistance for the population involved. However, a global *caretaking shortage* is occurring, because the assistance typically provided by the younger population or by caretaking institutions cannot meet the demands or can do so at high costs.

The reduction of home assistance by competent personnel leads to a consequent increase in hospitalizations with long stays, weighing on the health systems of nations.

To aggravate the situation, the latest health emergencies have led to further social and assistance isolation of the fragile population due to the massive redeployment of personnel specialized in emergency management.

A possible solution to this caregiver shortage comes from the Ambient Assisted Living (AAL) Joint Programme, which proposes to empower seniors' capabilities and their surroundings with technological solutions, aiming to enhance their quality of life and allow aging actively in their house.

The first line of enhancement proposed in the AAL framework consists of intelligent systems for the monitoring of the person and the environment, by using diffused indoor smart sensors. The second and more recent line is the further inclusion of a robotic agent within the AAL framework, introducing the AAL robot class. This choice has been corroborated by the emergency needs that have emphasized and pushed more to the use of robots both in care facilities and domestic assistance (+20% of the use of companion robots from 2020 to 2022).

In this context, several new robots have been proposed to improve hospital and domestic logistics. Some examples are the automatic wagons to move sterile instruments into the hospital area, or the automatic and secured cancer medicine transportation wagon both proposed by Mobile Industrial Robots. ABB proposed an automatic mobile robot to support the medical staff in repetitive laboratory tasks, e.g., loading/unloading centrifuges. During the last pandemic emergency, robots were employed for environment disinfection.

However, it should be specified that the use of social and assistive robots, although furtherly matured in recent years, had led - in the years preceding the health crisis - to the inclusion in the hospital and home environment of a large number of companion robots. The reason lies in the ability of these robots to fluidly interact with the human component

of the infrastructure. These capabilities allowed companion robots to be used as a receptionist in care facilities, or in more useful cases from an assistive point of view to their use to fight depression, loneliness, and, in some cases, dementia and related cognitive impairments.
In this global framework, the AMICO (Medical Assistance In COntextual Awareness) project was born. This project aims to create an intelligent infrastructure that integrates a home assistance and rehabilitation system including sensors distributed in the environment and a robotic agent. The innovativeness of this project stems from the desire to investigate and address AAL robotics aspects that, to date, are still challenging.
These include the creation of a pervasive environment that has the full psychophysical awareness of the infrastructure user and that demonstrates adaptability to different degrees of disability and emotional involvement of the patient.
The aforementioned project has chosen as a system robotic platform one of the most diffused autonomous robots in healthcare facilities and domestic assistance, i.e., Pepper by SoftBank Robotics. It counts more than 12000 employed units only in Europe due to the tradeoff between costs and provided services. The reason for this choice lies in the possibility of massive reuse of the proposed architecture in environments already equipped with this robotic platform for purely social involvement purposes.
During this thesis work, several new features have been introduced by Politecnico di Bari (and specifically by Design of Electronic Integrated System Lab) in the AMICO infrastructure.
First, the robot has been equipped with an additional user interface based on the *Brain-Computer Interface* (BCI). This interface opens the possibility of formalizing requests regardless of the severity of the user's disability, replacing mechanical or voice commands when needed.
Second, since the BCI is based on the analysis of brain signals for service selection, they have been re-used to provide a contemporary *assessment of the cognitive status* of the patient. This feature adds a step to the current state of the art of AAL robotics, which involves the administration of games aimed at maintaining or improving cognitive status without verifying the effects of the cognitive stimulation plans.
Third, the use of cortical signals, directly connected to the autonomic nervous system, has led to the possibility of using them for the extraction of the patient's emotional state. The *emotional state recognition* system has been used in the AMICO infrastructure to create an adaptive interaction system that can modify conversational tasks according to the emotion felt by the user.
Moreover, to recreate an infrastructure capable of operating standalone in the assistive context, the companion robot Pepper has been functionally reprogrammed to implement *non-native object manipulation* system that is able to operate without any internet connection to preserve sensitive data. Indeed, although it is equipped with two upper arms, it lacks object manipulation capabilities, because arms are only used for animated verbal interactions.

Finally, the robot has been programmed to check the correct administration of the drug, operating the first phase of *pharmacological compliance*.
The thesis is organized as follows.

Chapter 1 will analyze the context of robotics-empowered AAL, providing details about the current classification of AAL robotics and application fields. A section of the chapter will be devoted to a systematic review of the services provided by AAL robots. Finally, a summary of the features introduced by the present thesis in the AMICO AAL infrastructure will be provided.

Chapter 2 will describe, from an architectural point of view, the overall AMICO AAL framework, focusing on the contributions introduced by the thesis. The chapter will provide nominal use cases of the system, aiming to provide the readers with possible applications. Next, it will outline the main features of the architecture performers: user, robot, and environment.

Chapter 3 will be dedicated to the methodologic analysis of the employed BCI system. In this context, the chapter will provide a brief overview of the nature of the cortical signals involved in the application and the experimental and stimulation setups adopted in the framework implementation. Next, the chapter will analyze the pre-processing and processing stages of cortical signals, oriented to the implementation of the selection BCIs. This chapter will also define the algorithms behind the cognitive status extraction system and the emotion recognition system, both based on brain signals.

Chapter 4 will summarize the methods of the Pepper robot necessary for the realization of an automaton able to move autonomously in the environment and manipulate objects for pick-up and delivery procedures. Specifically, the thesis will detail an innovative object manipulation system, which operates offline.

Chapter 5 will provide experimental results and discussions by dividing them by macro areas, i.e., brain-robot interface and manipulation routines. In the first case, each BCI or emotional/cognitive recognition system will be analyzed in terms of standard performance metrics. In the second case, the results deriving from the optimization tests of the manipulation procedures will be provided.

Chapter 6 concludes the thesis by providing the breakthrough points introduced by the proposed features and analyzing related future perspectives.

Robotics in Ambient Assisted Living

The global increase of the elderly population (65+) percentage stressed the need for the introduction of new paradigms concerning sustainability of care and assistance. Besides the lack of caregivers, since people lives longer, the costs of care increase due to the high number of chronic diseases and comorbidity. In this framework, Ambient (*Active*) Assisted Living (AAL) field fits. This chapter will outline the global aging problem and how AAL technologies can address the related healthcare requests. Next, the introduction of a new agent in the 3rd generation AAL frameworks, i.e., robotic platforms, will be detailed providing their classification and their field of applications. Basing on this last classification, a detailed analysis of AAL robotics state of the art will be provided within the chapter. Finally, a section will be dedicated to AMICO project, an AAL infrastructure that prompted the development of innovative features that are the objectives of this thesis work.

1.1 Ambient Assisted Living (AAL)

According to the World Health Organization (WHO) report on global aging [1], the proportion of the world's population over 60 years will reach 22% by 2050 starting from the current 12%. Moreover, the Organisation for Economic Co-operation and Development (OECD) report [2] shows that, to date, the highest percentages of the elderly population (involving 65+ people) in 2022 are recorded in Japan (28.86% of the population), Italy (23.67%) and Germany (22.09%). While in Japan (54%), Germany (40.50%), and Italy (40.20%), the OECD collected the highest old-age dependency ratio, defined as the number of individuals aged 65+ per 100 people in working age (i.e., 20 to 64). Fig. 1.1 shows the above-described statistics, emphasizing the global aging numbers. A major challenge that must be addressed due to the longer life spans of humans is the problem of caring for elderly people because the activity and perception of these people are not keeping pace with the expected life span. As an increasing percentage of the population will be considered elderly in the near future, according to OECD data, there will be a caretaking gap since not enough young people will be available to sustain the elderly population [3].

Most importantly, when the population is becoming older, the chance for emerging chronic diseases is getting higher, further decreasing the mobility and flexibility of elderly people. According to the review in [4], 65+ people mainly suffer from total/partial mobility impairment (44% of women and 28% of men experience problem in walking, and moving), 20% has disabilities due to diabetes, 13% suffers from depression symptoms, 10% of women and 13% of men are forced to have hospital rehabilitation due to cardiovascular disease. People between 75 and 80 achieve the 3% of diagnosed dementia, with this number that increases up to 24% for 80+. This results in the need for long-term continuous care which cannot be provided by the younger population or the classical caretaking institutions due to high costs and the lack of experienced caretaking personnel [3].

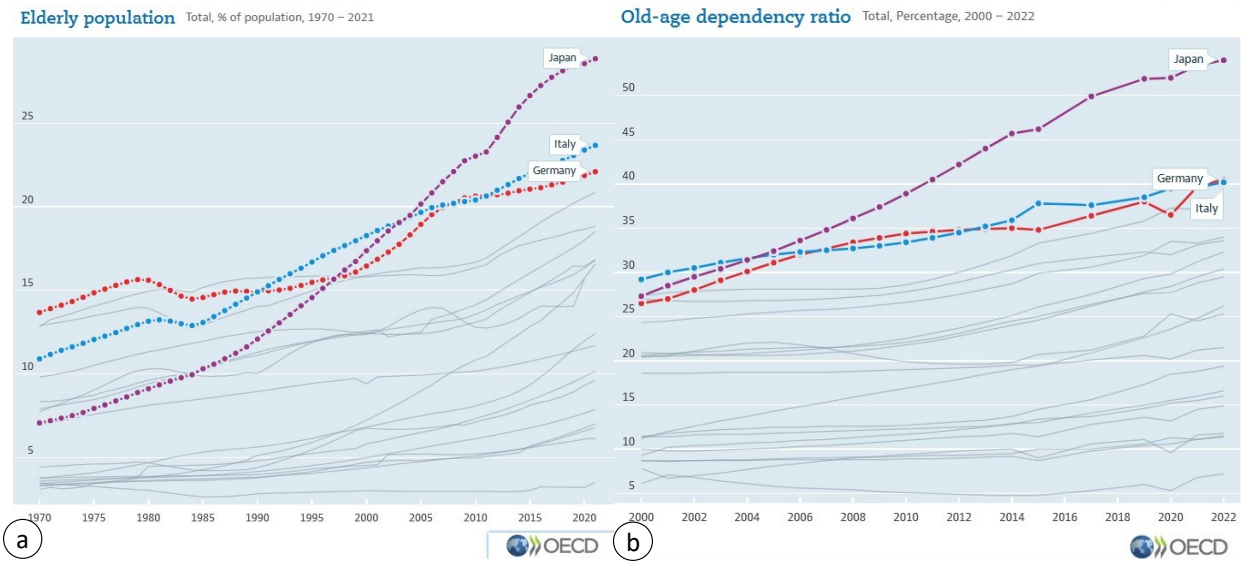

Figure 1.1 OECD Statistics. (a) Elderly population (% of population) available a; (b) Old-age dependency ratio (%)available at https://data.oecd.org/chart/6Q2W. Background gray tracks are the same parameters related to G20 countries.

A possible solution to this caregiver shortage is to empower seniors' capabilities and their surroundings with technological solutions, aiming to enhance their quality of life and allow aging actively in their houses.

In this vision, seniors' assistance is delegated to the intelligence embedded in the technology that should not represent a further obstacle for the user [5].

These technological solutions concur to define the so-called Ambient Assisted Living (AAL). AAL is a multidisciplinary scientific field involving the creation of high-level software and hardware systems to increase occupational health, quality of life, and well-being for an enhanced living environment. AAL generally employs information and communication technologies (ICT), stand-alone assistive devices, and smart home technologies in a person's daily living and working environment to enable individuals in staying active longer, remain socially connected and live independently into old age [6, 7].

AAL foresees the employment of pervasive home environments that embed sensors, actuators, smart interfaces, and artificial intelligence [6]. AAL gathers, under this umbrella term, traditional assistive technologies for people with disabilities, universal design approaches to favor usability, accessibility, and acceptability of interactive technologies, and the emerging ambient intelligence computing paradigm, which provides intelligent, unobtrusive, and ubiquitous assistance [8].

Concerning the AAL framework definition, in an attempt to provide a quantitative classification of the term, the authors in [9] stated that to date there are three generations of technology designed for supporting the independent living of older adults.

Within the first generation of AAL technologies, it is possible to identify the first wearable alarms (e.g., pendants or wearable alarm buttons). They were intended to allow older adults to press the button or pendant in order to raise the alarm in the case of an emergency situation, such as a fall [10].

Although this 1st generation alarms demonstrated several benefits related to security and safety, such as reduced stress among older adults, families, and caregivers, earlier hospital discharge and delayed entry into long-term care facilities [10], they had some important weaknesses. A person unable to move, as well as a person who forgot to wear the device, cannot trigger the alarm. This latter is a quite diffused scenario that an older person takes off the alarm when he/she goes to bed and forgets to put it back on when needing to use the bathroom in the middle of the night.

The second generation of AAL technologies integrates electronic components to address the limitations of the first generation. These technologies can detect potential emergencies, such as a fall or environmental hazards, by using sensors [11]. Seniors are monitored using sensors spread within the home, which can autonomously call for assistance without the need for the user to trigger the alarm [12]. This technologic category provides several benefits, however, some users feel the presence of unanimated spread sensor networks as intrusive and raised concerns about privacy problems [13]. These technologies are getting a big slice of the marketplace and are, currently, used by older adults within the home.

The most recent trend of AAL travels alongside the advances in ICT. To date, this new technology not only detects and reports problems, but also prevents problems using the main concept of AAL: make AAL services sensitive, responsive, and acceptably integrated into our daily environment to create an intelligent ambient.

These systems integrate computing systems and assistive devices into everyday living contexts for not only monitoring the home environment but also the older person. Environmental and wearable sensors monitor vital signs as well as changes in mobility and activity patterns. These may be indicative of changes in health status.

The surrounding is also supplied with actuators, which are mechanical devices used for controlling a mechanism or system that provide the older person with assistance, and smart interfaces to provide information, support, and encouragement.

Presently, this third generation is in development and several new intelligent systems and remote services are being increasingly adopted in this domain to improve quality of life, and support self-independence [14, 15].

In this framework, a new trend is taking hold aiming to address the caregivers' shortage and to support caretaking personnel work, reducing their effort when most of the seniors' independence is gone. This trend consists of embedding *acceptable* robotic platforms in the smart home environment [5, 14, 16, 17], which are able to: (i) improve the physical capability of the target user, (ii) assist the user in personal care, (iii) socially interact with patients; (iv) assist both primary user and caregiver with repetitive tasks such as drug delivery and (v) continuously monitor the user well-being.

1.2 Robotics-empowered AAL

1.2.1 Robot Platforms Classification

This section provides a compendium concerning the definition of robot categories belonging to the field of AAL.

According to [18, 19], a robot that assists people with physical disabilities through physical interaction is defined *Assistive robot* (AR), while those robots that support people in a nonphysical way through non-contact interaction are named *socially assistive robots* (SAR) [18, 19].

The standard ISO 13482:2014 provides a *personal care robot* (PCR) definition and classification [20, 21]. According to ISO13482, generally speaking, PCRs are robots that "typically perform tasks to improve the quality of life of intended users, irrespective of age or capability." [20].

The same standard subdivides these kinds of robots into three categories: (i) *Mobile servant robot*, (ii) *Physical assistant robot,* and (iii) *Person carrier robot.*

The first kind of robot is related to PCRs capable of traveling to perform serving tasks in interaction with humans, such as handling objects or exchanging information. Physical assistant robots are PCRs that physically assist a user to perform required tasks. They supplement and/or increase the personal capabilities of the primary user.

Finally, person carrier robots have the purpose of transporting humans to an intended destination (except for robots that travel faster than 20km/h).

For this purpose, the International Federation of Robotics coined the term *service robot*, defining it as a robot which operates semi- or fully autonomously to perform services useful to the well-being of humans and equipment, excluding manufacturing operations.

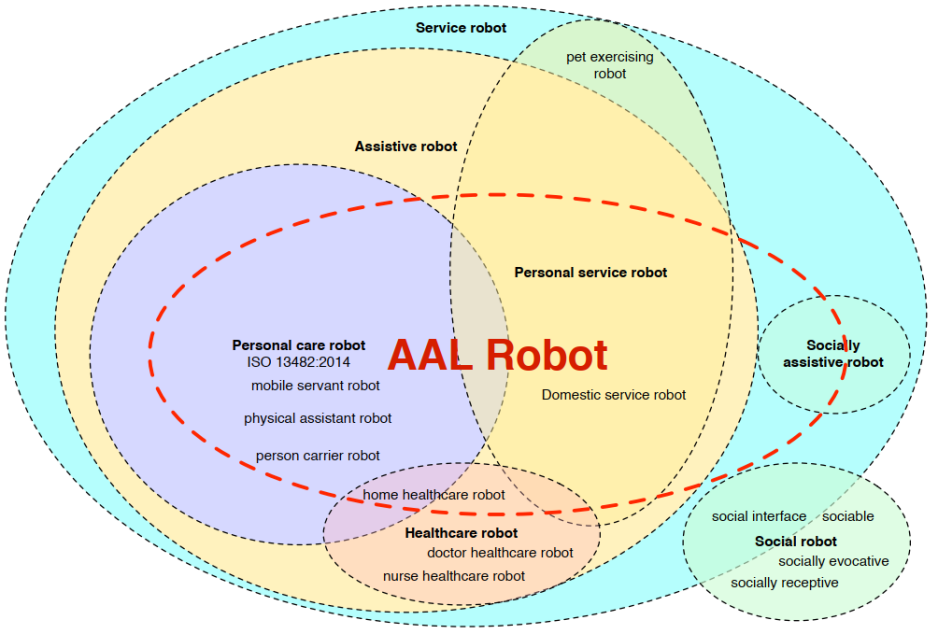

Figure 1.2 Subset of AAL robots (red dashed set) in the definition space for robotics.

To limit the inclusion batch the term *personal service robot (*or *service robot for personal use)* has been introduced. This latter concerns robots used for non-commercial tasks, usually by lay persons. Examples are domestic servant robots, personal mobility assisting robots, automated wheelchairs, and so on.
In general, AAL robots do not exactly fit into the pre-defined categories of *assistive robots* or *socially assistive robots* since they fall within the characteristics of both categories, but they limit their application to older people and people with disabilities in their living environments. Fig. 1.2 shows the overlap between the various classifications provided by [20, 22] and the subset of AAL robotics, which includes with its boundaries several kinds of robots.
Ultimately, referring to the European AAL Joint Programme outcomes [23] that reformulate AAL passing from Ambient Assisted Living to Active Assisted Living due to the robot presence, it is possible to state that an AAL robot is a robot that (i) assists the target group of older users including users with disabilities; (ii) supports the target group during daily life or work, (iii) improves or maintains the independent living of the target group, (iv) reduce the effort of caregivers (secondary user) when their employment is unavoidable.

1.2.2 Application Fields of AAL robots

AAL robotics definition gathers several need-oriented tasks that are currently addressed by different robotic platforms. These tasks range from supporting the elders in activities of daily living such as feeding, dressing, bathing, ambulating or in activities that involve other technological devices, e.g., preparing food or doing laundry and housekeeping, up to socially involving them with activities, learning new skills and participate in hobbies.
A classification based on the group of tasks already available on commercial and research-grade robots divides the AAL robots into 8 main classes: (i) robotic aids for mobility, (ii) robotics aids for manipulation, (iii) companion robots, (iv) specialized personal care robots, (v) household robots, (vi) fetch and carry the platform, (vii) emotional robots, (viii) telepresence robots.

Robotic Aids for Mobility. This kind of AAL robot is intended to augment the mobility capabilities of the primary users (elders) supporting their movements and navigation. Some examples are robotic wheelchairs, exoskeletons, and robotic walking frames. An example is the FRIEND II robot (Fig. 1.3.a) proposed by authors in [24]. It presents an 'intelligent' wheelchair with mounted manipulators that is possible to control by voice and mechanical commands (e.g., joystick). Under the activity of daily living support aspect, these robots allow elders to fulfill all those activities that require lower limb mobility (it is assumed full control of upper limbs). They permit also the improvement of the social sphere of the primary users.

Robotic Aids for Manipulation. Manipulators included in AAL robotics definitions typically refer to wearable robots for upper limbs. They are intended to support activities that require dexterity and strength for the hand and arm. Most of these

manipulators are embedded in other solutions or personal care robots. An example of an AAL manipulator (Fig. 1.3.b) is proposed by authors in [25,26], who propose a portable 6 DoF gripper, born to support patients with spinal cord injury, which properly installed near the specific target can permit to drink, brush teeth and supports elders in eating. The drawback of this kind of robot is the low degree of re-programmability and adaption to different real-life scenarios without any technical intervention. It limited the release of new devices in this field.
Under the activity of daily living support aspect, these robots have the potential of fulfilling all those activities that require upper limbs mobility, as well as completing some simple work or hobbies.

Companion Robots. Companion robots are platforms mainly designed to socially interact with people. They can accomplish several kinds of tasks such as, e.g., health monitoring, providing cognitive and psychological support, disseminating information, triggering alarms or warnings, socially involving the elder, opening a remote communication channel with family and so on. This kind of robot recorded a huge increase in the usage rate during COVID-19 pandemic, because of the intrinsic resistance to the virus and the possibility of continuous IR sterilization [27]. In this context, companion robots provided, *inter-alia*, a virtual bridge between patients in hospital or care facilities and their families [28].
One of the most diffused companion robots is Pepper (Fig. 1.3.c) by SoftBank Robotics [29]. It is an anthropomorphic robot categorized as a *social robot* because mainly designed to interact with people. It embeds a rich dialog engine supported by animation and an intrinsic (but poor) emotion recognition framework. The presence of a re-programmable tablet on its chest increases the number of applications of this robot. Considering the activity of daily living support aspect, these robots can provide a series of important services such as health monitoring, drug administration reminders, learning support, physical training suggestions and motivation, entertainment, loneliness fighting,

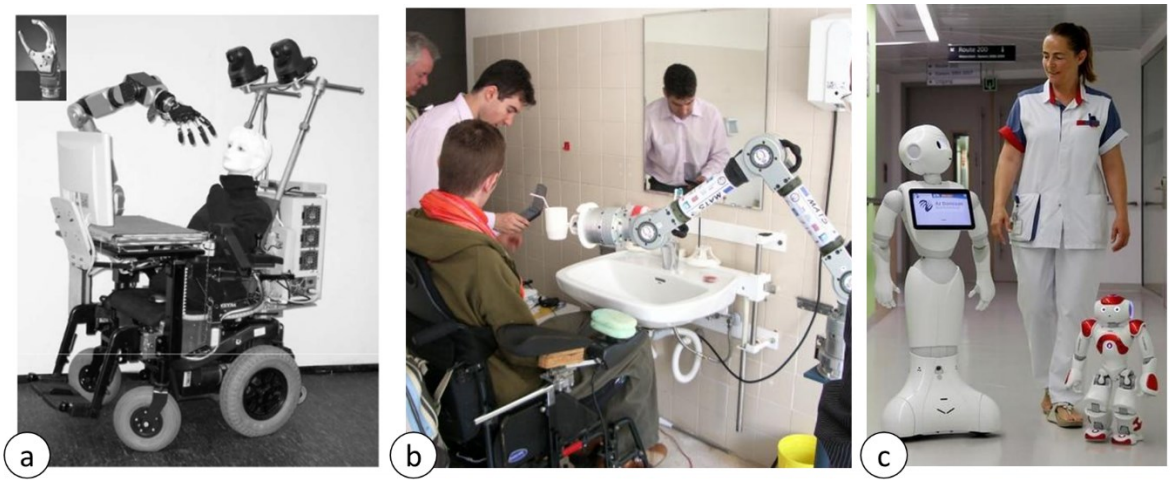

Figure 1.3 AAL robots classification for application field. (a) Robotic Aids for Mobility – FRIEND II rehabilitation robot by [24]; (b) Robotic Aids for Manipulation - ASIBOT by [25]; (c) Companion Robots – Pepper and Nao by SoftBank Robotics [29]

and so on. The drawback of social robotic companions is that their use is often limited to non-physical support because the almost totality of the commercially available companion robots do not embed these functionalities (manipulation or care support).

Specialized Personal Care Robots. This category refers to the robots designed to implement specific care tasks like support in bathing, brushing teeth, eating, and drinking.
An example is Bestic [30] (Fig 1.4.a), which is a robotic arm supplied with a spoon to support food intake. Another example is I-Support [31], a research-grade platform able to exploit innovative soft robotics arms and a set of stereo cameras to move and irrorate sponges with soap, supporting bathing operations.
These robots are highly specialized and, typically, final user-tailored, in terms of programming, at the installation time. Concerning the activity of daily living support aspect, PCRs can address specific problems, *singularly* covering all the needs of the elders. The drawback of PCRs lies in the number of robotic platforms that must be employed to cover even similar needs, e.g., eating and drinking.

Household Robots. This is a widely consolidated category of robots that is not specifically designed for older adults, but that fit the AAL robots' definition. Under this category, it is possible to find commercial robotic vacuums and floor cleaners. To date, the acceptance rate of this kind of AAL robots is high due to the provided benefits, making them widely diffused in our homes and cost-effective platforms.

Fetch and Carry Platforms. These robots are intended to support elder in carrying heavy payloads or bringing goods. Some examples of these platforms are robotic trolleys, which are devices able to autonomously follow the owner by transporting heavy contents. Another example is a mobile tray robot proposed in [32] (Fig. 1.4.b), which is supplied with a tablet as human interface, a tray, and a wheeled base to transport goods within a specific facility. Currently, this robot has been used in a hotel to bring goods to specific rooms. Anyway, its capabilities can be adapted for employment in care facilities or domestic environments.

Emotional Robots. This category gathers those "toy" robots designed to support caregivers in social work with elderly people. These robots have, usually, pet shapes (e.g., cats, dogs, etc.), because of the positive influence introduced by the interaction between elders and pet-like robots. Indeed, it has been studied that the benefits are similar (although to a lesser extent) to those introduced by real pet therapy [33]. An example of an emotional robot is the well-known Paro [34]. Paro is a seal-like robot able to react with flippers and sounds to different user actions. This kind of objective has been recently embedded in companion robots. Nevertheless, emotional robots are still finding large applications in people with dementia therapy.
Concerning the activity of daily living support aspect, emotional robots can only provide company and entertainment.

Telepresence Robots. These robots offer the possibility of creating a virtual bridge between patients at the bedside and care staff that can visit them remotely. Vita® [35] is an example of a telepresence robot that lets doctors interact with patients remotely. It provides two-way video and access to patient data and diagnostic devices such as otoscopes and ultrasound imagers. Also, in this case, it has been studied the possibility of supplying companion robots with specific diagnostic devices, borrowing the telepresence capabilities when the companion robot has a two-way video framework available. In terms of supporting activities of daily living, telepresence robots can only be used for health monitoring and administration warnings management.

Figure 1.4 AAL robots classification for application field. (a) Specialized Personal Care Robots – Bestic by [30]; (b) Fetch and Carry Platforms – A.L.O. Botlr by [32]; (c) Emotional Robots – Paro by [34]

1.2.3 AAL Robotics Investments

This section provides a quick overview of the European investments in AAL robotics, to chart the course of the current trend in the specific sector.

Limiting the analysis to projects funded by the EU Commission in the frame of the Horizon 2020 (H2020) program, it resulted in those 8 projects concerning AAL robotics receiving economical support (according to CORDIS database) for a total of € 31.000.000. If the analysis is extended to the FP7, the number of projects grows up to 35 projects. Four out of 9 projects from H2020 framework (44.4%) concern Companion Robotics application field. Specifically, RADIO project [36] proposed a study concerning the acceptance and unobtrusiveness criterion that allows robots to be accepted part of the user's daily life. ENRICHME [37] created a platform for AAL, in which the companion robot is used to keep the older adult active, by suggesting physical/mental activities. GrowMeUp [38] created a service robotic system that with a series of warnings and suggestions encourages and engage older persons to stay longer active, independent, and socially involved, in carrying out their daily life at home. Finally, MARIO [39] addresses the difficult challenges of loneliness, isolation, and dementia in elders through a well-structured dialog engine.

One out of 9 (11.11%) is dedicated to robotic aids for mobility. It is the ACANTO project [40] that proposes a smart robot walker that revisits the notion of robotic walking

assistants and evolves it into an activity vehicle. It executes a program of personal training embedding exercises in every-day activities.
In the context of fetch and carry robots it is possible to find the project RAMCIP [41], which proposes a servant robot aware of the users' activities profile in the various spaces composing the house, capable of monitoring the home environment, to optimally decide when and how to assist. The project is optimized for houses with people affected by Alzheimer's Disease or mild cognitive impairment.
The previously introduced I-SUPPORT [42] constitutes the only project under the PCRs umbrella.
Finally, the eNHANCE projects [43], which fit the robotic aids for manipulation, have been funded. It aims to assist the user in performing their daily-life interaction with the environment through an intelligent adaptive interface controlled by an intention detection interface and a personalized behavioral model. The eNHANCE active support orthotics enables the users to achieve their desired movement actions while motivating the users to maximize their own force contribution. This will maximize user performance relative to their personal capacity, and so maximize therapeutic effects.

1.3 Companion Robots for AAL: Related Works

According to Sec. 1.2.3 appears clear that, currently, most of the research and commercial interests concern the improvement of AAL frameworks that include companion robots. It is an interesting (although not unexpected) result because among the benefits of companion robots it is possible to find good versatility (there are a lot of non-physical support applications), as well as capabilities in adapting to users' behavior (e.g., by profiling daily activities). These capabilities make companion robots good interlocutors to fight loneliness [39], terminals to increase mental activities to address cognitive impairments [39], motivators for physical and psychological activities [37, 38], and mobile reminders (e.g., for drug administration, warnings, agenda).
Focusing on this field of application, the advances in sensors, wireless networks, the Internet of Things (IoT) and robotics led to the developing of innovative technologies for sustainable and high-quality home care for the elderly. Extensive research has been carried out in mobile autonomous companion robots operating in domestic environments. In this framework, the CompanionAble project [44] (Fig. 1.5.a) tried to address the issues of social inclusion and homecare of the older population. In this project, a custom wheeled totem with a height of 120 cm has been realized. The totem provides a 15.4 inches touch screen display for the presentation of information to the user and the input of user data. A series of entertainment can be selected (e.g., video, music) and provided via a display interface. The robot presents a tray where a user can deposit items useful for daily usages, like glasses, wallets, keys, or a mobile phone. The totem can navigate autonomously on specific paths reaching specific meeting zone.
Another remarkable solution is the one proposed in ENRICHME project [45] (Fig. 1.5.b). They proposed a system that consists of three main elements: the mobile robot platform, the Ambient Intelligence System (AIS), and the Networked Care Platform (NCP). The

first two elements are deployed in the elderly home, while the latter is accessible via the internet. The robot platform is realized via a custom version of TIAGO robot by PAL Robotics (no upper arms version). It acts as a companion and a physical system interface for elderly users. The AIS is a dedicated computer interfaced with the robot, the different smart home sensors, and the NCP. NCP includes a visualization tool accessible by the medical staff to monitor the data made available by the robot and the smart home environment. The robot is responsible for domestic monitoring, object finding (through RFID sticked on important objects), personalized reminders and administration of periodic cognitive games.

Another project in the field is MOBISERV [46] (Fig. 1.5.c). The system is designed to monitor the user's physical activity and health indicators via wearable hi-tech fabrics, and analyze your nutrition habits through smart home sensors. Starting from this data, MOBISERV provides older adults with: (i) nutrition assistance and dehydration prevention by eating and drinking reminders and encouragements; (ii) a personal health coach encouraging physical activity and specific exercises, and supporting telemedicine services; (iii) well-being services for cognitive stimulation and social inclusion, responding to the user's emotions; (iv) games for entertainment (v) camera-based fall detection with direct communication to a care center, and (vi) video communication service to allow connections with friends and family.

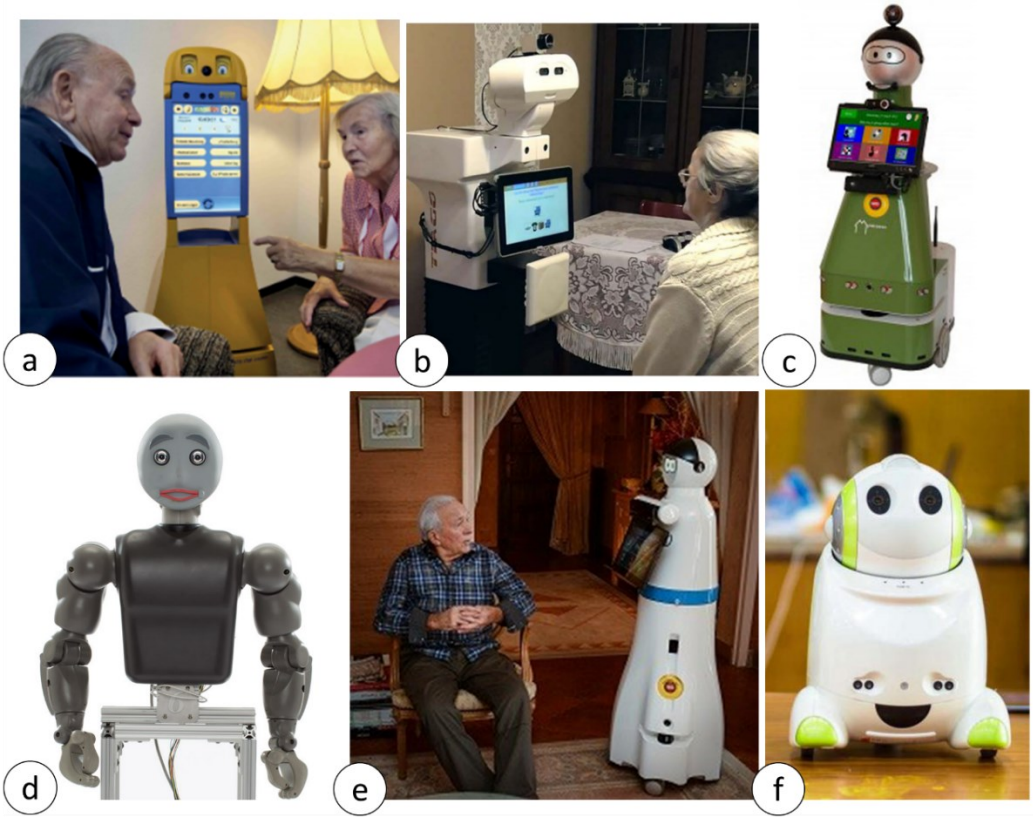

Figure 1.5 Companion robots for AAL – State of the Art. (a) CompanionAble project; (b) ENRICHME project; (c) MOBISERV project; (d) Fasola et al. [47]; (e) MARIO project; (f) Matilda project

Using the robotic platform Bandit, authors in [47] (Fig. 1.5.d) presented a companion robot that operates as a personal trainer. It has been designed to motivate and engage elderly users in simple physical exercises. Indeed, the study demonstrated that older adults express a clear preference for using a physically embodied robot coach, rather than a virtual one, in terms of enjoyableness, helpfulness, and social attraction.
MARIO project [39] (Fig. 1.5.e) exploited a Kompai platform and commercial footprint of ROBOSOFT, to realize a companion robot empowered with a semantic data analytics engine for personal interaction, and unique applications tailored to better connect older persons to their care providers, community, own social circle.
Each objective is developed with a focus on loneliness, isolation, and dementia. MARIO can call (or videocall) friends, family, and doctors, provide news from the world, open e-books, reproduce movies and music, and provide to-do lists and reminders. The companion robot also collects data useful for clinical characterization, by means of a touch screen interface.
Another interesting research is proposed by authors in [48], which realized a human-like assistive communication robot, Matilda (Fig. 1.5.f). This robot is oriented to improve the emotional well-being of older people by analyzing their emotions and moods.
Tables 1.1 and 1.2 summarize the above-presented projects emphasizing their main characteristics. In order to provide a general (*non-exhaustive*) overview of the state-of-the-art, tables report, per each analyzed study:

- **Navigation.** It refers to fully autonomous navigation capabilities.
- **Object Manipulation and Handling.** The parameter concerns the ability of the robot to manipulate objects via grippers/hand.
- **Health Monitoring.** It refers to the ability of the robotic system to monitor vital signs from patients (cortical, cardiac activity, blood pressure, etc.). Also, self-diaries and behavioral analysis are considered admissible for the purpose.
- **House Control.** The capability of the companion in interfacing smart home sensors and actuators (smart light, plugs, etc.).
- **Cognitive Involvement.** Functionality that allows the robot system an active involvement of the user to verify/stretch / improve cognitive abilities (via games, tests, etc.).
- **Cognitive Assessment.** The autonomous or semi-autonomous capability to record the cognitive response to assess progress or to evaluate impairments.
- **Physical Activities Involvement.** The robot can suggest, and propose new physical activities.
- **Physical Activities Assessment.** The robot is able to check the proper physical exercise fulfillment.
- **Entertainment.** The companion is supplied with a video terminal (e.g., tablet or display) for videos and movies reproduction and a loudspeaker for music playing.

- **Social Sphere Improvement.** The automaton can create a virtual bridge between users and familiars, friends, and doctors. The virtual bridge can be a call and/or video call system.
- **Programmable reminders.** The parameter refers to the robot's ability in storing and synchronizing events basing on a specific scheduling.
- **Drugs administration reminders.** It is a specific functionality of the previous point of the bullet list. It refers to a particular scheduling concerning drugs administration with related information.
- **Pharmacological compliance.** The robot embeds an approach to assess if the subject has taken the medicine, or it is able to check the reduction of symptoms related to the taking the drug.
- **Emotion recognition.** The system can recognize the user's emotions (or moods) to adjust its behavior adaptively.
- **Attention to severe disabilities (no movements, dysarthria).** The proposed solution can be applied to a catchment area that also involves older persons with severe disabilities. For instance, providing a specific input channel for selection different from the touch screen or lighten up the secondary users' (caregivers) effort in assisting elders.

Table 1.1 – Companion Robots State-of-the-Art: Part I

Project	**ENRICHME [37]**	**GrowMeUp [38]**	**MARIO [39]**
Navigation	✓	✓	✓
Object Manipulation and Handling	✗	✗	✗
Health Monitoring	✓ Behavioral Analysis	✗	✗
House Control	✓ Smart Sensors and Actuators Interface	✗	✗
Cognitive Involvement	✓ Periodic cognitive games	✓ Activities Proposals	✓ Cognitive games
Cognitive Assessment	✓ remotely via NCP	✗	✗
Physical Activities Involvement	✓	✗	✗
Physical Activities Assessment	✓	✗	✗
Entertainment	✓ Movies, Music	✓ Movies, Music	✓ Movies, Music, Books
Social Sphere Improvement	✓ Video-call, Call	✓ Video-call, Call	✓ Video-call, Call
Programmable reminders	✓	✓	✓
Drugs administration reminders	✓	✓	✓
Pharmacological compliance	✗	✗	✗
Emotion recognition	✗	✗	✗
Attention to severe disabilities (no movements, dysarthria)	✗	✗	✗

Table 1.2 – Companion Robots State-of-the-Art: Part II

Project	**MOBISERV [46]**	**CompanionAble [44]**	**Matilda [48]**	**Fasola [47]**
Navigation	✓	✓	✓ Desk navigation	✓
Object Manipulation and Handling	✗	✗	✗	✓ not employed
Health Monitoring	✓ Wearable fabrics	✗	✗	✗
House Control	✓ Smart Sensors and Actuators Interface	✗	✗	✗
Cognitive Involvement	✓ Cognitive games	✗	✗	✗
Cognitive Assessment	✗	✗	✗	✗
Physical Activities Involvement	✓ Activity proposal, Motivation	✗	✗	✓ Activity proposal, Motivation
Physical Activities Assessment	✗	✗	✗	✓ Exercise Check
Entertainment	✓ Movies, Music, Games	✓ Movies, Music	✓ Active Interaction	✓ Movies, Music
Social Sphere Improvement	✓ Video-call, Call	✓ Video-call, Call	✓ Video-call, Call	✗
Programmable reminders	✓	✓	✓	✗
Drugs administration reminders	✓	✓	✓	✗
Pharmacological compliance	✗	✗	✗	✗
Emotion recognition	✓ User self-assesment questionnaire	✗	✓ Face and Voice Analysis	✗
Attention to severe disabilities (no movements, dysarthria, etc)	✗	✗	✗	✗

Fig. 1.6 shows the percentage of solutions that offers the services reported in Tables 1.1 and 1.2. According to data reported in Fig. 1.6, it can be noticed that the main prerogatives of companion robots in AAL are: (1) autonomously navigate within the house (100%); (2) provide entertainment to the user (100%); (3) improve his/her social sphere (85.71%); (4) generate reminders (85.71%). Most of the solutions are also designed to submit the user cognitive tests (57.14%), however, only a few solutions (14.28%) use data from the cognitive games to extract parameters useful for related assessments. The same can be said for physical activities.

Fig 1.6 also emphasizes that despite almost all the robots being provided with drug administration reminder capabilities, no solutions embed a pharmacological compliance step. Moreover, most of the companion robots proposed in the context of the European projects have no arms, or if they have, they are only used for animated interactions.

Finally, no AAL robots' solutions are designed to address the problems of older adults with severe disabilities despite being widely diffused.

Another interesting result of the present survey concerns the robot's shape. Indeed, as it is possible to infer by observing Fig 1.5, most of the companion robots used for AAL purposes are research-grade versions of commercial social robots with anthropomorphic shapes (except for CompanionAble [44]). This shape is selected to improve the acceptance rate of robots in a domestic environment. Indeed, it has been demonstrated [49] that (under specific conditions[1]) anthropomorphic shape reduced the refuse rate in the phase of adaption (1 month from the robot inclusion in the domestic environment), incorporation (2-6 months), and identification (6+ months).

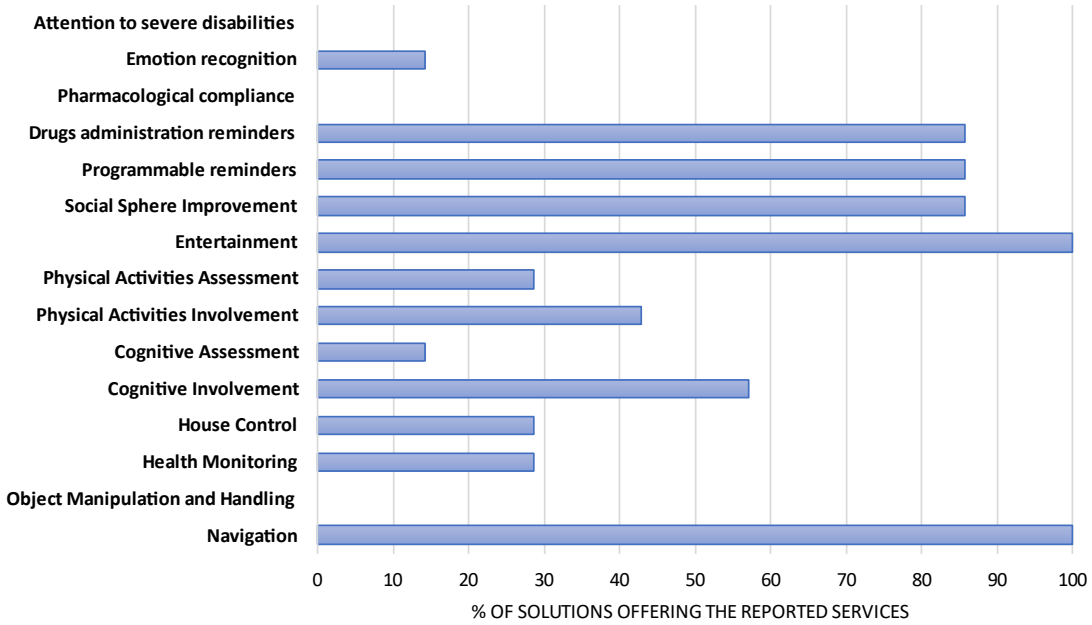

Figure 1.6 Percentage of state-of-the-art companion robot solutions that offer specific services

[1] Specific conditions concern the tone of voice (neutral is preferred to positive), the reduced theatrical of the animation is preferred.

1.4 AMICO Project

The present thesis was born in the context of a robotics-empowered AAL system funded and realized in the context of the NOP Research and Innovation 2014-2020, which finances industrial research and experimental development projects in the 12 specialization areas identified by the National Research Programme (PNR) 2015-2020, and specifically in the "Technologies for Living Environments" specialization area.

The project, named AMICO (Medical Assistance In COntextual Awareness), proposes the development of an AAL infrastructure consisting of a person (primary user), his/her domestic environment, both equipped with sensors, a telemedicine platform, and a robot that acts as a mediator between the user, the surrounding and the external environments [50, 51].

The companion robot employed for the AMICO project is Pepper by SoftBank Robotics [29]. Pepper robot has been chosen for manifold reasons. The first one lies in its previous large diffusion in European care facilities and smart homes (~12000 units only in 2018). It permits to easy transfer of the whole architecture in previously robot-empowered infrastructures.

The second reason lies in the high acceptance rate of the Pepper robot's physical characteristics (e.g., anthropomorphic shape) according to [49]. The third reason lies in the presence of upper arms designed to allow the robot an effective animated verbal interaction with users.

The proposed infrastructure, and specifically the robotic agent, provides services for people that cover entertainment (e.g., mobile terminal with, movies and music), reminders (including drug administration warnings), social sphere enhancement through video call/call channels accessible by tablet apps (when the internet connection is active).

The platform is also oriented to health monitoring through, *inter alia*, new wearable ultrasonic sensors for the measurement of hemodynamic parameters and contactless temperature sensors [52].

The health checking outcomes are, thus, sent through a secure connection to a Fast Healthcare Interoperability Resources (FHIR) server implementing a telemedicine platform for the remote monitoring by doctors and caregivers of the user.

As the main objectives of this thesis, several new features have been introduced by Politecnico di Bari (Design of Electronic Integrated System Lab) in the AMICO infrastructure:

- The robot has been equipped with an additional user interface based on the *Brain-Computer Interface* (BCI). This interface has among the advantages the possibility of use by any type of user, regardless of the severity of the user's disability (e.g., Tetraplegia that makes it impossible to use any mechanical interface, as well as tremors/dyskinesia associated with age, Parkinson's Disease and parkinsonisms, dysarthria that makes voice commands difficult to be implemented, etc.)

- The BCI, which is based on the analysis of electroencephalographic (EEG) signals for service selection, is used for two other contemporary purposes. The first of these, conducted in parallel with the selection of services, consists of the *quantitative* assessment of the cognitive status of the patient. The second one, operating at regular intervals throughout the day, concerns an emotion recognition system based on EEG in order to achieve an adaptive human-robot interaction.
- To allow a correct cycle of service selection that includes the receipt of the desired specific goods, the Pepper robot has been functionally reprogrammed to implement non-intrinsic object manipulation features.
- In the same context, when the good to be delivered is a pharmaceutical product for self-medication or due to related scheduling, the robot has been programmed to verify the actual reception of the drug, operating the first phase of pharmacological compliance.

It should be stressed that all the new features have been realized through routines that can operate offline (without an internet connection) to guarantee a working version of the robotic platform even to users who have raised concerns related to privacy [13].

Ultimately, the inclusion of the aforementioned capabilities allows not only to make the infrastructure more inclusive (bigger catchment area), but also to cover aspects not investigated or under investigation in the context of AAL robotics.

In addition, the introduction of infrastructures able to *quantify* the psychophysical state of the user, also and not only under the cognitive aspect, opens new and interesting AAL applications that can perfect the life cycle of the users in accordance with the advice of experts. At the same time, knowing the emotional state of primary users would allow them to make decisions oriented to the well-being of the same and to undertake adaptive processes of interaction by the entire infrastructure (both robot and environment).

The BCI-based AAL Framework

The AMICO project aims to provide an inclusive AAL framework in which three main actors, i.e., user, robotic platform, and environment, interact with each other in a novel and collaborative way. In this respect, this chapter will outline the overall AMICO AAL framework, focusing on the contributions introduced by the thesis work. Next, a working principle will be proposed, to provide the readers with possible applications of the system. Finally, the chapter will detail – through three dedicated sections – all the main characteristics of the main actors of the architecture.

2.1 Overall Architecture

As stated in Sec. 1.4, the AMICO infrastructure has introduced innovative features in the context of robotics-empowered AAL, to date, still weakly analyzed at the state of the art. These features make the infrastructure inclusive towards bedridden users with severe disabling pathologies, allowing a contextual analysis of the cognitive and psychophysical state of the user, keeping him constantly updated on the state of the instrumented house.

Fig. 2.1 shows the overall architecture of the proposed infrastructure. Specifically, it reports *only* the functionalities introduced in the context of the present thesis work, delegating the analysis of the general potential of the system to the brief description in Sec. 1.4 and to the previous work available at [50].

According to Fig. 2.1, the proposed system can be divided into four main blocks: (i) the user (or patient); (ii) the Brain-Computer Interface (BCI); (iii) the Robotic Platform, and (iv) the Environment.

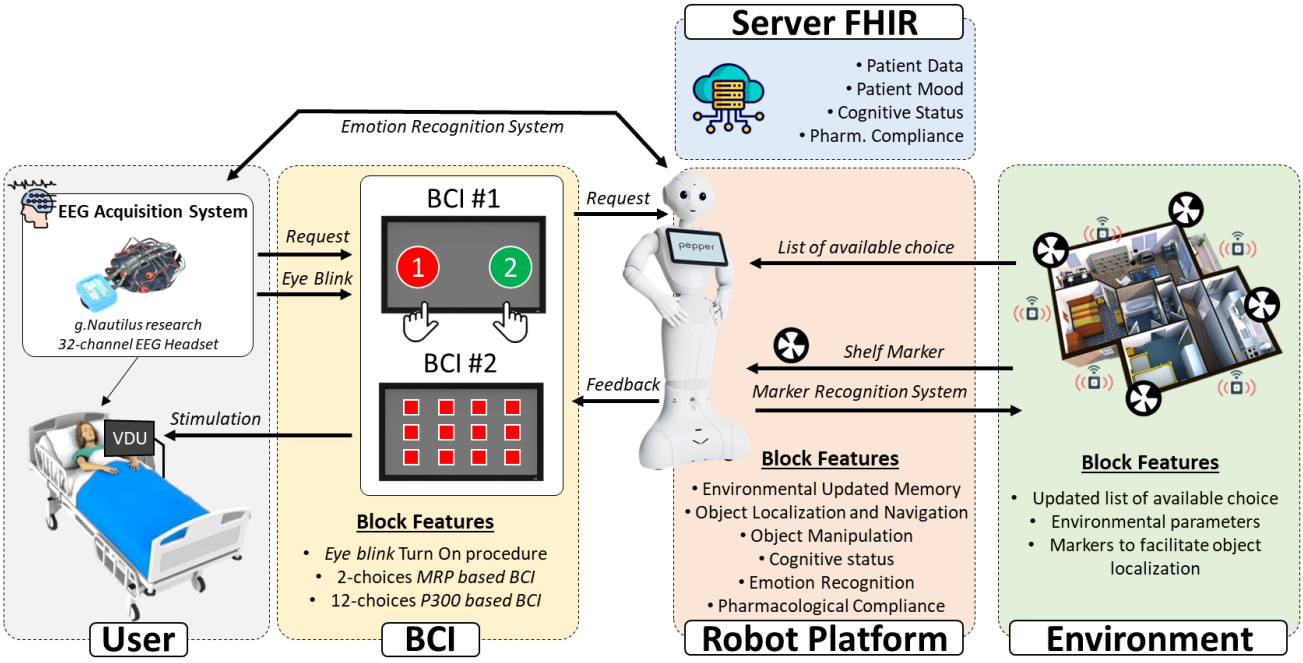

Figure 2.1 Block diagram of the multi-platform architecture introduced in the context of the AMICO project

The first section (i.e., the User) is composed of the user himself, the EEG acquisition system, and a generic video terminal unit used to stimulate the user's cortical potentials through video/audio.
The second functional block, i.e., the BCI, can be embedded both in the video terminal (if it is a computation unit such as a PC) and the robotic platform. For the proposed application, the BCI block is implemented on the robotic platform. The block is designed to "translate" brain activity into practical and unique commands for the robot. This section of the architecture also oversees the management of the stimulation protocol to be submitted via video terminal.
The Robotic Platform consists of the Pepper robot. In this framework, it is programmed to: (i) implement BCI algorithms; (ii) elaborate user's requests; (iii) navigate to reach and manipulate objects; (iv) analyze EEG data to extract the cognitive status of the user; (v) examine EEG to check the user's mood; (vi) send data to telemedicine services (Server FHIR – Fig. 2.1).
The Environment block consists of a series of sensors and markers distributed in the instrumented environment. The use of markers facilitated the integration of a preliminary phase of autonomous navigation for object localization. The distributed sensors used for the specific application permit the identification of the presence/absence of specific goods on a dedicated shelf, allowing the continuous updating of a list of available services.

2.2 Working Principle

The following section will describe a general principle of operation of the proposed architecture to provide readers with a schematic overview of the potential of the infrastructure.
The infrastructure includes the bedridden user wearing the intended EEG headset, and lying in front of a video terminal.

2.2.1 Service Selection via BCI

Wake-up and BCI selection. First, the user/patient who wishes to interact with the robot must activate its wake-up state. This wake-up procedure is realized through a functional sequence of prolonged eye blinks. In particular, the patient should blink his/her eyes 4 times in less than 4 s.
Once wake-up state is achieved, the video terminal shows a system message in which the user is required to re-propose, after an audio signal, 3 eye blinks and 1 second pause to activate the procedure of BCI #1, otherwise 4 eye blinks and 1 second pause to activate BCI #2. The two BCIs differ from cortical pattern to be "translated". These cortical patterns determine the number of possible choices and the speed of the BCI.
More details concerning cortical patterns employed to implement the BCIs are detailed in Sec. 3.1.
The BCI #1 allows the user to select between 2 available choices per session, while BCI #2 permits to select among 12 choices. In the first case, the selection is formalized in ~ 1 s, while in the second case the selection lasts ~21 s.

BCI #1 Use Case. BCI #1 is structured to show the user two lists of possible actions to be taken. As an example, these selections could be {*Bring me...* , *Call...*}. Through the specific stimulation protocol, detailed in Sec. 3.1.3, the user selects one of the two possible choices. This selection will open an additional nested binary choice. As an example, suppose the user chose *Bring me...* . In this case, the second choice could be {*Food, Drink*}. Continuing with the decision tree realized via *nested binary choices* it will be possible to provide a large number of final requests at the expense of selection times. This BCI requires a minimum movement of the fingers (extensor/flexor muscle tone lower than a mouse click).

BCI #2 Use Case. BCI #2 presents a screen with 12 choices. Through the specific stimulation protocol, detailed in Sec. 3.1.3, the user can directly select one out of 12 choices. BCI #2, like BCI #1, can also be used to nest choices. At present, this feature has only been implemented for food and drink requests.

Actuation Routine (Navigation and Manipulation). Once the BCI-based request has been uniquely formalized via BCI #1 or BCI #2, the robot first enables the optical/acoustic sensors for autonomous navigation and, finally, plans the movement from the starting point in the vicinity of the user. This reference in space is identified in the following as Home.

Having in memory the coordinates of the various objects to be taken, or the paths to follow for the various procedures to be completed, the robot can navigate autonomously up to the prescribed position. More details will be provided in Sec. 4.3.

In the case of objects to be picked up and delivered to the user and/or caregiver, the robot will navigate to the shelf that contains the specific object reaching a position called the approach area.

Starting from this position, the robot completes some preliminary steps which involve the initialization and alignment of the outputs of the RGB cameras, and a 3D sensor integrated into the Pepper robot. As a second step, the proposed system estimates the height of the shelf containing the objects. It is useful to better plan the subsequent movements in order to avoid collisions. Afterward, the robot starts scanning the frontal area by means of an RGB camera and a pre-trained object detection algorithm. Once identified the target, the robot selects a specific manipulation routine accordingly to the position of the object (i.e., placed centrally or laterally on the shelf). More details about the implemented routines are available in Sec. 4.4. The robot is now ready to properly deliver the object to the user or caregiver.

The above-described procedure is shown in Fig 2.2 and Fig 2.3. In the first figure, i.e., Fig. 2.2 the procedure is represented as a temporal diagram reporting the involvement of the architecture actors, while Fig. 2.3 shows the procedure through a five-frame demonstrative sequence.

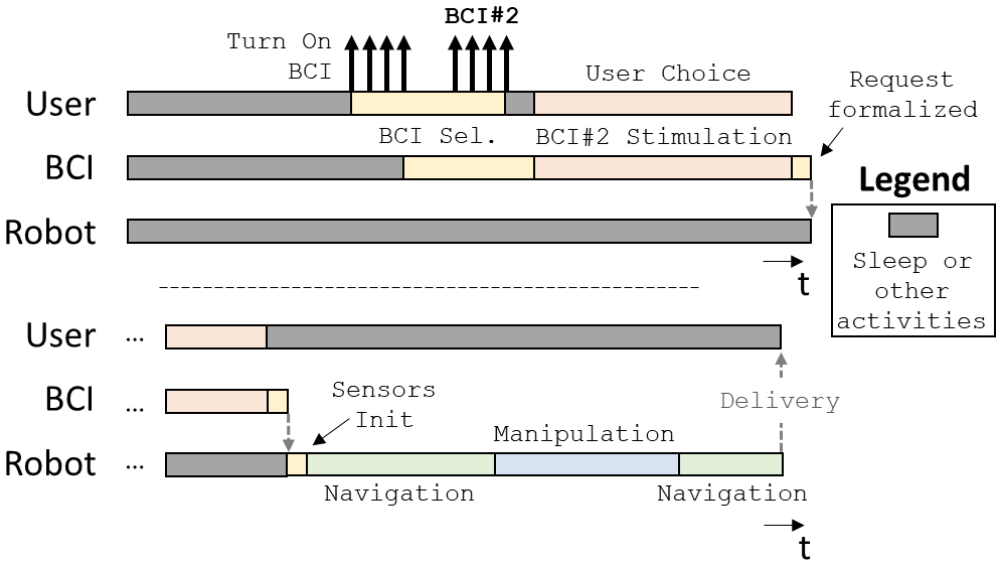

Figure 2.2 Temporal diagram of the service Selection via BCI use case

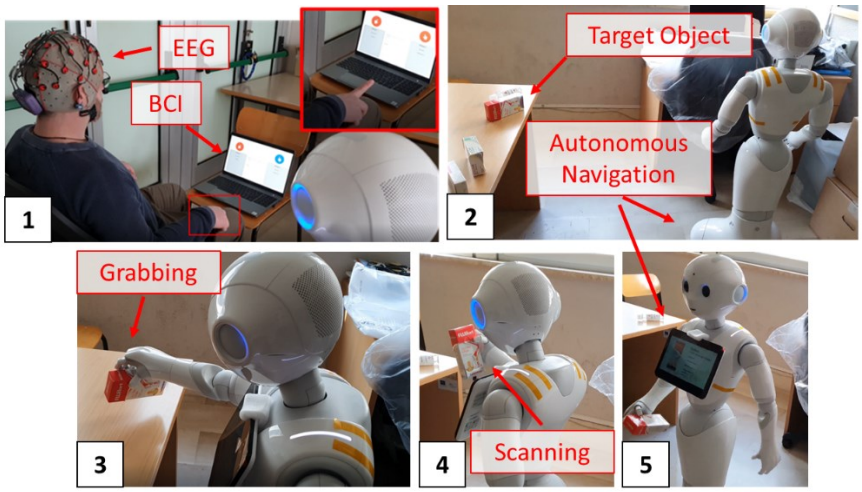

Figure 2.3 Five-frame demonstrative sequence of service Selection via BCI: (#1) BCI-based request formalization; (#2) autonomous navigation to approach area; (#3) object manipulation procedure followed by (#4) Scanning process for checking the grabbing procedure; (#5) autonomous navigation to approach the patient/user for goods delivery.

Stop BCI action. To block any operation (i.e., generate an interrupt in the robot's action cycle), the user must eye blink 5 times in a short time (< 5 s). In this case, the robot will stop the operations, asking the user for a confirmation of the stop, opening to an action list: {*Yes, No*} depending on the BCI previously enabled. If *Yes* is selected via BCI, the robot will retrace the recorded route back to the Home. Then he goes back to *Sleep* waiting for a new Wake Up signal. If *No* is selected, the robot continues with operations.

Robot-to-Environment Interface. During the selection phase (regardless of source BCI#1 or #2), the user may ask the interface for a list of foods, drinks, and drugs available at that time or some information about the home environment (temperature, humidity, lighting). In this case, Pepper will query its memory at specific event locations, in which -asynchronously - the environment block stores data as soon as they are updated (via a dedicated Memory Raise Event interface) [53]. In this way, Pepper is always

updated on the temperatures/humidity/lighting of the various rooms, and on the foods, drinks, and medicines available on the shelves predisposed to manipulation.

2.2.2 Pharmacological Compliance

One of the possibilities provided by the new features of AMICO is to receive pharmacological compliance feedback that does not stop at a simple reminder of drug administration.
It is possible, in fact, that the user must receive some medicines at specific times, or that the same user requests some medicines for self-medication through the use of one of the BCIs introduced in the previous paragraph.
In this case, the navigation phase toward the target object retraces the steps already addressed in Fig. 2.3.
Once it reaches the patient, the robot will provide the user (patient or designated caregiver) with all the indications for the recovery of the selected object. The pharmacological compliance steps are shown in Fig. 2.4 as a temporal diagram and in Fig. 2.5 as a three-frame demo sequence.
At present, the medicine recovery procedure requires the user to place the hand under the robot one to receive the object (Fig. 2.5 – step #2). An alternative to this step is to place the object on a plane close to the user (e.g., a table or bedside table). The deposit procedure has also been implemented.
Following the deposit or delivery, if the delivered item is a medicine, Pepper will wait in close proximity to the user to complete the confirmation phase of the administration.
Once the drug is taken or provided to the patient (in the case of caregivers), the box of the same medicine should be placed in front of the robot's camera. The latter will provide for its recognition by completing the confirmation of administration and notifying it to the FHIR server integrating the telemedicine services. The confirmation of administration of the drug procedure is shown in Fig. 2.4 as a temporal diagram and Fig. 2.5– step 3.
Future works will concern the analysis of the content of the drug envelope, through - for example - the counting of the blisters used compared to the starting ones, or the recognition of the action by the user.

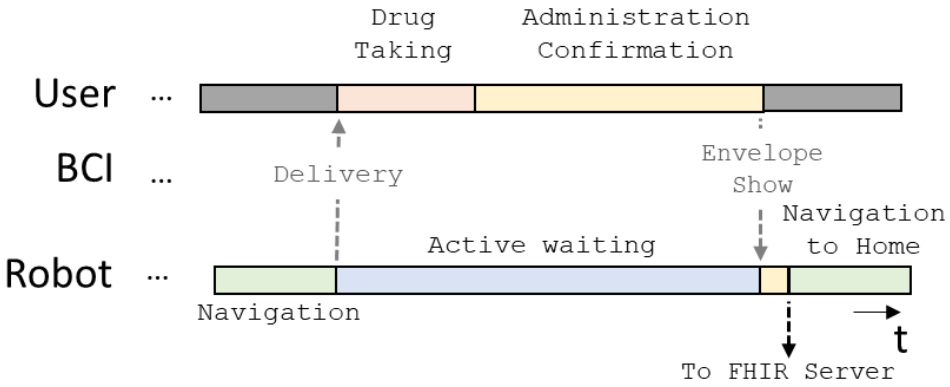

Figure 2.4 Temporal diagram of the Pharmacological Compliance use case.

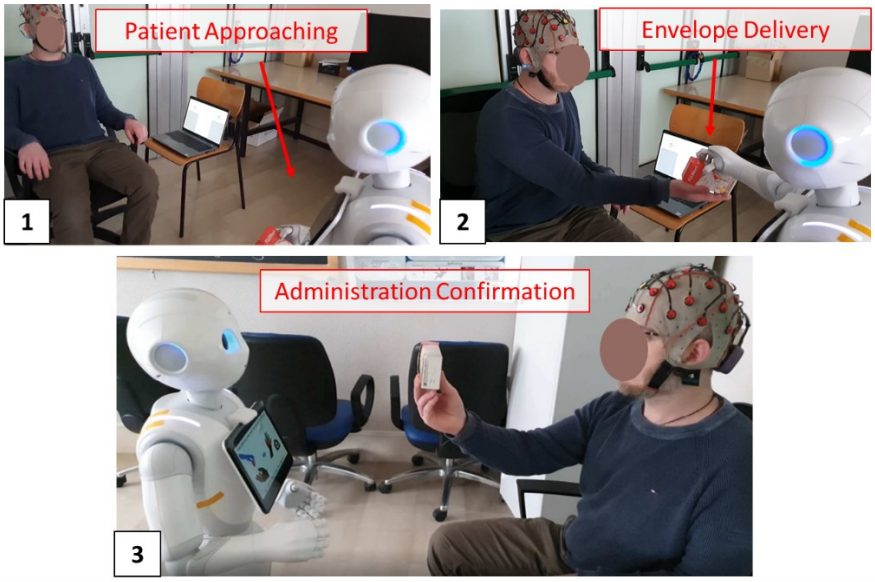

Figure 2.5 Three-frame demonstrative sequence of Pharmacological Compliance: (#1) patient approaching procedure with embedded face recognition; (#2) drug delivery through guided procedure; (#3) confirmation of the administration after taking medications.

2.2.3 Cognitive Status Extraction

The second innovative feature introduced by the here-presented AAL framework is the possibility of providing a quantitative assessment of the user's cognitive status. This feature adds a *cognitive assessment* functionality, often lacking in state-of-the-art solutions as shown in Sec. 1.3.

At the implementation level, the cognitive assessment is carried out every time the user formalizes a choice via BCI #2. The reason lies in the fact that the cortical pattern on which this BCI is based is also often used as a biomarker in the analysis of cognitive status at the clinical level.

In the context of the proposed application, the system accumulates data while using BCI #2, providing - after a preset number of observations - a Figure of Merit (FoM) to the telemedicine hub. This figure of merit can be used to support an additional and definitive cognitive assessment that is left to the doctor who follows the patient remotely. Fig. 2.6 shows a demonstrative temporal diagram of the cognitive status extraction use case.

2.2.4 Emotive Status Extraction

One of the fundamental roles of an AAL infrastructure is to improve the well-being of primary users and lighten the effort of secondary users. Recently, the aspect related to the improvement of well-being in AAL contexts has approached the concept of adaptive improvement of the assistive infrastructure. In particular, AAL infrastructures have evolved in favor of a mood/emotion-driven approach [5]. This type of approach allows to improve of the Perception – Comprehension – Action cycle (often named *affective computing*) in the field of social Human-Robot Interfaces (sHRI).

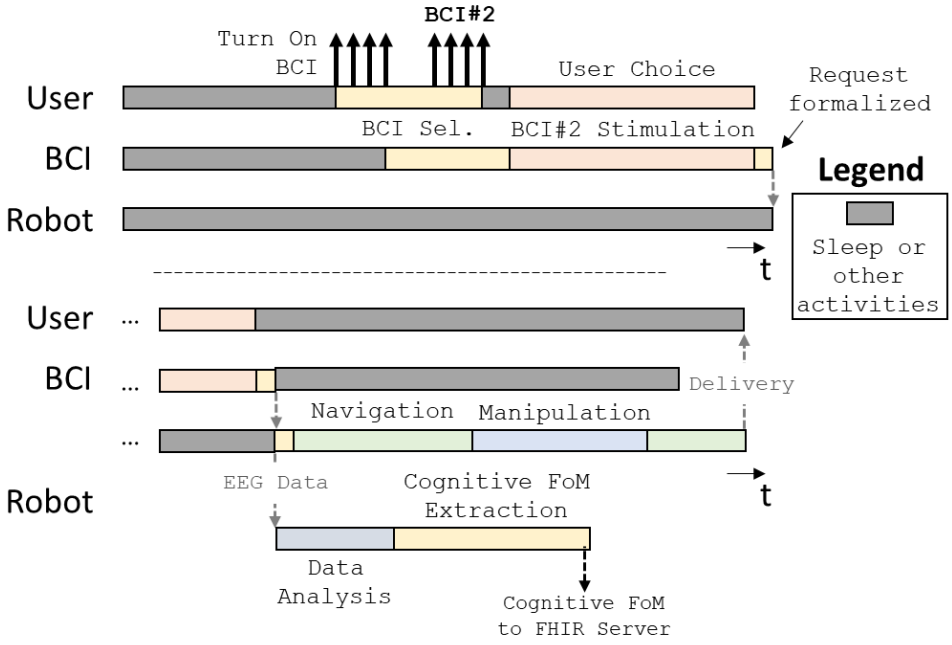

Figure 2.6 Temporal diagram of the Cognitive Status Extraction use case.

In this context, a new feature introduced during this thesis work concerns the design and the implementation of a new inference engine that aims to improve the Perception – Comprehension phase of the affective computing cycle.

The introduced method involves a first phase of system calibration, during which the robot will interact with the primary user verbally, with animations, videos and music. At the end of each stimulation session, the user (or caregiver on behalf of the user) will be provided with a short survey to be filled out. It consists of the assignment of some scores to three descriptive parameters of emotions. These parameters are: Arousal, Valence, and Dominance.

According to the model in [54], the *valence* parameter is the level of the pleasantness of the emotion. Low values of the valence correspond to negative psychophysical effects on the subject, while high values correspond to emotions with positive effects. A*rousal* is the level of activation of an emotion. Low arousal values represent a mood with low emotional intensity, while the high level corresponds to a mood with high emotional intensity.

Finally, *dominance* is the level of control "on-and-by" an emotion. Low values of dominance identify a situation in which the subject is controlled by the emotion, while high values correspond to the capability of the subject in controlling the emotion.

Based on the self-assessment diaries and depending on the EEG data collected during the calibration phase, the system extracts an inference model capable of identifying a specific emotional state for the patient.

Once the inferential engine has been instantiated, during nominal operation, the robot will approach the user with random pace during the day, undertaking short verbal interactions with him/her.

Following the interaction, Pepper will analyze the patient's EEG signals, proceeding with the classification of his current emotional state. This parameter will first be sent to the FHIR server for retransmission to the telemedicine service and, subsequently, will be transmitted to the adaptive conversational/interactional system that is not covered by this thesis work. Fig. 2.7 depicts the temporal diagram of the above-presented procedure.

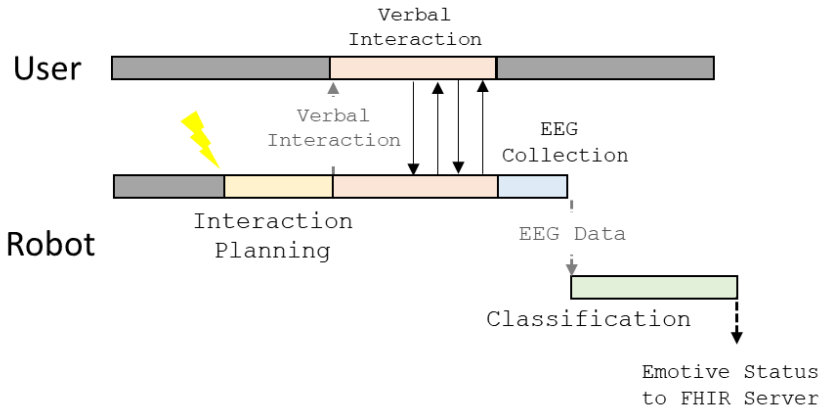

Figure 2.7 Temporal diagram of the Emotive Status Extraction use case.

2.3 The User

As can be seen in Fig. 2.1, the first main actor of the presented infrastructure is the User. Specifically, in this application, the User consists of a video terminal and the EEG acquisition system.

As for the video terminal, in the proposed application, is implemented through a notebook connected via Bluetooth (or WiFi) to the robotic platform. The video terminal is used for the reproduction of stimulation protocols made in the Unity environment.

The Bluetooth/WiFi interface is mainly used for the piloting of the stimulation protocol by the robotic platform.

Concerning the EEG acquisition system, the hardware selected for the experimental phase is g.Nautilus Research by g.tec medical engineering GmbH [55].

The g.Nautilus system consists of an EEG headset with 32 active electrodes with a wet skin interface and a transmitting device connected to the back of the headset. Fig. 2.8.a shows a snapshot of the experimental setup used to test the proposed BCI. The EEG headset is labeled as #1 in Fig 2.8.a.

The overall g.Nautilus Research system also consists of a static reception unit (base station) that can be connected via USB to the work terminal (PC, notebook, robot, etc.). In this specific application, the base station is connected to the video terminal (i.e., a notebook), which oversees the streaming of the EEG data toward the robotic platform, acting as the gateway in transparency. The actually proposed data exchange map is shown in Fig. 2.8.b. Two alternatives have been proposed in the context of the AMICO project by using dry electrode EEG headsets: Unicorn Hybrid Black by g.tec neurotechnology GmbH [56] and Enobio by Neuroelectrics [57]. These two headsets allow the EEGs to be directly sent via Bluetooth or TCP/IP toward the robotic platform, without passing

through the video terminal. The connection map related to these alternatives is presented in Fig. 2.8.c.
The distribution of the electrodes network of the g.Nautilus Research responds to the international regulation imposed by the 10/20 system [58]. The interelectrode connections and the connections between the electrodes and the transmitting station are ensured by flexible and non-removable PCB tapes reducing non-physiological artifacts related to bad wiring [59].
The g.Nautilus Research system guarantees the minimization of electrode-skin impedance, allowing a reliable measurement operation, reaching values between 30kΩ and 100kΩ through the use of electrolyte creams and conductive gels (in this case g.GAMMAgel)[2].
High impedance values are instead guaranteed by the integrated preamplification stage (>100MΩ) which makes the electrode classifiable as active.
The headset embeds an integrated Analog-to-Digital Converter (ADC) per channel (for a total of 32 ADCs) with a declared resolution of 24 bits (with a Sensivity ranging from ±185mV to ±2.25V). Each converter performs a 1024MHz oversampling.
A 32-to-1 multiplexer channels the information content of the selected channel outbound. The signal extracted downstream of the multiplexer is the input of the XVV-MEGA22M00 RF module.
The radio module integrated into the transmitting station (developed by Dresden Elektronik) combines an ATmega128RFA1 and a ceramic onboard chip antenna.

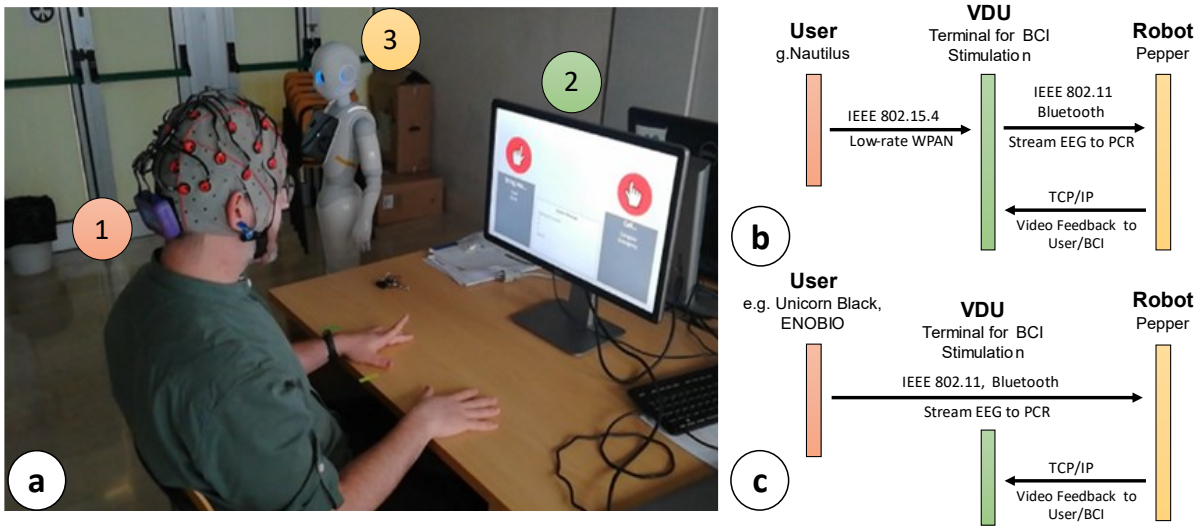

Figure 2.8 Experimental Setup of AMICO infrastructure. (a) Experimental setup for BCI-based selection with BCI #1. Labels: #1 g.Nautilus Research EEG headset; #2 Video terminal (VDU); #3 Robotic Platform: Pepper. (b) Current connection map. (c) Alternative Connection map.

[2] Two dry alternatives have been investigated during the AMICO project with results that do not differ substantially from the wet equivalent. These two alternatives are Unicorn Hybrid Black by g.tec neurotechnology GmbH and Enobio by Neuroelectrics.

The antenna transmits with IEEE 802.15.4 standard protocol (low-rate Wireless Personal Area Network) at 2.4GHz, reaching transmission channel power between 2.3 and 2.9dBm within an indoor range of 20m.
The user is allowed, during the configuration of the device, to choose the sampling rate parameter between 250Hz and 500Hz, to provide output data.
Through dedicated C-language Application programming interfaces (APIs) provided by the manufacturer, it is possible to choose the sampling rate and full scale of the measurements on each individual channel.
It is also possible to set digital pre-filtering actions and Noise Reduction protocols. The user can choose among 15 pre-set digital bandpass and a Notch filter (48-52Hz or 58-62Hz).

2.4 The Robotic Platform

The robotic platform employed for the specific AAL framework is Pepper v1.8a, a human-shaped social robot developed by SoftBank Robotics [29]. The robot has a height of 121 cm and weights approximately 29 kg.
Pepper is supplied with an Intel Atom E3845 (quad-core) 1.91 GHz processor, 4GB DDR3 (2 GB dedicated to Operating System (OS)) of RAM, 8GB eMMC (not available to user), and 16 GB of micro SDHC. All the routines proposed in the following run on Pepper's native operating system, NAOqi OS, which is a Gentoo-based GNU/Linux distribution [29]. For the sake of readability, only the details related to the sensors that are involved in the here proposed routines will be described in the following.

RGB Camera. Pepper is provided with two RGB cameras and a 3D depth sensor as per Fig. 2.3. The RGB cameras are two OV5640 by Omnivision. The first one, identified as OV5640top is placed on the forehead with an angle of 90° with respect to the head frame axis (0° considering the x-axis reference in Fig. 2.9). The second camera, labeled OV5640bot is placed on the robot mouth with an orientation of -40° with respect to x-axis reference in Fig.2.3. The OV5640 provides a maximum resolution of 5 Mp with 55.2° of the horizontal field of view (FOV), 44.3° of vertical FOV, and a diagonal FOV of 68.2°. The output frames are set to kVGA (640*480 px) with 10 fps.

Depth Camera. The embedded 3D sensor, located in the left eye of the robot, is an ASUS Xtion with 0.3 Mp of resolution.
The Xtion provides 58° of horizontal FOV, 45° of vertical FOV, and a diagonal FOV of 70°. The depth camera output has been set to be kQVGA (320*240 px) with 5 fps.

DC Motors. The routines also involve motors management to move the arm segments, represented for the sake of clarity only for the right side as S1, and S2 in Fig. 2.9. Specifically, routines manage six different motors for the different robot sections: head, arms, and hands.
A brush DC coreless SE24PCTCA is driven for the head yaw and the shoulder pitch. A SE24M1RTCA is used to drive the head pitch and the elbow roll.

Finally, the hands section is managed by two 17N88208E for the wrist yaw and the hand open/close. For navigation purposes, Pepper is also supplied with 3 holonomic wheels (spheric shape) placed at the base of the automaton.

Lasers. For navigation purposes, Pepper is equipped with 6 laser line generators emitting at 808nm with a framerate of 6.25 Hz per laser. As per Fig. 2.9, there are 3 main laser scanners pointing toward the front, right and light directions. Each laser scanner is composed of 15 laser beams able to update the directive distance value in ~160 ms. According to the constructor directives the Horizontal Field of View (HFOV) of each laser scanner is 60°, while the Vertical Field of View (VFOV) is 40°. It leads to an overall HFOV of 240° with 60° of blind angle equally distributed across 45° and -45°. There are also 3 laser scanners used for front-ground evaluation. The maximum distance of detection is defined as up to 10 meters with a maximum height of 10 cm for the objects that the lasers can detect.

Tablet. The pepper robot is provided with an LG CNS Tablet fixed on the chest frame. The tablet directly connected to the robotic system is supplied with an Android 4.0 Mini PC TCC8925 Cortex A5 1.0GHz 1080p Full HD VPU + Mali400 3D/2D GPU. The tablet has 1 GB RAM and 4 GB eMMC. The tablet has a built-in predisposition for IEEE 802.11 b/g/n/a WiFi, and it has been modified to receive also via Bluetooth dongle. Softbank Robotics permits developers to realize custom Android apps to be updated via a dedicated platform. The manufacturer provides Android API (Pepper SDK plugin) to manage all the routines/methods provided by NAOqi.

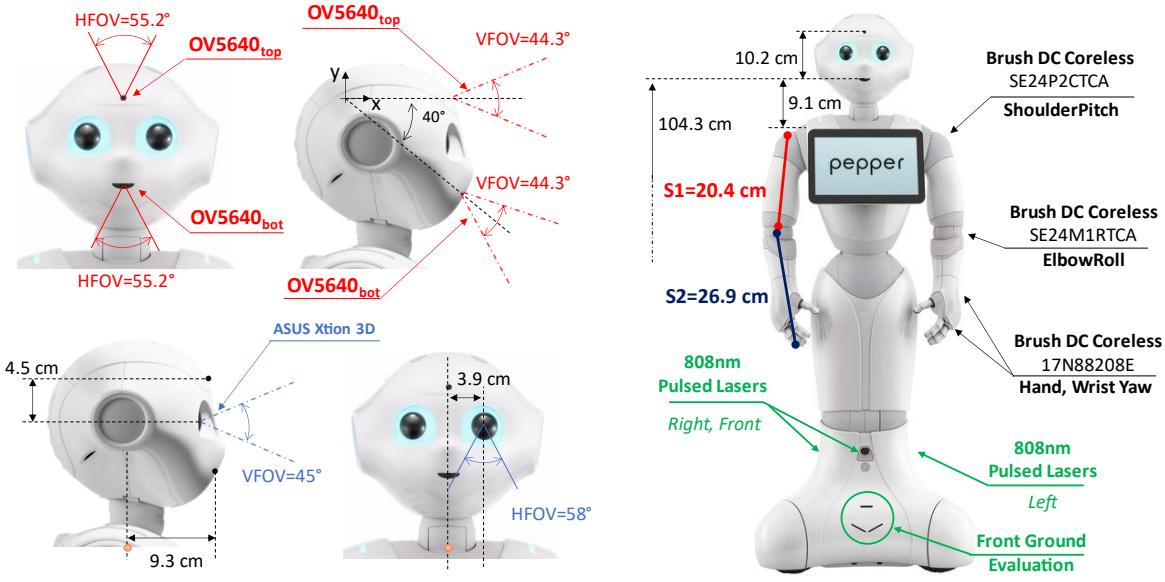

Figure 2.9 Overview of employed Pepper robot hardware features

2.5 The Environment

The functional block called Environment consists of sensors and navigation markers distributed in the smart home environment.
Specifically, one of the roles of this block is to favor a correct shelf approach phase for the manipulation of objects by the robot. It is realized through the specific use of textured images known as NAO marks [60], as will be deepened in Sec. 4.4.
As seen in Sec. 2.2.1 (Robot-to-Environment Interface), another main role is to update, in real-time, the memory of the robot about: (i) the services available at any time of the day; (ii) environmental parameters related to rooms and shelves used to store goods.
For this purpose, the house is littered with several multi-sensor nodes capable of extracting environmental parameters such as temperature, humidity, and ambient lighting. An example of a multi-sensor node realized in the context of the present thesis work and used for the specific application is shown in Fig. 2.10a [61].
Each multi-sensor node consists of a *low-cost and components availability-oriented* custom board. The microcontroller that constitutes the computation core of the node interfaces via I2C three sensors: (1) a sensor dedicated to temperature (T) and relative humidity (RH) acquisition; (2) an Ambient Light Sensor (ALS) and (3) an Inertial Measurement Unit (IMU) breakout board. The board also provides some headers to exploit the unused General Purpose I/O (GPIO). Each node includes a module for Bluetooth Low Energy (BLE) transmission. The main features of the multi-sensor node are reported in Fig. 2.10a.
The device uses a 3.7V LiPo battery for power supply. Let us consider: (i) a LiPO battery capacity of 2000 mAh; (ii) one read every 30 s from all the sensors and their BLE transmission; (iii) IMU sensor down (not used); (iv) no sleep mode for the serial interface circuitry. Considering the above-mentioned condition, the node can work for a minimum of 110 days (worst case).
The nodes' firmware is designed to be updated via a dedicated GUI exploiting an *over-the-air* framework. There are two sensor node versions: the integral one (see Fig. 2.10a) which measures 60 mm x 48 mm and the compact version (without serial interface) which measures 39 mm x 48 mm.
Each environmental monitoring node can be reused as a shelf controller for the detection of the conservation parameters of the goods and the verification of their presence. An example of functional reuse of the multi-sensor node for shelf monitoring is shown in Fig 2.10b.
For this purpose, the multi-sensor node provides some inputs (A0-A3 in an Arduino-like pinout of ATMega) suitably conditioned for the insertion of Force Sensitive Resistor (e.g., FSR402 by Interlink Electronics [62]).
Such sensors are used to detect the presence on them of weights greater than 11g. Exceeding this threshold is detected as the presence of the specific object.
In this way, the multi-sensor node is aware of the goods available for delivery, and – at the same time - their storage conditions. In this respect, our work [61] has shown the

applicability of the node also in the dynamic redefinition of the goods' expiration date starting from the above environmental parameters. In that case, the node is also equipped with the computing capacity of the shelf life, to prevent food waste [61].
Since the objects are distributed in specific locations of the smart home (known by the robot), each FSR point is associated with only one type of object. The change of state linked to each FSR point (e.g., removal of a previously present object or repositioning of the same), will trigger the multi-sensor node to send the updated list of available choices.

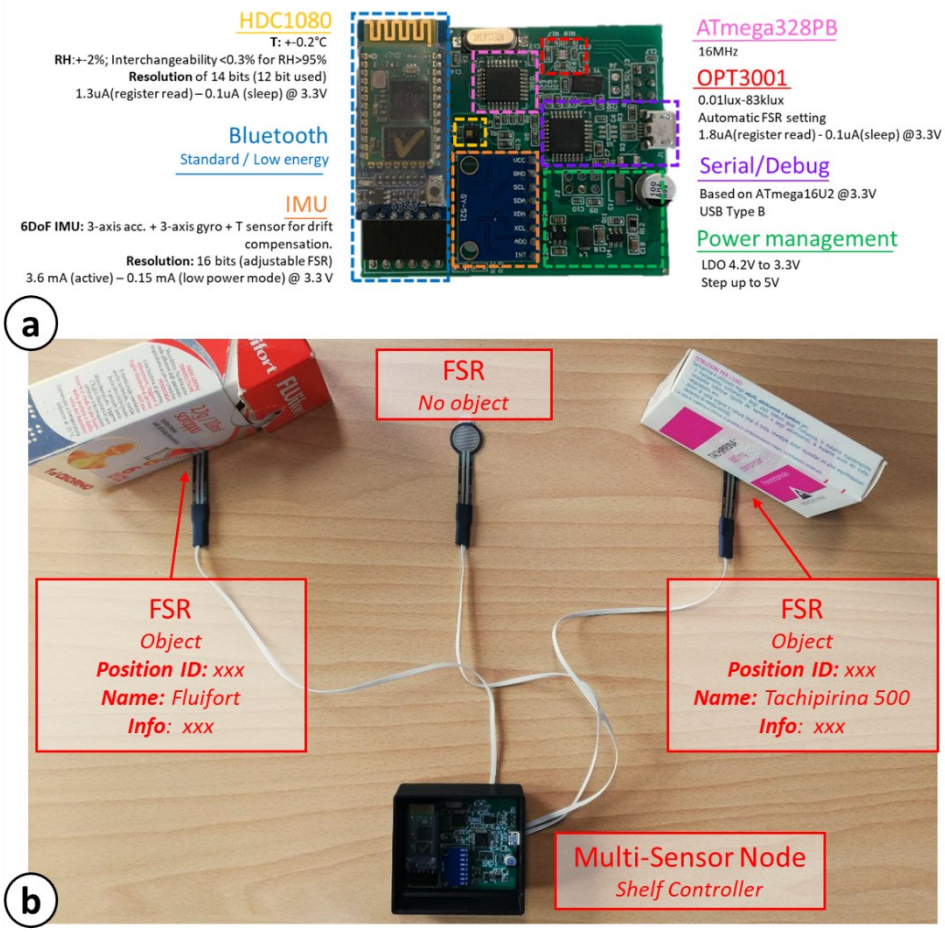

Figure 2.10 Overview of employed Pepper robot hardware features

The transmission of these parameters to the robot is carried out via BLE to an intermediate gateway that collects the data from the various multi-sensor nodes and retransmits them via WiFi/BLE to the robot's memory system. The gateway was built in this context using a Raspberry Pi 4.

The AAL Brain-Robot Interface

The most important novelty introduced in the context of the proposed AAL framework is certainly the *brain-robot interface*. This interface can be used for multiple reasons such as providing choices or assessing the cognitive and emotional state of the primary user. These are essential features for improving the Perception-Comprehension-Action cycle in assistive and social robotics.
In this chapter the cortical patterns useful for the realization of BCI will be analyzed. These will be studied under the physiological and implementing aspect. A methodological and implementation definition of the pre-processing steps, as well as of the presented BCIs will follow. A section will then be dedicated to the system that extracts a figure of merit related to the cognitive state of the patient. Finally, a final section will analyze the emotion recognition system implemented by the AMICO infrastructure.

3.1 Cortical Patterns for BCI

The aims of the AMICO infrastructure range over two fundamental fronts at the care level, i.e., the clinical-diagnostic and the decision-making (for the formalization of needs).

In this regard, the interposition of a Brain-Computer Interface (BCI) between the primary user and the AAL robot is well suited to address both project constraints.
In fact, it is possible to identify some cortical patterns able to operate both as a diagnostic biomarker and as a support to decision-making systems. Specifically, two types of cortical patterns will be addressed in the context of this thesis.
The first pattern will exploit a category of signals called Movement-Related Potentials[3] (MRP). These cortical potentials are related to the planning of specific movements and can be generated endogenously by the user without the need for a specific stimulation system. This type of signal is used for fast decision-making BCI applications with a low number of choices, even asynchronous [63].
The second pattern will exploit a category of cortical signals called Event-Related Potentials (ERP). These cortical potentials arise in response to particular visual/auditory/tactile stimulations from the external environment (e.g., via stimulation video terminal presented in Sec. 2.1). These evoked potentials are also often used in clinical practice to quantify patients' cognitive impairment [64]. The use of ERP for decision-making BCI applications is also widespread [65]. These BCIs can guarantee a large number of choices at the expense of a reduced speed.

3.1.1 BCI #1 Cortical Pattern: Movement-Related Potentials

During voluntary movements, activation of proper muscle sequence is preceded by a cerebral preparation process. This movement planning process usually starts about 1 s before muscle contraction and is realized, *inter alia*, by some EEG potentials named MRPs [15].

[3] The same potentials can be identified as Movement-related Cortical Potentials (MRCP) in the literature.

The fundamental component of these MRPs that contributes to the definition of a pattern in the time domain that is simple to "translate" into a command for the interactive application is the Bereitschaftspotential (BP).
The BP is a slow negative component that decreases progressively in amplitude starting from 2 s before the movement actuation [66]. The BP waveform is shown in Fig. 3.1.
The initial slow segment of BP (BP1) is called *early BP*. It begins ~2 s before the movement onset and originates in the pre-supplementary motor area (pre-SMA) with no site-specificity and the SMA proper according to the somatotopic organization, and shortly thereafter in the lateral premotor cortex bilaterally with relatively clear somatotopy [67, 68]. About 400 ms before the movement onset, the steeper negative slope, called *late BP* (BP2), occurs in the contralateral primary motor cortex area and lateral premotor cortex with precise somatotopy. These two phases of BP are differentially influenced by various factors, especially by complexity of the movement which enhances only the late BP. As shown in Fig. 3.1, BP deflection also appears during imaginary movements. BCIs based on this last approach is called motor imagery BCIs, but their use is out of the thesis scope.

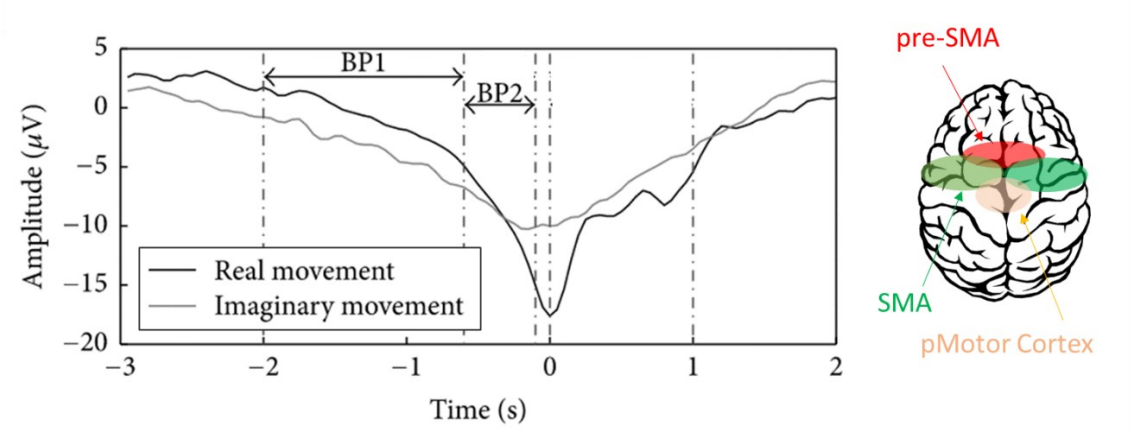

Figure 3.1 MRP waveform in the time domain and related brain topography. Picture extracted by [66] and modified.

3.1.2 BCI #2 Cortical Pattern: Event Related Potentials - P300

Event-related potentials (ERPs) are small voltages generated in the brain structures in response to specific events or stimuli. They are EEG changes that are time-locked to sensory, motor or cognitive events [69]. Event-related potentials can be elicited by a wide variety of sensory, cognitive, or motor events.
ERPs can be divided into two categories: early ERP and late ERP. The early waves, which peak roughly within the first 100 ms after stimulus, are termed ‘sensory’ or ‘exogenous’ as they depend largely on the physical parameters of the stimulus. While ERPs from later parts reflect the way the subject evaluates the stimulus. They are termed ‘cognitive’ or ‘endogenous’ ERPs as they examine information processing.
For the proposed application a single ERP component is considered due to its proved correlation with cognitive decline [70]: the P300.
P300 can be evoked using a clinical protocol known as the *oddball paradigm* in which a subject is actively and cognitively involved in the discrimination of low-probability-

occurrence target elements mixed with high-probability (or "standard") non-target elements [71, 72].
Fig. 3.2 shows the ERP waveform in the time domain and related brain topology. Specifically, Fig. 3.2 emphasizes a window in the time interval where the P300 component peak. It is interesting to see that according to waveforms from Fig. 3.2 [70], there is a difference in terms of amplitude and latency (temporal onset of the component) of the P300 among younger, older adults and people with a risk of cognitive decline.
For auditory/visual stimuli, the P300 latency range is 250-400 ms for most adult subjects between 20 and 70 years of age. The latency is usually interpreted as the speed of stimulus classification resulting from the discrimination of one event from another. Shorter latencies indicate superior mental performance relative to longer latencies. P300 amplitude seems to reflect stimulus information such that greater attention produces larger P300 waves [69]. Reduced P300 amplitude is an indicator of the broad neurobiological vulnerability that underlies cognitive and memory disorders. P300 wave spatially concentrates in the cortical midline with high incidence in the parietal lobe.

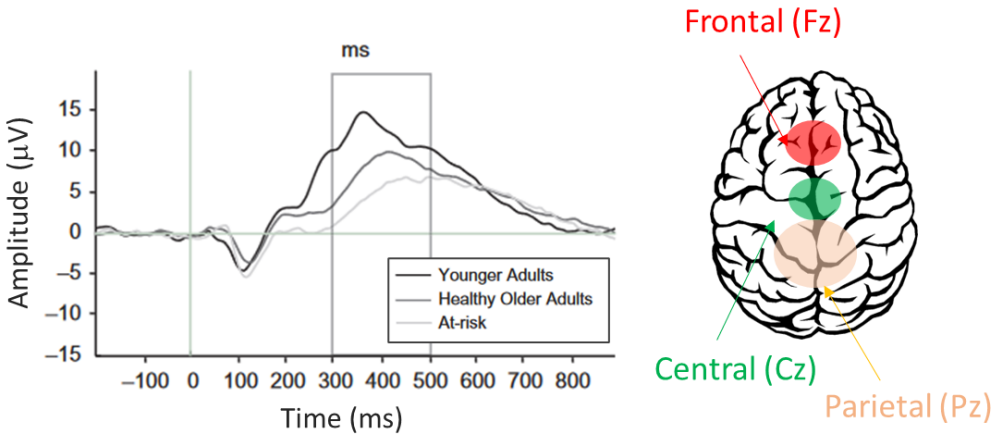

Figure 3.2 ERP waveform in the time domain and related brain topography. Picture extracted by [70] and modified.

3.1.3 Experimental Setup and Stimulation Protocols

3.1.3.1 Experimental Setup

As anticipated in Sec. 2.2, during the nominal operation of the system, the user wears a wireless EEG headset and sits, or is lying down, in front of a stimulation video terminal. This section will first provide details of the electrode sites selected for cortical activity monitoring according to the cortical topologies introduced in Sec. 3.1.1 and 3.1.2. Secondly, the stimulation protocols adopted for the two BCIs will be analyzed.
To identify the best channels to be monitored for MRP or P300 extraction purposes, a state-of-the-art analysis has been conducted through Google Scholar, IEEExplore databases. Databases have been queried with keywords “P300” “MRP” “Movement-related Cortical Potential”, “MRCP”, “BCI” and the contributions have been limited to the period 2018-2022. Ten recent studies per cortical pattern (MRP: [73-82] P300: [83-

92]) have been selected and statistically analyzed. All the nomenclature below refers to the 10/20 international EEG positioning system.
The experimental setup investigation led to the following result:

- For an application suitable for detecting the P300ERP component, the channels to be monitored are typically: Fz, Cz, C_3, C_4, Pz, P_3, P_4, Oz (Fig. 3.3a)
- For an application suitable for detecting the MRP for hand movements (see somatopy), the channels to be monitored are typically: T_7, C_3, CP_5, CP_1, CP_2, CP_6, C_4, T_8 (Fig. 3.3b)

In light of the above, it was chosen to use a common setup that considers 14 channels for the experimental phase: Fz, T_7, Cz, C_3, CP_5, CP_1, CP_2, CP_6, C_4, Pz, P_3, P_4, T_8, Oz. Both applications require the use of the AFz electrode as GND for a monopolar reading and the right ear lobe as the reference electrode (REF). The analog input range of ± 187.5 mV with 24-bit resolution, the sampling rate of 500 Hz and a preventive bandpass filtering between 0.5 Hz and 35 Hz (8^{th} Butterworth order) before transmission are also shared.

3.1.3.2 Stimulation Protocol

The use of two BCIs in the context of the AMICO infrastructure, each one with specific cortical responses, led to the need for designing and implementing two different stimulation protocols.
All the stimulation protocols below were made in Unity 2018.3.0f2 environment with C# coding.

MRP-based BCI. As reported in Sec. 3.1.1, several studies have shown that any voluntary movement is preceded by a routine of neural preparation (motor planning) that defines the muscle sequence to be implemented to perform the specific gesture [93].

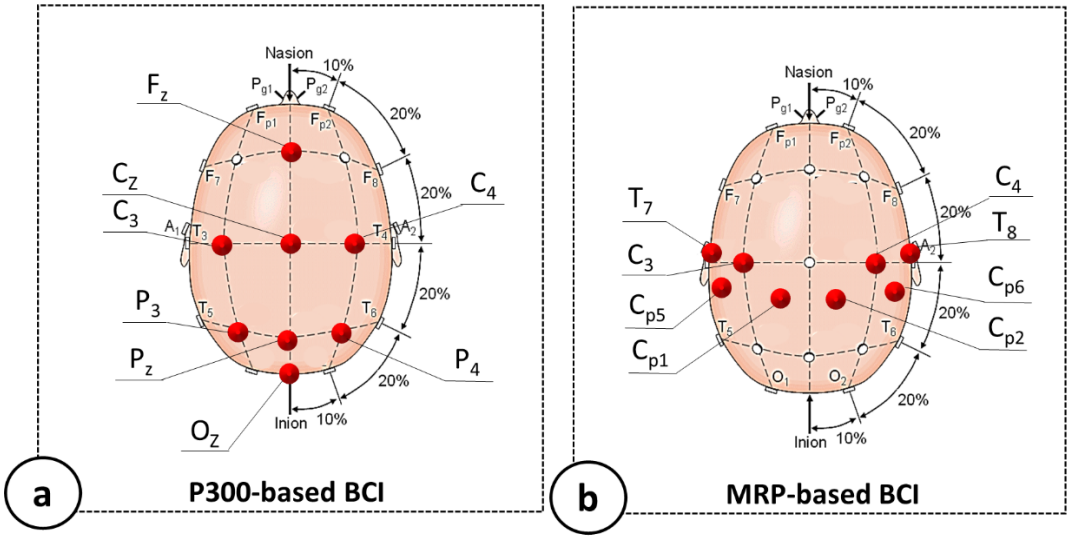

Figure 3.3 EEG headset channel selection for BCI applications setup. Reported electrode sites belong to 10/20 EEG positioning system.

For this reason, to emphasize the voluntariness of the movements on the modulation of brain activity, the user is asked to slightly move the index finger of the ipsilateral hand to the choice he/she intends to undertake.
To ensure repeatability, each movement must take place, e.g., following an. The acoustic signal has no stimulative purpose (and can be eliminated as will be illustrated in the specific preprocessing section). Following the movement, the related BCI will decode the EEG content providing as output the selected output.
The procedure described above is then repeated until the final choice is reached, through the use of nested choices (each of a binary nature) as described in Sec. 2.2.1.
Fig. 3.4 shows a screenshot of the stimulation protocol in a BCI service selection and a snapshot of real-life use.

P300-based BCI. As previously introduced, the P300 is an event-related potential, which must - therefore - be appropriately evoked at the sensory level (visual, auditory, or tactile). In this case, an experimental protocol for the elicitation of the P300 is the oddball paradigm. In general, an oddball paradigm consists of a sequence of stimuli that can be divided into two categories: rare or *target* stimuli and recurrent or *not target* stimuli. Precisely, the physiological condition of classification of a rare stimulus determines the onset of the potential P300 [12].
For this purpose, a stimulation protocol based on a visual paradigm known as the *matrix P300-speller* [94] has been created during a collaboration with g.tec neurotechnologies GmbH. The realized protocol is shown in Fig. 3.5. According to what is reported in Fig. 3.5, the proposed matrix protocol consists of 3 rows and 4 columns. Each row and column is randomly flashed once (without repetition) per run. Each run consists of 7 flashes that cover all the rows and columns involved.
For the specific application, there is an interstimulus time (ISI) of 500 ms (between two consecutive flashes). The flash lasts 200 ms, so the fully gray view lasts 300 ms.

Figure 3.4 MRP-based BCI stimulation protocol: example of selection interface and a snapshot of real-life use.

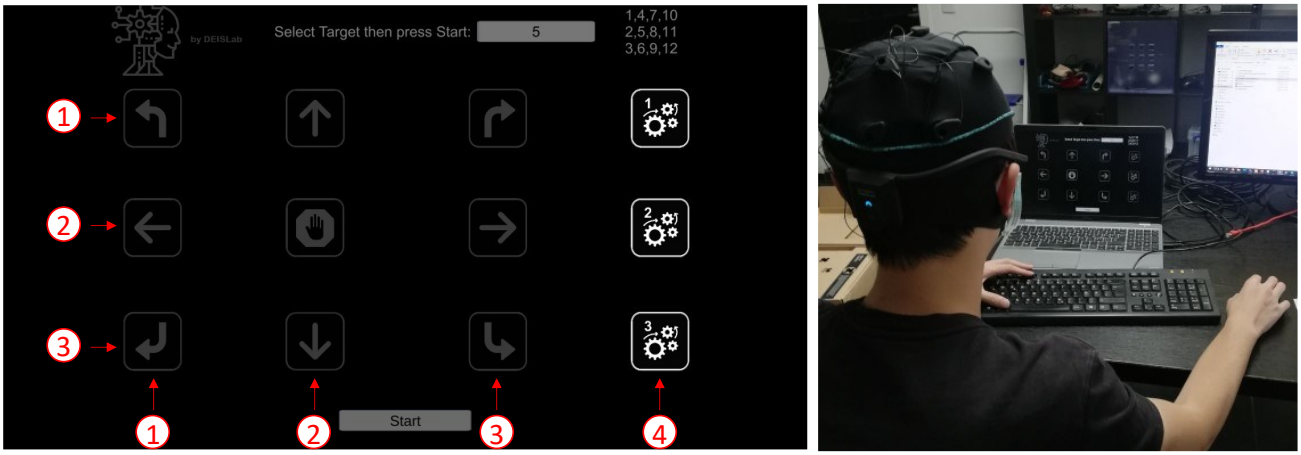

Figure 3.5 P300-based BCI stimulation protocol: example of selection interface and a snapshot of real-life use.

During the protocol, the subject is asked to focus on a specific element of the matrix. Each element is identified by a row and a column index. For instance, let us consider the subject is focusing on the element {1,2}, which is a straight arrow up arrow in Fig. 3.5.
In this case, when row 1 or column 2 (targets) flash, the P300 component will be detected at the cortical level. In all other cases (not target ones) this ERP component will not be present.
After each run, the related BCI will return a single-run classification. BCI predictions are then cumulated for 3-6 runs (10.5 s – 21 s) to provide the decoded choice.
It is necessary to specify that the matrix selection panel proposed in Fig. 3.5 derives from a previous choice that evolved the protocol in this direction. In fact, it presents arrows with directions linked to a choice of robot driving through BCI. To provide a complete overview of the infrastructure, an example of nesting of choices proposed by MRP-based BCI is presented in Pseudocode section 3.1, while a selection panel of the P300-based BCI is provided in Pseudocode section 3.2.

Pseudocode 3.1: MRP-based BCI choices example

- **Services**
  - **Bring me…**
    - Food
      - List of choices *(Robot-Environment connection)*
      - Request
        - ✓ 1, 2
        - ✓ 3, 4
    - Drink, Drugs
      - Drink
        - ✓ List of choices *(Robot-Environment conn.)*
        - ✓ Request
          1. 1,2,3
          2. 4,5,6,7
      - Drugs
        - ✓ List of choices *(Robot-Environment conn.)*
        - ✓ Request
          1. 1,2
          2. 3
  - **Call** …
    - Caregiver
    - Emergency

- **Entertainment**
  - **Music**
    - List of choices
    - Request
      - 1,2,3
      - 4,5
  - **Video**
    - List of choices
    - Request
      - 1,2,3
      - 4,5,6

Pseudocode 3.2: P300-based BCI choices example

- **List of Services**
- **Call Caregiver**
- **Call Emergency**
- **Bring Food**
  - List of choices *(Robot-Environment connection)*
  - Request 1
  - Request 2
  - …
  - Request 10, 11
    - Request 10
    - Request 11
- **Bring Drink**
  - List of choices *(Robot-Environment connection)*
  - Request 1
  - Request 2
  - …
  - Request 4
- **Bring Drugs**
  - List of choices *(Robot-Environment connection)*
  - Request 1
  - Request 2
- **Robot Driving**
  - x9 directions
  - Routine 1 (Manipulation support)
  - Routine 2 (Manipulation support)
  - Routine 3 (Manipulation support)
- **Room Information**
  - Room 1
  - Room 2
- **Conversational Task**
- **Entertainment**
  - Music
    - Request 1
    - Request 2
    - …
  - List of choices Music
  - Video
    - Request 1
    - Request 2
  - List of choices Video
  - Video call
  - Favorite Music 1
  - Favorite Video 1
  - …
- **Temperature Taking**
- **Blood Pressure Taking**

3.2 BCI Pre-processing

Data from acquisition hardware (i.e., g.Nautilus Research EEG headset) are streamed to a dedicated Android application (min SDK = API 22 Android 5.1 Lollipop) running on LG CNS Tablet on Pepper's chest. The application takes care of implementing the pre-processing stage on the received data and the real-time inference phase. The training of the used models is instead realized in dedicated environments.
The following section deals with defining the preprocessing phase carried out on the EEG signals acquired during the operation of the implemented BCIs.
Three specific pre-processing sections will be presented: the shared preprocessing operating on the input signals to both BCIs and two specific data processing subsystems, one per implemented BCI.

3.2.1 Shared Pre-processing

The signals from the acquisition hardware (i.e., g.Nautilus Research EEG headset) are streamed to a circular buffer of 2560 samples (about 5 s of EEG signal sampled at 500 Hz) instantiated locally to the robot.
The choice of a circular buffer of about 5 seconds is due to the time required for the detection of the anomalous ocular blink used for the generation of the Wake-Up signal (see Sec. 2.2).
The subset of data progressively present within the circular buffer is subjected to a functional block that deals with rejecting the most common physiological artifacts (eye movement, contraction of the jaw, cardiac artifacts, etc.) and non-physiological (50Hz power line, impedance shift, etc.) [95].
In this context, this functionality has been entrusted to an EEG signal reconstruction algorithm known as riemaniann Artifacts Subspace Reconstruction (rASR) [95, 96].
This method has been selected due to its ability to extract, in real-time, an EEG signal without the major known artifacts, isolating the independent components (ICs) that overlap to generate the raw EEG via an Independent Component Analysis (ICA) [95].
The delay introduced by the rASR algorithm, in relation to the parallel analysis of 14 EEG channels, required on average 12.84 ± 2.72 ms (experimental results on repeated tests) including conveying, rejection and return of independent components and signal of interest [96].
Among the ICs it is possible to extract, distinctly, the only contribution of some specific artifacts, such as the ocular ones of our interest.
Thus, the rASR has a dual purpose in the proposed architecture:

- remove known physiological and non-physiological artifacts from the EEG of interest, improving the performance of BCIs in terms of accuracy;
- extrapolate the ICs related to the eye artifacts, making it possible to isolate the contribution of the blinks and, consequently, improving the robustness of the Wake-Up signal generation system.

Once BCI is selected, the EEG transmission is streamed back to the chosen BCI (P300 or MRP). The EEGs sent to the BCIs are continuously subjected to the real-time artifact rejection phase carried out by the rASR algorithm. The above procedure is summarized in Fig. 3.6.

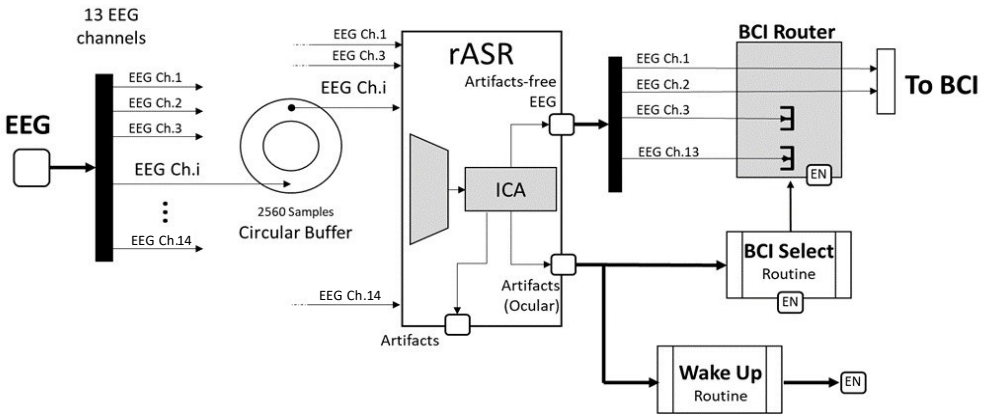

Figure 3.6 Representation of the shared pre-processing block.

3.2.2 MRP-based BCI Pre-processing

When enabled to operate by the related blinking protocol (see Sec. 2.2.1), the MRP-based BCI allows the streaming transfer of EEG data from 8 channels out of the 14 monitored (T_7, C 3, CP_5, CP_1, CP_2, CP_6, C_4, T_8), according to what is reported in Sec. 3.1.3.1.

Filtering. The MRP-based BCI does not require any additional filtering blocks functional to the analysis of motion-related potentials, because they operate in the band provided by the pre-filtering stage operated before the transmission: 0.5 Hz – 35 Hz (see Sec. 3.1.3.1).

The subset of EEG data from the rASR block is preemptively sent to a circular buffer of 640 samples (~1.25 s @ 500 Hz) and updated in a First-In-First-Out (FIFO) way at every new sample.

Trial extraction. Once the BCI is started, it provides an acoustic signal at each choice to be taken, according to what is reported in Sec. 3.1.3.2. This beep lasts ~500ms (i.e., > human reaction time). The falling edge of the command to sound the buzzer is used as a trigger of the proposed system.

Once the trigger signal is detected, the BCI waits for about 250 ms (128 samples) after which it enables the parallel transfer of EEG data from the circular buffer to a dedicated memory unit. This subset of EEG data consisting of 640 samples of which 512 samples are before the trigger and 128 following the same, performs the so-called *trial* (observation) of this type of BCI.

3.2.3 P300-based BCI Pre-processing

When enabled to operate by the related blinking protocol (see Sec. 2.2.1), the P300-based BCI allows the streaming transfer of EEG data from 8 channels out of 14 (Fz, Cz, C3, C4, Pz, P3, P4, Oz) to a filtering block useful for the analysis of event-related potentials.

Filtering. The implemented filtering stage consists of a 4th order Butterworth low-pass filter with cut-off frequency fc=10 Hz.
The subset of filtered and artifacts-free EEG data is preemptively sent to a circular buffer of 320 samples (~640 ms @ 500 Hz) and updated in a First-In-First-Out (FIFO) way at every new sample.

Trial Extraction. The resulting EEG signal is then subjected to a synchronization phase with the visual stimulus (see Sec. 3.1.3.2). This step consists in extrapolating, from the streaming of EEG signals, a subset of data for the duration of 600 ms starting from the rising edge of the stimulation signal generated by the visual protocol and 40 ms prior to it. This data block is the trial for the relevant BCI.

Baseline Removal. A baseline signal is extracted as the average of the values assumed by the EEG in the range [-40 ms, 0 ms((20 samples). This average value is then subtracted from the entire trial.

Downsampling. The resulting waveform is downsampled from 500 sps to 20 sps, using a DIF algorithm (Downsampling by Integer Factor – F = 25). It results in a total of 13 samples per EEG channel for each observation.

Flattening. The resulting matrix of 8 channels x 13 samples is flattened on a single row, resulting in a feature vector composed of 104 features.

3.3 MRP-based BCI

To date, the cortical pattern realized by MRP has found few applications in the field of selective BCIs [73-82]. Such cortical signals are, in fact, often used to support other types of primary BCIs, for example, those based on sensory-motor rhythms [76] to improve their accuracy. Another wide use of these cortical signals concerns BCI applications for rehabilitation [96], to study gait initialization phases in patients with coordination disorders [97], or to analyze movement intention and voluntariness [93, 98, 99].
In this context, a first-of-a-kind MRP-based BCI application for service selection has been realized and implemented in the AMICO infrastructure and is here proposed.

3.3.1 Local Binary Pattern-based MRP decoding

Fig. 3.7 shows the working principle of the implemented MRP-based BCI. The figure shows the phases of cortical pattern recognition during a real-time classification, neglecting those strictly related to the training phase conducted offline.
Although no mention in Fig.3.7 is reserved for the offline training stage, most of the processes in the block diagram are shared between the two branches of computation (offline training and real-time classification).
As can be deduced from the principle diagram in Fig. 3.7, the proposed algorithm operates on two main branches: (i) the off-line training process that calibrates the system, extracting *reference masks* and training a *classifier based on Support Vector Machine* (SVM) and (ii) the recognition and classification of the cortical pattern in real-time, which deals with comparing the observations to be classified with the reference masks extracted during offline learning.

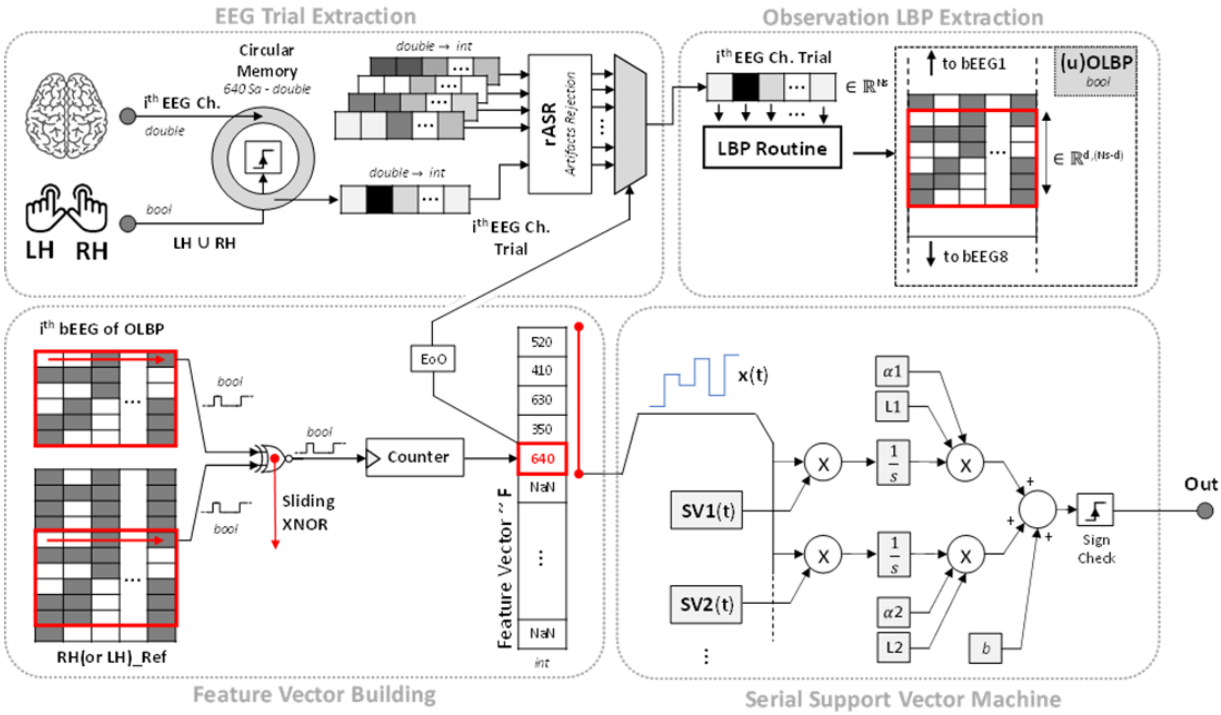

Figure 3.7 Overview of the real-time cortical pattern recognition system for the MRP-based BCI. The EEG Trial extraction block details the extraction of the trial only for an i-th EEG channel

3.3.1.1 Offline Learning

Data preparation. The offline learning phase implemented for MRP-based BCI consists of four main phases: (i) the identification of the average cortical behavior, for each EEG channel, in relation to the movements of the left hand (LH) and right hand (RH); (ii) the extraction of reference masks through a routine based on the Local Binary Patterning (LBP) method; (iii) the feature vector extraction and (iv) SVM classifier training.

The first step of the supervised learning process is to label each trial, extracted according to the guidelines provided in Sec. 3.1.3.2, with the movement that generated the cortical modulation, i.e., LH or RH movements.

EEG data from left- and right-hand movements are then stored in two dedicated memory spaces. The two resulting matrices have the same conformation and consist of 3D data structures, $\mathbf{Tr} \in \mathbb{R}^{S,N,M}$ where S is the number of samples that compose a single trial, N is the number of monitored channels and M is the number of observations composing the training dataset, or equivalently, of movements of the user's hands (right or left). For the specific application, S=640 samples, N=8 channels according to Sec. 3.1.1 and M varying between 80 and 100 trials per user.

Extraction of mean cortical responses. The 3D data structures **Tr** from left and right-hand movements are singularly compressed into two dimensions matrices by computing, for each EEG channel, the average waveform on a part of collected observations composing the training dataset (i.e., 40% of the training dataset is used for the mean cortical responses extraction).

This last step produces two matrices: a first matrix containing the average EEG signals from the 8 monitored channels, extracted at the movements of the left hand, **LH** $\in \mathbb{R}^{N,S}$, and a second matrix containing the average EEG signals from the 8 monitored channels, extracted at the movements of the right hand, **RH** $\in \mathbb{R}^{N,S}$**.**
By their own definition, the rows of these matrices represent the average behavior of each individual EEG channel at the movement of a specific hand. This approach is made possible by the marked repeatability of the acquired waveforms, which maintain a typically phased structure for all observations on which the average operation is performed.

Local Binary Patterning. During the second phase of the learning process, each row from the **LH** and **RH** matrices is numerically treated using a data analysis routine typical of image processing. This procedure, called Local Binary Patterning (LBP) is a particular methodology belonging to the category of symbolization algorithms. The purpose of these algorithms is to transform experimental measurements, typically multidimensional, into two-dimensional data structures composed of binary strings. [100].
The use of end-to-end binary operations leads to a reduction in the computational complexity of the algorithm following the symbolization stage (in this case the Feature Extraction phase) and to a consequent speeding up of the classification stage in real-time. Specifically, in this context, the LBP deals with encoding a one-dimensional time series (EEG signal), typically double-precision, in a two-dimensional binary matrix according to a differential analysis between two contiguous values of the EEG signal [101].
In particular, the here implemented LBP procedure uses a d = 6-bit encoding, according to the following relationship:

$$\begin{cases} bEEG = 1 & \leftrightarrow EEG(ch, i+1) - EEG(ch, i) > 0 \\ bEEG = 0 & \leftrightarrow \quad otherwise \end{cases} \qquad (1)$$
$$\mathrm{i} = 1{:}\,\mathrm{d} - 1, \mathrm{ch} = 1{:}\,\mathrm{N}$$

where $EEG(ch, i)$ is the i-th sample of the average EEG trial extracted on the ch-th channel, while N represents - according to common notation - the number of monitored channels.
The LBP routine is applied to all 8 channels for both **LH** and **RH** matrices, resulting in two *reference masks* named **LH Ref** and **RH Ref** respectively. Both reference masks have dimensions {d*N, S} [102].
Fig. 3.8.a shows an example of applying the LBP routine to a one-dimensional time series (e.g. an EEG). Fig. 3.8.b shows the binary structure of the **LH Ref** and **RH Ref** reference masks. In Fig. 3.8 the elements of the matrix that have yellow color identify a logical 1, while the logical 0 is identified by the blue color.

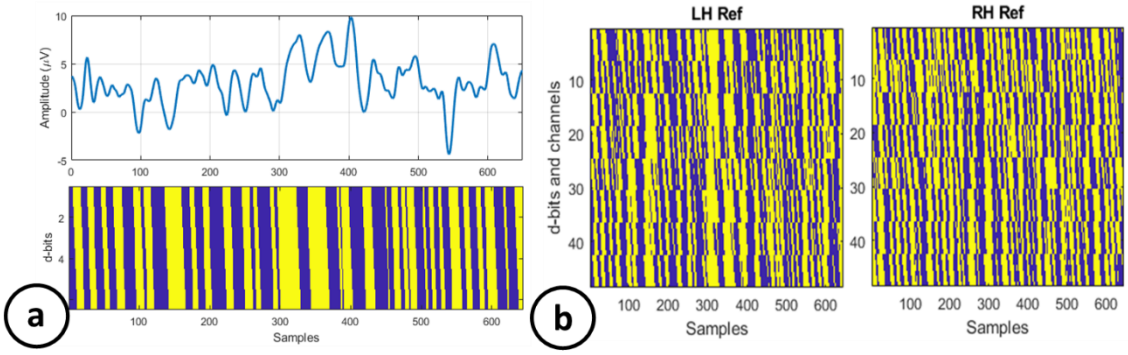

Figure 3.8 Routine LBP. (a) Esempio di codifica LBP per un singolo trial (non medio); (b) Esempio di maschere di riferimento **LH Ref** e **RH Ref**

Feature Vectors Extraction. During the third phase of the learning process, all the trials acquired and not involved in the extraction of the reference masks (**LH Ref** and **RH Ref**), are treated individually by the LBP routine according to the eq. (1). The procedure is identified in Fig. 3.7 by the block called Observation LBP Extraction.
The output of this phase is a two-dimensional binary array, called an Observation Local Binary Pattern (**OLBP**), which shares the same dimensions as the reference masks.
In the following, I will refer to **OLBP** for LBP matrices extracted from supervised observations in the learning phase, while I will use **uOLBP**, where the character 'u' stands for unlabeled, for LBP matrices extracted from observations to be classified (therefore still unlabeled).
Each element of the **OLBP** matrix is then compared, via a plane of XNOR ports, with the corresponding element of the masks: **LH Ref** and **RH Ref**.
This comparison step results in two matrices that provide information about the degree of similarity between the single **OLBP** observation and each of the **LH** (or **RH**) **Ref** reference masks.
The last step is to count the number of "1s" on each row of the comparison matrices.
The final feature vector must take into account the degree of similarity of the **OLBP** with both **LH Ref** and **RH Ref**, for this reason, it was decided to rearrange the feature vector as follows:

$$F = \begin{bmatrix} @1^{st}\,\text{row}: \sum_{j=1}^{Ns} \text{uOLBP}(1,j) \,\overline{\oplus}\, \text{LH Ref}(1,j) \\ \dots \\ @48^{th}\,\text{row}: \sum_{j=1}^{Ns} \text{uOLBP}(48,j) \,\overline{\oplus}\, \text{LH Ref}(48,j) \\ @49^{th}\,\text{row}: \sum_{j=1}^{Ns} \text{uOLBP}(1,j) \,\overline{\oplus}\, \text{RH Ref}(1,j) \\ \dots \\ @96^{th}\,\text{row}: \sum_{j=1}^{Ns} \text{uOLBP}(48,j) \,\overline{\oplus}\, \text{RH Ref}(48,j) \end{bmatrix} \quad (2)$$

The conformation of the feature vector as reported by eq. (2) allows to obtain immediate discrimination in the subspace of the features allowing the application of SVM to the linear kernel.
Indeed, if an **OLBP** is coming from a movement of the left hand, it will turn out that all features from 1 to 48 will have a high value (high number of 1 in each row), while features

from 49 to 96 will have a low average number of 1, making the logical classification immediate.
Finally, the vector **F** is defined by eq. (2) is used to train a linear kernel-based SVM.
More details about the LBP routine and the employed SVM, as well as their implementation in the AAL framework, are provided in our previous work [101-105].

3.3.2 Real-time Classification

At present, during the real-time classification phase, the **LH** Ref, **RH Ref** reference masks, and parameters of the trained SVM model are stored in static memory allocations into the reference app.
During the online classification phase, the EEG trials coming from the pre-processing system reported in Sec. 3.2.2 and triggered by a movement to be classified are subjected to the **uOLBP** creation phase according to what is reported in the previous paragraph.
The uOLBP is then subjected to the Feature Extraction step that compares the **uOLBP** to the reference masks according to eq. (2). The resulting vector is then forwarded to the linear kernel SVM [105] for trial classification.
Once the inference is over, the app formalizes a unique command for the robotic platform through a dedicated NAOqi API. It will start on the robot the requested routine according to the use cases proposed in Sec. 2.2.

3.4 P300-based BCI

To date, P300-based BCIs are the most widespread neural interfaces. The reason lies in the simplicity of elicitation of the P300 component through the oddball paradigm and in-depth knowledge of the nature of this physiological deflection [106].
Despite these advantages, P300 presents a very low signal-to-noise ratio (SNR), making difficult its recognition in individual EEG trials (single-trial). The common practice to address the problem in most state-of-the-art P300-based BCIs is to take advantage of the event-locked nature of the P300 component. In this respect, most solutions take a progressive average approach by accumulating trials. The effect of the average on a trial collection is to emphasize only the components in phase (ERP), reducing the EEG background noise. Nevertheless, this approach slows down the BCIs because several trials need to be buffered (15-50 target trials on average).
In the recent past, a growing interest has been dedicated to deep learning applications in P300 detection. Despite deep learning approaches strongly impacting natural language and computer vision processing, allowing the resolution of up to that moment still unsolved real-world problems, Neural Network (NN)-based application for P300 detection is still in an evolving stage, far from a consolidated methodology [107].
The first authors who exploited the deep learning capability of capturing hierarchical features from EEG signals were Cecotti and Graser in their work [108]. They implemented a four-layer convolutional neural network (CNN) to capture spatial and temporal features from the raw EEG signals in input. Their architecture was able to reach an accuracy of 25.5% on a single trial and ~95% after 15 repetitions in a standard 6x6

speller problem [94]. This approach was exploited by the authors in [107], which extended the model by introducing a batch normalization layer to prevent CNN overfitting. The proposed method, named BN3, analyzed EEG data as a bidimensional matrix with the {channels × time} shape, acting on it as an image structure and reaching the same results as the authors in [109] after 15 repetitions, with an improvement of +10% of accuracy on a single repetition. A similar approach was adopted by the authors in [110], who—differently from [107]—analyzed an EEG trial as a 3D matrix. Their work exploits spectral features from three different frequency bands (i.e., θ, α, β) creating from each time instant an RGB image given by the linked spectrogram. Then, they stacked convolutional layers with a long short-term memory (LSTM) neural network, increasing the system complexity without any substantial improvements against the above-listed state-of-the-art results. Similar results were achieved in [111], where a 3D recurrent neural network (3DRNN) was used to detect a P300 signal in a single trial. Finally, authors in [112] proposed the introduction of a preprocessing stage based on principal component analysis (PCA) to improve the CNN results. The proposed system demonstrated a slight improvement in the results, starting from 8 repetitions up to 15.
Ultimately, these approaches have provided promising results in the context of single-trial or low-trial classification. However, almost all the proposed NN-based solutions exploit large input data spaces and architectures with a high number of parameters (not suitable for devices with low computational resources or embedded systems).
To bridge these gaps, in the context of a collaboration between the Design of Electronic Integrated System Lab of the Politecnico di Bari and g.tec medical engineering GmbH, a framework that allows the automatic identification of a user-tailored Dense Neural Network with low complexity (control over the number of maximum parameters) has been created. The NN considers a low number of input channels (i.e., 8 versus 64 of solutions at the state of the art), compatible with most of the low-cost EEG acquisition devices available on the market. A further advantage of the NN extracted through the framework implemented is the robustness to acquisitions in an uncontrolled (noisy) environment, however, the explanation of this advantage is beyond the objectives of the present thesis work.

3.4.1 Dense NN Extraction Framework

The EEG data pre-processed according to what is reported in Sec. 3.2.3 are organized during the training phase in a 2D matrix $\mathbf{Tr} \in \mathbb{R}^{N_{obs}, N_f}$, where is the number of N_{obs} observations (trial) foreseen for the classifier training phase, N_f is the number of features that are intended to be submitted to the classifier. In accordance with Sec. 3.2.3, for this application. $N_f = 104$. The features correspond to the flattening of downsampled signal from the 8 monitored channels.
The **Tr** matrix is then used as a training dataset for the framework that identifies the best user-tailored NN topology.

The topology selector has been realized by means of Python and specifically via Keras API [113] with Tensorflow backend. For the topology selector definition, the Keras Tuner toolkit [114] has been employed.
The framework operates *offline*, computing data extracted and pre-processed via Android App according to Sec. 3.2. The result of the topology selection (i.e., the best NN model) is then extracted with the Tensorflow lite extension (*.tflite*).
Fig. 3.9 provides a schematic overview of the topology selector working principle.
From Fig. 3.9 it is possible to notice that the core of the proposed system is a model training function with relative validation. In this context, validation is conducted by means of a k-fold validation approach with k=3 (or k=4 if the dataset exceeds 1500 target+not target observations). The extraction of the weights of the fully connected NN by the training process and its validation is iterated, in the context of hyperparameter search, using the Hyperband research algorithm proposed in [115].
The compiling settings for the model training block include:

- Binary cross-entropy loss function.
- Nadam (Adam with Nesterov momentum) gradient descendent optimizer. The optimizer learning rate was set to 10^{-4} with no decay, while its parameters β1, β2, and ε were set, respectively, to 0.9, 0.999, and 10^{-8} as per [106].
- The batch size for the stochastic gradient descendent was set to 64.

The objective of the implemented Hyperband research algorithm has been defined as the minimization of the average loss function on all the considered validation batches (Average Validation Loss in Fig. 3.9).

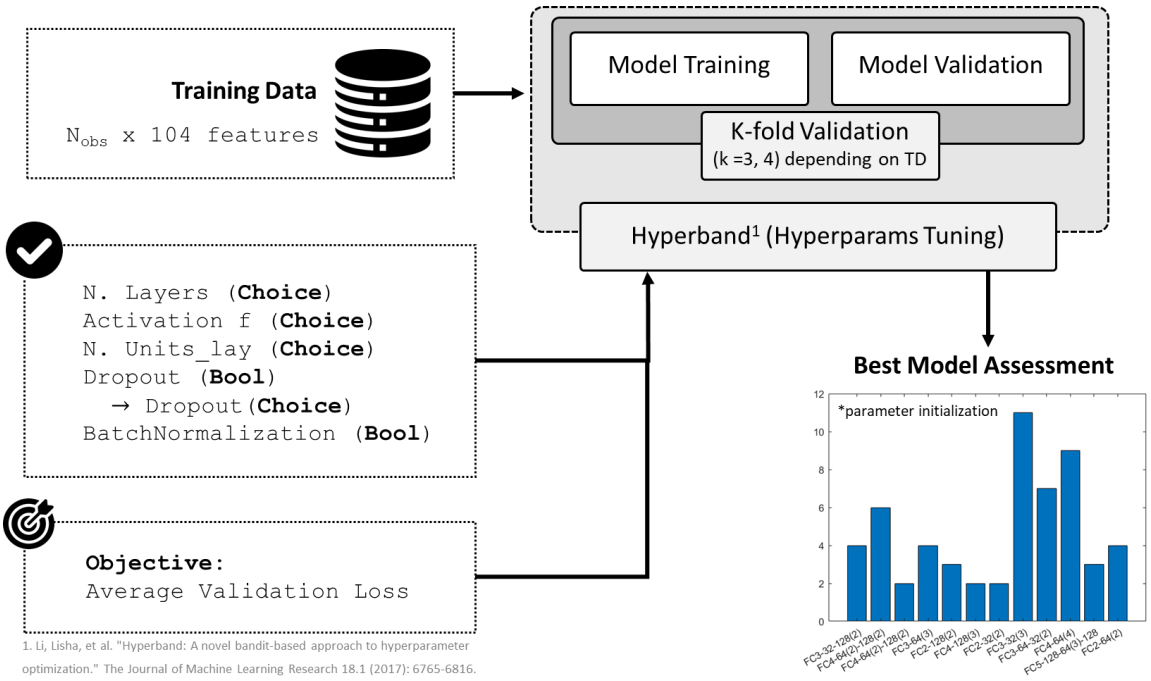

Figure 3.9 NN topology selector based on Hyperband approach for hyperparameters tuning.

In order to reduce the inference and actuation times of the developed Android application[4], the Tuner has been set with project constraints related to the complexity of the model itself.
To date, for ease of implementation, only fully-connected NN topologies have been considered in the tuning model. Starting from this limitation, the tuner analyzes – through a Hyperband approach – the impact on a specific metric (Average Validation Loss) of the following parameters:

1. *Number of hidden layers* constituting the dense NN (Fig 3.10 – Step 1). The maximum number of allowed hidden layers is 4. This limitation is due to an arbitrary constraint on the number of NN parameters that can be used, considering the chosen number of units per layer parameter.
2. *Number of units per hidden layer* (Fig. 3.10 – Step 3). The number of units per layer is limited to a maximum of 256 (see parameter 1) and provides a choice with steps that respect the powers of 2 to speed up NN on CPUs by leveraging SIMD instructions[116].
3. *Activation function per layer* (Fig. 3.10 – Step 4). Two different activation function are analyzed by the Tuner: the Rectified Linear Unit (*ReLU*) and the *Tanh* [117].
4. *Dropout layer presence and dropout rate* (Fig. 3.10 – Step 4). This tuning step analyzes the effect of a dropout layer in correspondence of different hidden layers. The dropout rate can range between 0.1 and 0.5 with a step of 0.05.

The NN incorporates some fixed layers (i.e., not subject to Tuner parameterization). These layers are the two input layers and an output layer. The two input layers consist of an Input layer, used to derive the size of the input data, and a second Batch Normalization layer. The latter implements a normalization phase of the incoming data based on the average and standard deviation values, channel by channel, from the training dataset.
The output layer is a single-unit dense layer with *Sigmoid* activation function. This layer outputs a number between 0 and 1 with a binary discrimination threshold set to 0.5.
Exceeding this threshold means that a P300 waveform has been detected in the analyzed single trial (labeled as a target stimulus trial), while not exceeding indicates that the analyzed trial belongs to a not-P300 category (labeled as a non-target stimulus trial).
The proposed tuner exports the best model at the end of the tuning process. The initialization parameters of the extracted model are saved, the weights are reset, and the tuning process is repeated several times to create a best models occurrence function (histogram on the bottom right in Fig. 3.9). Only the best model among these is selected for the final training. For best models are intended those models with the highest occurrence in the iterated process.

[4] Another reason lies in the future perspective of moving the inference algorithm on portable devices based on microcontroller to be directly connected to the robot.

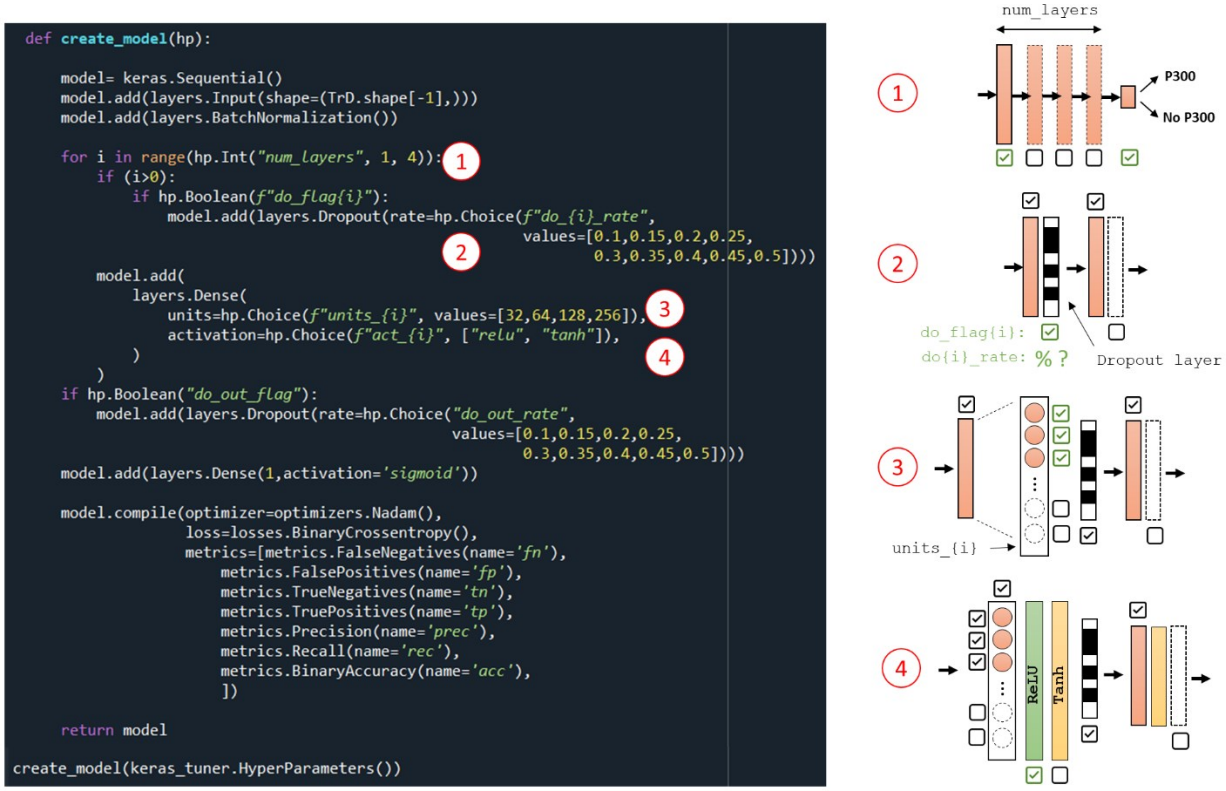

Figure 3.10 Demonstrative Keras Tuner code snippet with the graphical explanation of main steps.

3.4.2 Real-time Classification

As described in Sec. 3.4.1, the best NN from the topology selection is extracted as a Tensorflow lite model. These models can be imported as assets in the Android environment. Using protobuffer is then possible to exploit the extracted NN as a real-time classifier integrated into the app by creating a dedicated inference function.

During the real-time classification working mode, the Android app that implements the classifier, receives input data from the EEG acquisition system, and pre-processes them according to what has been seen in Sec. 3.2.3, thus, using them as feed for the imported NN.

After the 7 flashes composing the run, the app returns the values provided by the NN output layer. A value per each row and column is provided.

The system extracts the best row and best column, identifying an element of the 3x4 matrix (see stimulation protocol at Sec. 3.1.3.2). The values provided by the output layer are then saved in memory and incremented progressively with each run.

The real-time classification phase is to be considered completed only if for 3 consecutive runs the element chosen as the highest score does not change.

Once the inference is over, the app formalizes a unique command for the robotic platform through a dedicated NAOqi API. It will start on the robot with the requested routine.

3.5 P300-based Cognitive Status Extraction

Nowadays, the P300 analysis is a widely used diagnostic tool for diagnosing and monitoring neuro-degenerative pathologies [118]. As we have already seen in Fig. 3.2 of Sec. 3.1.2, it has been widely demonstrated that the latency of the P300 component and its amplitude is related to phases of cognitive decline [70, 118].
The main problems related to the P300 extraction methods for clinical purposes are that: (i) they are performed only in specialized centers scattered only in large urban centers; (ii) the protocol to derive the P300 is time-consuming (> 10 min) and when performed by electroencephalography (EEG) it requires the processing of at least 16 EEG channels.
Many solutions have been already proposed in the literature [83-92] for a correct and fast P300 extraction and detection starting from EEG raw data, in particular in BCI applications. However, since all these methods [83-92] are based on machine learning algorithms, they are not directly suitable as a diagnostic tool since their aim is only the detection of the P300 pattern, but they are not oriented to the P300 characterization in terms of amplitude, latency and involved brain area. For this reason, the most used approach to measure and characterize the P300 in the clinical environment still remains the Independent Component Analysis (ICA) [119], and the 'grand average'. Nevertheless, these approaches start their computations from some "a priori" assumptions that very often are not at all verified and valid for the P300.
In this context, exploiting the forced use (for other purposes) of the P300-based BCI in the AMICO framework, it is possible to collect in parallel many observations containing the P300 waveform. This collection, if properly analyzed, can allow a *remote* spatio-temporal characterization of the ERP component, to effectively propose a Figure of Merit (FoM) useful to support a cognitive follow-up. In this respect, this section describes an EEG reconstruction method to improve the P300 signal-to-noise ratio that is able to provide a clinically recognized FoM.

3.5.1 P300 as Biomarker for Cognitive Impairment Diagnosis

P300 latency and amplitude reflect the degree of cognitive decline in dementing illness [118]. A single P300 pulse is anticipated by further ERPs (i.e. P100, N100, P200, and N200 – see Fig. 3.2) which classify the cognitive process. The P300 characterization as a biomarker for cognitive impairment is based on the simultaneous evaluation of amplitude and latency. For this aim, the proposed FoM can be defined as [120]:

$$\frac{V_{pk-to-pk}(P300-N200)}{Latency}\left[\frac{\mu V}{ms}\right] \tag{3}$$

where the relative peak distance between P300 and N200 potentials is considered. According to [120, 121], it is possible to extract threshold values for amplitude and latency for 26 healthy subjects (aged 64.9 ± 10.9 years): the P300 amplitude should be >5.3μV, while the latency < 349ms [120]. From these clinical values, it is possible to estimate that for healthy subjects the FoM is >0.01 μV /ms. FoM, ranging between

	Amplitude [μV]	Latency [ms]	FoM [μV/ms]
Healthy	>5.3	< 349	≥ 0.01
MCI[1]	1.4 < A < 3.1	>389	0.003≤FoM≤0.008
HCI[2]	< 1.4	>400	FoM≤0.003
[1] Mild Cognitive Impairment; [2] Heavy Cognitive Impairment			

Figure 3.11 P300 clinical reference for diagnosis. Picture extracted by [120]

0.008μV/ms < FoM < 0.01 μV /ms.. Fig. 3.11 reports threshold values to discriminate Mild Cognitive Impaired patients and heavily cognitive impaired ones [120, 121].

3.5.2 P300 Spatio-Temporal Characterization Method

The use of conventional ERP averaging (grand average), although it is the most widespread method in clinical practice, is inappropriate since the intrinsic variability of the ERP leads to distortions of latencies, reduction in maximum amplitude (peak) and a broadening of the component.

Residue Iteration DEcomposition. The method used for the proposed application is a tuned version of the residue iteration decomposition (RIDE) which is a hybrid approach based on linear superposition and iterative residual calculation. The RIDE approach allows to detect spatio-temporal ERP characteristics with no limitations in terms of number of electrodes and number of target stimuli [120].
RIDE considers a linear superposition model of single-trial ERPs. Except for the noise ε, the single-trial EEG is decomposed into two components: stimulus-locked (S) and cognitive-locked (C) components. A single-trial EEG, including EEG background activity and noise, can be expressed as:

$$EEG_i(t) = S(t) + C(t + \tau_i) + \epsilon(t) \tag{4}$$

Where τ_i is the latency of component C in the i-th trial and is characterized by a distribution $\rho(t)$ assumed to be Gaussian (but this is not a limitation hypothesis). A conventional average over N trials would result in:

$$\begin{aligned} \langle ERP \rangle &= S(t) + \frac{1}{N}\sum_{i}^{N} C(t + \tau_i) + \frac{\epsilon(t)}{\sqrt{N}} = S(t) + \int C(t + \tau_i)\rho(\tau)d\tau + \frac{\epsilon(t)}{\sqrt{N}} \\ &= S(t) + C * \rho + \frac{\epsilon(t)}{\sqrt{N}} \end{aligned} \tag{5}$$

Eq. 5 shows that, although noise is reduced, the average creates a broadening of the C component which is convolved with its distribution. Neglecting ε, it is possible to consider the residues in a single-trial:

$$\boldsymbol{Res_i(t) = EEG_i(t) - \langle ERP \rangle = C(t + \tau_i) - C * \rho} \tag{6}$$

If the residues are aligned to their τ_i through cross-correlation jitter-latency estimation and averaged again, the distortion is reduced and the first estimation of C is computed as:

$$\boldsymbol{C_1(t) =< Res >= C(t) - (C * \rho) * \rho} \quad (7)$$

By replacing C_1 in eq. (5), it is possible to obtain a first estimation of S_1. The procedure is then iterated using a first ERP estimation [$ERP_1=S_1(t)+C_1(t+\tau_1)$] leading at the end to a more precise S and C estimation. After the n-th iteration, the components C_n is given by:

$$C_n(\boldsymbol{t}) = C - C * \rho_0 * \rho_1 \ldots * \rho_n \rightarrow C \quad (8)$$

After n iterations, C_n and S_n converge to C and S since the iterative convolution by ρ approaches zero. Differently from similar iterative methods (i.e. Takeda et al. [122]), the RIDE method does not introduce systematic artifacts and its convergence is fast (≈ 10 iterations). The RIDE algorithm has been tested for different trends of ρ and it has been verified to be robust and accurate [120]. RIDE advantages comprise the low number of target stimuli, few electrodes needed, good accuracy, and information regarding single-trial. For the AAL framework application the tuned version of RIDE, the t-RIDE, proposed in our previous work [120] has been employed.

Tuning of RIDE. The t-RIDE algorithm is made up of two phases: (i) window optimization and (ii) results extraction. RIDE algorithm is a generic approach for ERP extraction, but it needs to be tuned for P300 calculation. To reduce the computational effort and since it is not known 'a priori' the source of P300, at first only one signal derived from the average of Pz and Cz is considered for the window optimization. A first default rectangular window is set to 250 ms – 400ms after the target stimulus according to clinical reference [70]. The first window sizing can be remotely customized by the physician. A starting estimation of the latency of C component is performed using the Woody's Method and a first characterization of the P300. Based on template matching, the cross-correlation between the template of P300 and a single trial EEG is performed, and residuals are calculated separating the S and C components. The procedure is iterated until the latency C in single trials stops changing monotonically. When the C latency convergence is reached, the results are stored, and the procedure is iterated again with a different windowed EEG signal. The start of the window is iteratively 4ms right shifted while the end of the window performs 8ms right shift. Seven different windows are considered in order to cover the full-time range in which the P300 can occur: the last computation is performed on the rectangular window of 278 – 456ms after the target stimulus. At each iteration, the evaluated P300 maximum amplitude related to the particularly used window is stored. At the end of the window definition cycle, the window that has led to the highest P300 peak is considered the optimized window. After the window optimization phase, the procedure for results extraction is performed on all the pre-processed channels. The results extraction phase involves the application of the RIDE method on the optimized window previously computed.

For each channel, the P300 is totally reconstructed in the window hooking the S component to the rising edge of the target stimulus, while the C component is appended to S using the estimated value of its latency.

For each channel, information about latency and peak are presented. The presented approach needs to estimate the latency of the C component (which coincides with the P300) for each single-trial so information about latency and peak variation trial-by-trial can be estimated. A statistical analysis of the data can be sent to the physician with the medium latency, peak and FoM.
In the AMICO context, the t-RIDE algorithm activates only after the fulfilment of the user's request as described in Sec. 2.2.3. Before passing to Sleep status, the robot runs the t-RIDE computation on the collected waveforms. Then starts sending the characterizing outputs of the P300 to the FHIR server to enrich the telemedicine service.
The t-RIDE outputs consist of:

1. The time-domain waveforms of target stimuli compared to not-target for each channel;
2. Information about P300 generation site and propagation (topography);
3. A flag for the presence/absence of P300, peak values, latency values, and FoM for each channel.

It should be specified that the software does not express any formal diagnosis, which will be performed by the specialized physician that operates remotely.
More details about the t-RIDE method and implementation are provided in our previous works [72, 120, 123, 124].

3.6 EEG-based Emotive Status Extraction

Emotions are the main engine that drives humans' capability in engaging and interacting with the social world [125]. In this context, despite the deep knowledge of human emotions is still in an early stage under the neurophysiological aspect (due to the lack in a systematic method for emotion recognition), several attempts in making machines able to understand basic human emotions have been documented in the last decades [125]. These latter attempts of improving human-machine interactions (HMI) constitute the affective computing field. Identifying emotion patterns is critical in several fields such as healthcare, robotics, medicine, education and entertainment. Specifically, concerning recent healthcare applications, emotion recognition found applications in the diagnosis and the treatment of mood disorders like depression, perinatal mood and anxiety disorders, sleep disorders, as well as the social disorders identification during the COVID-19 pandemic [126].
In the assistive context, emotion recognition has been employed in emotion-aware robotics platform for domestic rehabilitation, pharmacologic compliance assessment, and pervasive recommending systems for telemedicine [127].
Given the large spectrum of application of emotion recognition systems, recently, the scientific community started employing smart technologies and high-performance computing to automatize, in a reliable way, this recognition process [125]. The first step in this direction was finding a proper quantitative method to identify and characterize basic emotions. In this respect, the widely used method in the common practice is the

circumplex model described in Sec. 2.2.4. Briefly, the circumplex model permits to characterize emotions starting from three psychometric parameters: Arousal, Valence and Dominance.
A second step towards the automatic recognition was the identification of proper emotion indices. They range from physiological signals (such as EEG, electrocardiogram (ECG), and galvanic skin response (GSR)) to external indicators (such as facial expressions, voice changes, postures and body motion).
Without prejudice to the generality of the method proposed therein, this brief overview focus on EEG signals only as system input, because of EEG availability in AMICO infrastructure.
A wide variety of feature extraction approaches are used at the state of the art to characterize emotions. These approaches range among time [128], [129], frequency [130], [131], [129], time-frequency [132], [131], [133] and spatial domains [134]. Currently, there is not preferable approach for the specific application, but features based on the frequency domain remain the most used [132], [131], [133]. Concerning the employed classifiers, the most recurrent classifiers are the Random Forest (RF) [133], [129] and the Support Vector Machines (SVMs) [132], [128]. Nevertheless, NNs are also positively impacting the emotion recognition field [131].
Although the above-presented emotion recognition systems are increasingly becoming accurate, they are designed to run on laptop PCs with high performance. No focus is typically reserved for the complexity of these models, as well as for their applicability in wearable, portable, and real-time operating systems for real-life scenarios. Complexity control also permits to the creation of routines that operate without the need for an internet connection and online operative frameworks, for example, in domestic assistive robotics. In this latter case, a low-complexity application running offline in the background allows to protect sensitive data like patient history and home maps, usually stored for navigation. To bridge this gap, a framework that permits to extract an emotion recognition model with an automatic "*accuracy versus complexity*" characterization has been realized in the AMICO project context.

3.6.1 Emotion Recognition Architecture

Fig. 3.12 provides an overview of the proposed emotion recognition architecture using block diagram representation.
The working principle of the proposed architecture can be briefly summarized as follows. The proposed framework. shown in Fig. 3.12, operates with two main modalities: offline and online. At the first use, the overall architecture must be calibrated. This offline procedure consists of asking the involved subject to watch several videos or to start conversational tasks. These tasks have been selected to elicit different types of emotions [135]. At the end of each video clip/conversational task, the subject must provide an assessment of his mood by using a self-assessment manikin approach [136]. This approach is the official rating scale for emotion measurement.
It is composed of three panels, one per each AVD model parameter: arousal, valence, and dominance. Each panel consists of a rating that ranges from 1 to 9 (integer).

According to [135], the combination of these values defines the subject's macro emotion for the classification ground truth. During this working mode (i.e., Offline System Calibration – Fig. 3.12), the architecture collects the EEG data from all the available EEG channels (i.e., 14 channels), sending them to the Feature Extraction/Selection block. The outcomes of the self-assessment test are, instead, directly sent to the Train & Test block for cross-validation purposes.

As per the Fig. 3.12 workflow, the Feature Extraction/Selection block requires two inputs: the EEG signals and the Feature Extraction (FE) database (i.e., FE DB). The FE DB is a database that collects a set of scripts implementing several FE approaches. For this application, the database embeds some of the most recurrent FE techniques according to the state-of-the-art review in [134] and related works presented in Sec. 3.6. These approaches range from simple statistical features in the time-domain, up to wavelet-based frequency analyses. Nevertheless, it must be specified that this database can be easily manually expanded by adding new FE approaches.

Once the EEG data are received, the first automatic screening step consists of extracting a series of complexity metrics for each analyzed FE technique, as well as their combination. By setting functional thresholds on these metrics (e.g., memory availability on the target hardware, memory reserved for the application), it is possible to reduce the number of suitable FE approaches for the specific target, which can be an application logic or hardware that will implement the emotion recognizer. When the under-thresholds FE approaches have been selected, the proposed architecture analyzes the extracted features through a classifier-in-the-loop approach for the final feature selection.

Despite this is not a limiting choice, in the here proposed application, the feature selection step has been entrusted to the Neighborhood Component Analysis (NCA) approach, because of the technique's capability in weighting the selected features according to their contribution in terms of the final model accuracy. This last step aims to furtherly reduce the size of the features vectors, speeding up the online feature extraction and emotion classification steps, while keeping accuracy almost unaltered.

As the last calibration step, the optimized features are sent to a set of classifiers. Here, as per Fig. 3.12, a second grid check is carried out, to find the best combination of classifiers that maximizes the prediction performance, keeping in mind their implementation ease. The extracted classifier committee is composed of three classifiers that singularly analyze the parameters: arousal, valence, and dominance. The optimized processing chain (i.e., selected FE techniques + classifier committee) resulting from the double grid check, can be - thus - used for the online emotion recognition (Online Emotion Recognition - Fig. 3.12).

While the calibration phase is carried out in *offline* mode on a dedicated computer, the last operating branch (i.e., online emotion recognition – Fig. 1) runs on the robot computation core (considering a constraint of 256 MB for the background application).

In the nominal operation of the architecture, the robot is designed to randomly start a scheduled stimulation routine (red arrow in Fig. 3.12).

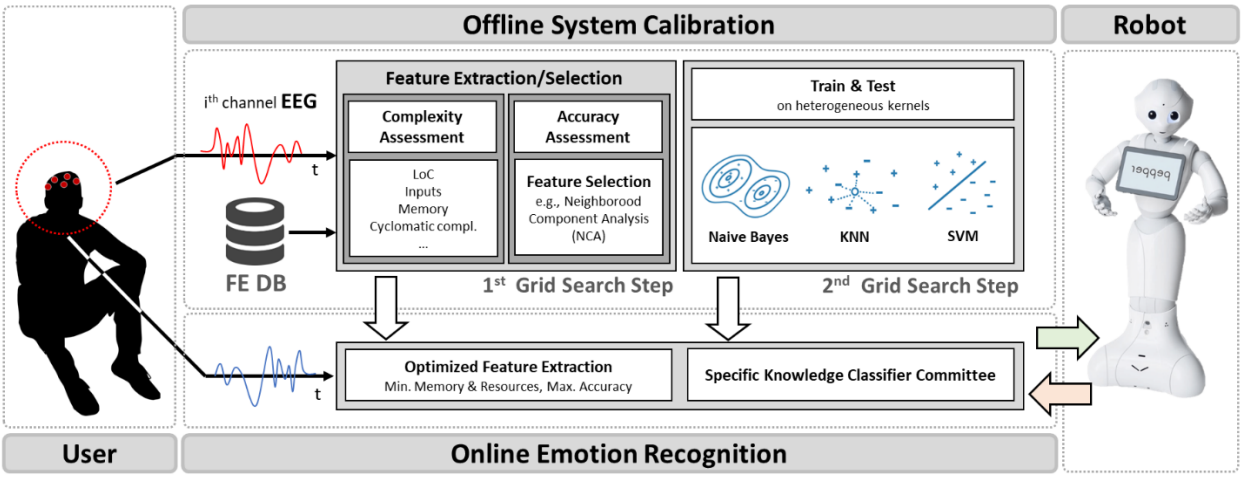

Figure 3.12 Block Diagram of emotion recognition architecture

The stimulation triggers the online emotion recognition branch that returns the predicted arousal, valence and dominance and, as a direct consequence, the estimated emotion (green arrow in Fig. 3.12).
The outcome of the emotion recognition system is, thus, transmitted to the FHIR server enriching the physician dashboard.
Locally, the robot also embeds an adaptive conversational task [137] realized in the context of the AMICO project.

3.6.1.1 Emotion Recognition System Pre-processing

When the emotion recognition system is enabled according to Sec. 2.2.4 procedure, the robot starts receiving streaming of 14 EEG channels (Fz, T_7, Cz, C_3, CP_5, CP_1, CP_2, CP_6, C_4, Pz, P_3, P_4, T_8, Oz), according to what is reported in Sec. 3.1.3.1.
The received EEG underwent the rASR-based artifact rejection stage according to Sec. 3.2.1.

Filtering. The EEG signals employed for the emotion recognition are subjected to an additional filtering stage. This stage consists of a 8^{th} order high-pass Butterworth filter with cut-off frequency of 4 Hz [126]. Overall, the signal is filtered in the band 4 Hz – 35 Hz. The upper bound is related to the pre-filtering stage operated before the transmission (see Sec. 3.1.3.1).

Downsampling. The resulting waveform is downsampled from 500 sps to 100 sps, by keeping 1 sample every 5 samples.

Trial Extraction. The resulting EEG signal is then subjected to a trial extraction phase. For the application it consists of a progressive storage of 33 s, of which 30 s after the conversational/video onset trigger (see Sec. 2.2.4) and 3 s before [126]. The buffer is composed of 3300 samples.

Baseline Removal. A baseline signal is extracted as the average of the values assumed by the EEG in the range [-3 s, 0 s((300 samples). This average value is then subtracted from the entire trial.

To furtherly reduce the amount of data to be analyzed during the complexity/accuracy assessment steps, the EEG signals, from each specific cortical area, are re-referenced and averaged, resulting in a single waveform per cortical group. In this respect, the architecture considers five main groups: frontal, central, parietal, occipital, and temporal. According to 10-20 international EEG positioning system, the considered cortical groups are composed as follows:

- Frontal: Fz
- Central: Cz, C_3, CP_5, CP_1, CP_2, CP_6, C_4
- Parietal: Pz, P_3, P_4, CP_5, CP_1, CP_2, CP_6
- Occipital: Oz
- Temporal: T_7, T_8

3.6.1.2 Feature Extraction/Selection

As shown in Fig. 3.12, the Feature Extraction/Selection block requires two inputs: EEGs and the FE DB. The FE DB consists of a compendium of the most recurrent FE techniques according to the related literature [134]. To generalize, in this application, the database embeds thirteen FE methods, which corresponds to those techniques that are used in more than two different papers by literature review in [132]. It must be stressed that this database can be enriched by including new scripts of primitive FE techniques.
The Feature Extraction/Selection block consists of three main steps: (1) the complexity assessment routine, (2) the accuracy assessment that completes the multi-optimization step and (3) the feature selection step via the NCA-based classifier in the loop algorithm. The first two steps, constituting the multi-optimization routine, are oriented to the identification of a combination of features properly embeddable on a target application logic or hardware. These steps are used to describe each combination with its conservative memory usage and script complexity, taking in the account the maximization of the accuracy. Instead, the third step aims to potentially reduce the FE chain memory usage and its complexity, leaving more room to the classification step implementation.

Complexity and Accuracy Assessment Routines. The FE DB consists of a database composed of a set of Python (or C-language) scripts. These scripts implement several primitive FE techniques summarized in Table 3.1. Their application of the EEG signals, for emotion recognition purposes, exploits the methodologic description provided by the references reported in the "Definition" column.
The FE DB also includes functional combinations of the previously reported FE primitive scripts for a total of C(n=13,k=1…13)= 8191 combinations (with no repetitions).
All the scripts reported in Table 3.1 have been realized in MATLAB® 2017a according to the related reference, and – thus – converted into Python codes employing the Compiler SDK (Python package) app with a specific focus on execution efficiency.
In this context, the parameter of implementation ease has been defined as a function of the number of code lines, the number and format of the overall involved variables, the

algorithm's memory usage, and the cyclomatic complexity as per McCabe's definition in [138].

This screening procedure based on the software metrics is summarized in Fig. 3.13 and supported by a purely demonstrative code example. According to Fig. 3.13, a script is extracted by the FE DB and undergoes four different assessment steps according to the above-mentioned methods: lines of code (loc(ex.c)), input format evaluation (inputs(ex.c)), memory usage (memory(ex.c)) and cyclomatic complexity (cyclomatic(ex.c)). The resulting values are then normalized to the target hardware availability or resource constraints for application logic. In the current application, constraints are manually defined by the user.

Table 3.1 – Primitive FE techniques with application reference

Id	Features Extraction Techniques	Definition
A	Wavelet Packet Transform (WPT)	[130]
B	Entropy on a frequency band (δ, θ, α, β)	[130]
C	Discrete Wavelet Transform (DWT).	[131]
D	Information Potential (IP) from the flexible analytic wavelet transform (FAWT)	[133]
E	Short-Time Fourier Transform (STFT)	[139]
F	Fast Fourier Transform (FFT) based Power spectral Density (PSD) on a frequency band	[132]
G	Auto-Regression (AR)	[140]
H	Common Spatial Pattern (CSP)	[141]
I	Built-in statistical operators: mean, median, standard deviation, max, min on the reference signal*	[132]
J	Logarithmic conversion of PSD	[132]
K	Hjorth features	[142]
L	Higher-Order Crossings (HOCs)	[143]
M	Derivative analysis	[143]

* It can concern several different input signals involved in the selected combination (EEG, Spectrum, etc.)

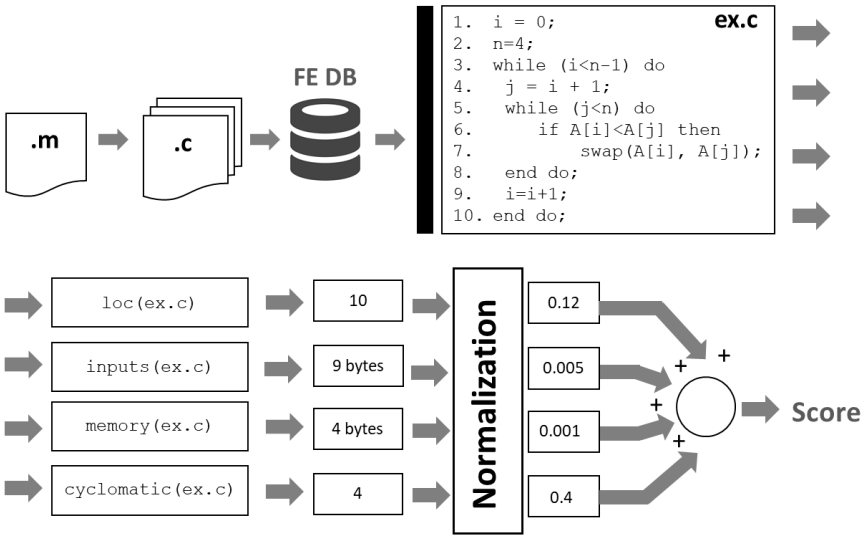

Figure 3.13 Automatic screening process based on software metrics

Specifically, before the calibration, the user sets a group of thresholds, one per each analyzed software metric. The normalization of the parameter values consists of the ratio between the current software metric value and the dedicated user-imposed threshold. The normalized values are then summed up to define a Score parameter (Fig. 3.13) as per the following equation:

$$Score = \left(\sum_{i=1}^{N_{sm}} \frac{sm_i}{Thr_i} \right) \Big/ N_{sm} \tag{9}$$

where N_{sm} is the number of the analyzed software metrics, sm_i is the i-th software metric, while Thr_i is the user-defined threshold for the i-th software metric.

All those primitive FE techniques, or primitives' combinations, which overcome one or more thresholds (normalized parameter >1) are automatically excluded from any further computation. Once the constraints-compatible FE methods are defined and the EEG signals are properly grouped resulting in only 5 signals per trial, the Feature Extraction/Selection block (Fig. 3.13) starts assessing the accuracy for each suitable combination.

Considering the training dataset, the five EEG signals (cortical groups) composing each trial are sent to a dedicated FE stage that selectively implements the FE techniques combination identified during the complexity assessment routine.

The resulting features are then used to train a set of sixteen classifiers. Among these, the best three classifiers are selected (according to the last step of the calibration pipeline furtherly detailed in the following), one per each AVD model parameter, resulting in a classifier committee. The final accuracy value, considered for the multi-optimization routine is, thus, derived by means of the lowest among the three cross-validation accuracies (conservative assessment). At the end of this step, a matrix containing the combination index and the related score, and accuracy is provided.

Multi-optimization routine. The complexity and the accuracy assessment steps provide the software metrics and the accuracy of a set of possible FE combinations. In the simplest version of the multi-optimization routine, each combination can be described as a point with coordinates (Ci_{Score} , Ci_{Acc}) in the bidimensional space {Score, Accuracy}. Specifically, Ci_{Score} represents the score of the combination of the i-th features, while Ci_{Acc} the accuracy related to the i-th combination. The multi-optimization routine consists of extracting a set of straight lines defining the edges of some areas here named priority areas. The optimal features combination selection is based on a weighted system in which, each priority area comprises between two straight lines has its own priority value (Pr – Fig. 3.14). The combination that falls within the zone with the highest priority value is automatically assumed as the best solution. In its current form, the use of straight-line equations allows rewarding those solutions with a minimum score with a tolerance of a maximum of 5% of accuracy. The linear kernel of the multi-optimization routine can be easily substituted with other kinds of polynomial kernels. Fig. 3.14 shows a demonstrative

distribution of 10 feature combinations in the 2D space {Score, Accuracy}. Moreover, Fig. 3.14 reports - in different colors - the first 4 priority areas with the related value. The straight lines defining the edges of the priority areas are also reported as dashed black lines.

According to the demonstrative routine in Fig. 3.14, the combination C4 resulted to be the only one in the priority area with Pr=13. In this case, it is chosen as the best solution and sent to the feature selection step.

It is important to note that, if the C4 is excluded and the priority area with Pr=12 is considered, it is possible to find three different solutions: C5, C6 and C9. In this case, the user must define the predominant metric for the selection.

It can be the accuracy (i.e, maximum value on y-axis), the score (i.e. minimum value on x-axis), or the user should manually choose the combination through the MATLAB GUI.

It must be specified that the multi-optimization routine can take into account also every single software metric, creating a weighted system with N_{sm} bidimensional subspaces (e.g., Accuracy-Memory, Accuracy-Lines of code, and so on). In this case, each software metric must be properly weighted to favor the user needs.

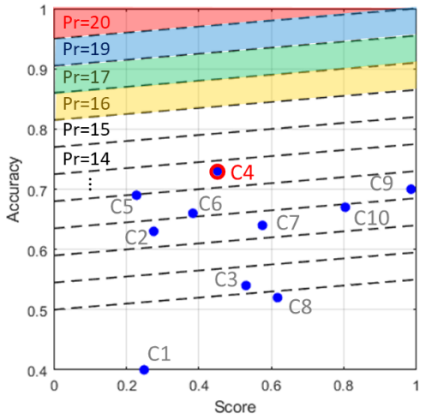

Figure 3.14 Priority areas and features combination distribution in the 2D space {Score, Accuracy} for the implemented multi-optimization routine

Feature Selection. Once the optimal combination of features is derived by the multi-optimization routine, the system embeds a feature selection procedure, aiming to furtherly reduce the classification problem complexity (number of features involved in the classification) and the memory usage of the FE stage in the online application of the architecture, leaving room for the classification step. For this step, only those algorithms, able to quantify the contribution of each feature in the overall system accuracy, have been considered [144]. For this reason, two feature importance approaches have been initially analyzed: the NCA and the Recursive Feature Elimination (RFE). The first approach (i.e., NCA) consists of assigning weights to the considered features according to the performance shown by a classifier trained on a features' subset. While the second one, RFE, exploits a wrapper-type approach to define a priority value for a fixed number of selectable features.

The explorative analysis of the feature selection methods showed that the NCA achieves higher accuracies than the RFE, for those classifiers that share the "neighborhood"

classification principle, such as the Naïve Bayes (NB), SVM, and the k-Nearest Neighbor (kNN). While, RFE was demonstrated to be more performant with linear threshold based classifiers, such as Decision Tree and RF. Since the here considered classifiers fall within the NCA applicability, this approach has been selected for the current version of the system application.

3.6.1.3 Classifier Committee

According to the rating scale meaning provided by [135], all the available emotions in human nature can be considered to operate in the spectrum of eight main emotions. They can be properly identified by studying the quadrants of the arousal, valence and dominance parameters. Specifically, considering a single parameter, e.g., arousal, ratings below (≤) 5 are classified as "Low Arousal" or LA, while the "High Arousal" classification, or HA, corresponds to ratings above (>) 5. This assignment procedure can be applied in the same way to valence and dominance (i.e., LV, HV and LD, HD). It leads to eight main emotions: anger = {HA, LV, HD}, joy = {HA, HV, HD}, surprise= {HA, HV, LD}, fear= {HA, LV, LD}, unconcern= {LA, LV, HD}, satisfaction= {LA, HV, HD}, calm= {LA, HV, LD} and sadness= {LA, LV, LD}. Due to the strong inter-variability among arousal, valence, and dominance parameters, the proposed system embeds a committee of classifiers, which consists of 3 different binary discrimination classifiers specialized on a single parameter of the AVD model.
In this application, to provide a complete and practical overview of the classification problem, we chose to analyze the accuracy of three different types of classifiers: the NB, SVM, and kNN. For the purpose, we selected classifiers that are intrinsically protected by the overfitting phenomena. In addition, to reduce memory consumption, we excluded those classifiers that need to resort to bagging or boosting techniques (such as Decision Tree or RF). It would computationally weigh down the overall system. In this context, it should be stressed that all the classifiers' parameters (e.g., support vectors, Lagrangian multipliers, and labels in SVMs) must be included in the overall memory consumption computing and should respect the constraint set by the user in the early part of the calibration step (Sec. 3.6.1.2).
For investigation purposes, the current system calibration phase embeds 16 classifiers among NBs, SVMs, and kNNs. Each main type of classifier has been implemented with different kernel functions. Table 3.2 summarizes the implemented classifiers, defining the acronyms.
Considering – for the sake of clarity - the only arousal parameter, the supervised learning approach returns a set of 16 cross-validated accuracy values, one per each evaluated classifier. Finally, the proposed architecture selects the best classification model (the one with the highest accuracy) and starts evaluating the valence and dominance parameters with the same rationale.
At the end of this procedure, the system has selected the three most accurate classifiers that will compose the classifier committee in an online emotion recognition context.

Table 3.2 – Employed Classifiers and Acronyms

Main	Kernel	Acronym	Main	Kernel	Acronym
NB	Gaussian	GNB	KNN	Fine	FKNN
	Kernel	KNB		Medium	MKNN
SVM	Linear	LSVM		Coarse Gaussian	CGKNN
	Quadratic	QSVM		Cosine	CosKNN
	Cubic	CSVM		Cubic	CKNN
	Fine Gaussian	FGSVM		Weighted	WKNN
	Medium Gaussian	MGSVM		Subspace	SKNN
	Coarse Gaussian	CGSVM	NN	Dense	DNN

3.6.2 Emotion Recognition: Real-Life Scenarios

Considering the assistive application in Sec. 2.2.4, once the robot application logic task terminates the calibration (Sec. 3.6.1), any new unlabeled observation can be accepted.
As a demonstrative workflow, the robot can, first, require the user's emotion verification through a neutral conversational task (no specific emotion elicitation). The assessment leads the system to the online collection of 33 s of brain activity from the 14 monitored EEG channels. According to the guidelines in Sec. 3.6.1.1, the resulting waveforms are averaged and re-referenced for cortical grouping purposes. The grouping procedure results in 5 waveforms, which are analyzed by a set of selected FE methods.
These FE techniques are derived through the multi-optimization routine and the NCA-based screening.
It must be specified that the features resulting from the previous step comprise all those features that overcame the NCA threshold for each single AVD model parameter. They are, thus, distributed to the related classifier composing the committee. Finally, each classifier involved in the committee returns a binary prediction. According to the arousal-valence-dominance status combination, as per Sec. 3.6.1.3, the committee determines the subject's emotions. Let us suppose that the arousal classifier returns A="Low" as well as the valence and dominance ones (V="Low", D="Low"), the robot will receive a sadness= {LA, LV, LD} label. The AAL robot can, now, plan feedback [137].

The AAL Robot Routines

The formalization of requests through the brain-robot interface enables automatic processes that aim to fulfill the user's needs. These automatic processes have been *ad-hoc* developed on Pepper robot and consist of autonomous navigation in the domestic environment and object manipulation. In the first case, Pepper natively provides a substantial number of methods that allow it to move smoothly in the surrounding environment. In the second case, Pepper's inherent abilities lack manipulation. This function has been expanded in the context of the present thesis work.
This chapter will provide an overview of Pepper's main methods, and the implemented navigation routine. The designed and implemented object manipulation routines will then be detailed in dedicated sections.

4.1 Main Pepper Methods

Using NAOqi, it is possible to create a wide range of applications with Pepper. The supported languages for this purpose are Python and C++. These applications can be developed through the modules that NAOqi provides. These modules have default methods built for specific processes. These methods are included in an API (Application Programming Interface) that allows the designer to easily handle various commands that Pepper can understand. Those that have been used in the development of this thesis work are detailed in the next sections. Further information on this framework and a full list of all API methods can be found in [145].

4.1.1 Navigation-related Methods

NAOqi Motion is the main tool for managing robot movements within an environment. This tool includes two specific APIs directly related to the robot's navigation control: *ALNavigation* and *ALMotion.*
The first API (i.e., *ALNavigation*) has low flexibility in terms of routine reprogramming, but it provides the Pepper robot with a sequence of already complete routines useful for indoor navigation.
Some examples are the intrinsic ability to automatically reprogram the trajectory to reach a target point avoiding obstacles (method: `ALNavigationProxy::navigateTo`), or the possibility of creating a dynamic map of obstacles by means of Source Localization And Mapping (SLAM) approaches that are "black-box" provided by the `ALNavigationProxy::explore` method. According to this last possibility, the robot also integrates a proprioceptive system to navigate inside the extracted map.
Several tests with these methods showed that the ALNavigation API performance are far from optimal. When navigating an area of no more than 4 m², Pepper stopped more than 1 m far from the desired destination on several occasions. Collision avoidance performed adequately, but in much test-runs Pepper would opt to stop upon finding an obstacle, rather than trying to recalculate a trajectory around the obstacle. Unfortunately, there is no way of modifying the navigation algorithm or tuning the intrinsic parameters. The lack of information about the base-algorithms limits the understanding of the problems.

Ultimately, the use of the ALNavigation module is discouraged to navigate within the map in a controlled way, nevertheless, the explore method with metrical map extraction is, instead, useful for object area identification purposes. An obstacle-free environment is preferrable for the specific application.
The second API (i.e., *ALMotion*) permits greater programming flexibility by providing a high number of movement primitives. Specifically, the API deals with the control of the stiffness of the various joints of the robot, the management of the degrees of freedom of the same, as well as the robot's locomotion. The ALMotion API also allows the definition of joint trajectories by implementing an intrinsic system of smooth interpolation.
In the context strictly related to navigation safety, ALMotion provides control over parameters such as Self-collision avoidance, External-collision avoidance, Fall manager, and Smart Stiffness. Concerning the first two parameters, ALMotion allows altering the motion to avoid self (see Self-collision avoidance [146]) or external collision (External-collision avoidance [147]).
Concerning the latter, ALMotion allows setting a tangential security distance as the minimum distance between any point of the robot and any obstacle during a move. It has been set to 0.1m (default).
The orthogonal security distance, i.e., the frontal distance at which the robot will stop when moving directly towards an obstacle at full speed, is set to 0.3m [147].

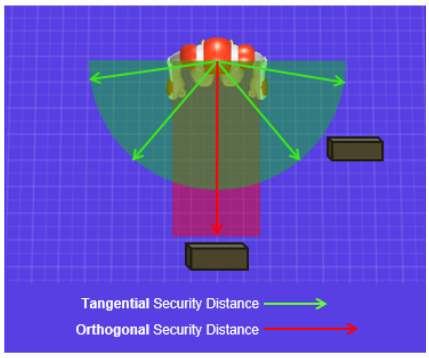

Figure 4.1 ALMotion safe distance for Pepper Robot. Picture extracted by [147]

ALMotion can be also employed to adjust the stiffness to spare power when the robot is not moving. The method is active by default. Another active service by ALMotion is the Fall Manager. The robot may fall, in this case, the service allows the system to detect falls, actuating routine to protect arms before touching the ground. The robot can therefore fall, without damage to sensible parts. The most used and customized method for navigation purposes is the `ALMotionProxy::moveTo` method which allows the user to set the relative displacement (in terms of meters) with respect to x, y, and ϑ considering the torso frame of the robot as a reference. More details about Pepper body frames are available at [148]

4.1.2 Manipulation-related Methods

The here proposed manipulation routines partially exploit some methods related to the ALMotion API, during the phases of approach objects (`ALMotionProxy::moveTo`) and

adjustment of the various joints in the phases of manipulation. (`ALMotionProxy::angleInterpolationBezier` o `ALMotionProxy::angleInterpolation`).

These last methods are useful to create a trajectory interpolation of the various segments involved in the movement by providing the interpolation method with the times and angle values that a particular articulation assumes during movement [149]. Fig. 4.2 shows an application of the interpolation with Bèzier method, which concerns the definition of 3 points, unevenly distributed in time and interpolated, related to the Pitch value of the right shoulder (RshoulderPitch). Two points are chosen for the right elbow movement (the roll parameter of the joint is set). The manipulation is mainly based on data acquired by Pepper's RGB cameras placed on the forehead and within the mouth, and the 3D sensor placed within its left eye as introduced in Sec. 2.4.
The access to raw RGB cameras is managed by *ALVideoDevice* API. The API provides access to raw images for framerates to 30 fps. The function is activable by exploiting the `ALVideoDeviceProxy::getDirectRawImagesLocal` methods to extract the frame to be furtherly processed, and `ALVideoDeviceProxy::releaseDirectRawImages` to release the processed image [150].
Although this feature was not used for the specific application, Pepper is also enabled for remote streaming of images captured by the RGB camera. It is possible to use gstreamer (already installed on the robot) to create a UDP (+RTP) stream of the front (/dev/video0) and bottom (/dev/video1) camera from Pepper. The stream is then retrieved on the host (PC) with gscam.
The management of the outputs related to 3D images is instead entrusted to the *ALSegmentation3D* API inside the NAOqi Vision tool. ALSegmentation3D extracts the objects present in the field of view of the robot by doing a segmentation of the depth image (returned by the 3D sensor) in blobs of similar depth. In order to segment the depth image into blobs of similar depth, a region growing based algorithm is used: pixels in the depth image can be linked to their neighbors to form blobs if the difference in depth between them is below a threshold (DepthThreshold). DepthThreshold can be changed with the function `ALSegmentation3DProxy::setDeltaDepthThreshold`.

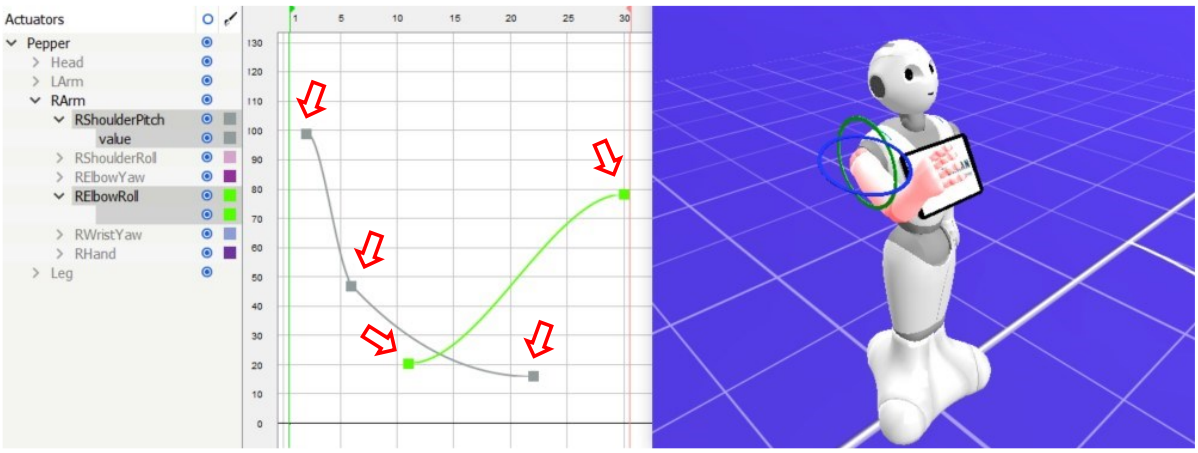

Figure 4.2 Demonstrative example of ALMotion angleInterpolationBezier method usage on the pitch of right shoulder joint and roll of the right elbow joint.

It is therefore possible to derive the list of blobs present in the Pepper field of view, and from these, it is possible to extract the average distance or point of the extracted image. Fig. 4.3 shows from the left to the right a raw image from top RGB camera, the same perspective but employing the 3D sensor and overlap of both frames.

4.2 Navigation Routine

In the proposed application, the navigation routine has been mainly entrusted to the native AL Navigation and ALMotion API provided by the NAOqi.
At first use, the robot enables the exploration method. It consists of investigating the surrounding in a circle of programmable radius. In this context, the robot starts moving from its absolute initial position (0,0) enabling the built-in SLAM algorithm for the 2D position computation. To improve the native SLAM algorithm precision, a novel procedure of body rotation angle correction has been introduced in this application.
It consists of evaluating the discrepancy between the target rotation angle and the actual rotation of the robot. It allows compensating for bad rotation management by reducing the odometry error.
Once the exploration is completed, Pepper stores a 2D matrix representing the extracted map. The robot is now able to identify the absolute position of the shelves containing all the goods to be manipulated. Pepper can, thus, estimate the absolute position of the object, or at least a rough estimate of it. Pepper will move toward the identified area starting the object manipulation routine.

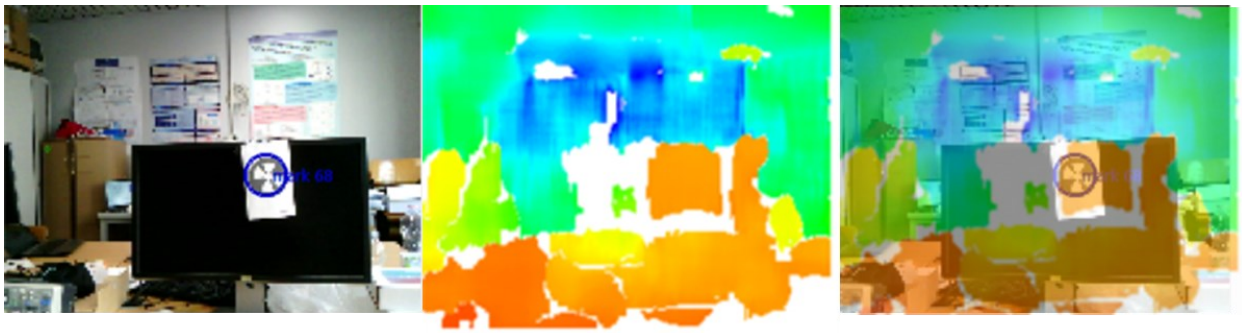

Figure 4.3 Demonstrative examples of the raw image from the ALVideoDevice API, 3D blob representation by ALSegmentation3D API, and a merge of both frames to check correspondence.

4.3 Object Manipulation Routines

The automation of object and materials handling activities showed a high impact on efficiency and productivity increment in several different applicative areas spanning from healthcare to assistive ones, where most of these operations are still performed manually. Let us think, a survey conducted in European assistive facilities and hospitals [151] in 2018 observed that about 40% of the nurses' and caregivers' time is dedicated to none-value-adding activities that involve repetitive delivery, correction of erroneous drug delivery, and so on [2]. The advances in robotics technology, such as the high number of sensors equipped on the automatons, the increasingly powerful on-board computers, the artificial intelligence-based algorithms, and the even more performant simultaneous

source localization and mapping (SLAM) approaches, paving the way to the employment of robots to help medical staff, reducing their workload by carrying out easy and repetitive tasks. It improves – as a consequence – the efficiency of the overall healthcare structure (both in hospitals and in the domestic assistance framework).
As introduced in Sec. 1.2, a category of robot widely used in healthcare facilities and domestic assistance is the social robots (under the acception of companion robot). These kinds of robots are mainly designed for verbal and animated interactions with users and patients. Nevertheless, this kind of robot did not find practical application in an assistive context, because most of the time they have two upper arms, but they lack object manipulation capabilities. An example of a social robot widely present in healthcare structures is, in fact, Pepper (Sec. 1.4). Despite its large use, the presence of a single servomotor to control the entire hand, and the absence of sensing feedback for force management, prevents its use in an assistive and support context for medical staff.
In this section, a capabilities expansion of Pepper will be proposed, through the introduction of a set of automatic object manipulation routines.
Due to Pepper's large diffusion and employment in healthcare facilities, the proposed functional re-design aims to make the robot suitable also for medical staff support and for the patient's assistance both in domestic and ambulatory frameworks.
The developed framework has been named MONOCULAR (eMbeddable autONomous ObjeCt manipULAtion Routines) [152] and embeds a set of manipulation routines based on the in-loco functional combination of data from two RGB cameras and a 3D depth sensor for a reliable object grabbing. The MONOCULAR framework allows the robot to recognize objects, approach them, grabbing and scan them for check purposes. MONOCULAR working is based on frames captured by an RGB camera and used to create an input stream for an embedded version of the real-time object detection engine "You Only Look Once" (YOLOv3) by [153, 154]. MONOCULAR uses 3D sensors to improve some grabbing processes. Through in-loco data fusion of RGB cameras and 3D sensor outcomes, MONOCULAR can autonomously drive several servomotors to complete the grabbing procedure, without the need for internet connections or external frameworks like ROS.

4.3.1 Preparation Steps

Fig. 4.4 reports a schematic overview of the first preparation steps for realizing the MONOCULAR framework. The workflow starts with the acquisition of a sequence of frames from the embedded top RGB camera (the forehead one) and the depth map frame from the 3D sensor placed in the left eye of the robot.

1st step: Align and Correct. The frames from the top RGB camera and the 3D sensors are displaced between them due to the mutual positioning and the dedicated field of view (FOV) as per Fig. 2.9. For this aim, a dedicated step for image alignment and correction step is needed. The procedure is composed of two main steps as proposed in our work [152]: (i) border correction and (ii) frame extraction.
The first step, i.e. border correction, consists of evaluating the mutual FOV overlap of the two cameras. Specifically, it defines the number of pixels to be removed to realize a matched image between the RGB and the 3D sensor frame (yellow areas in Fig.4.4).

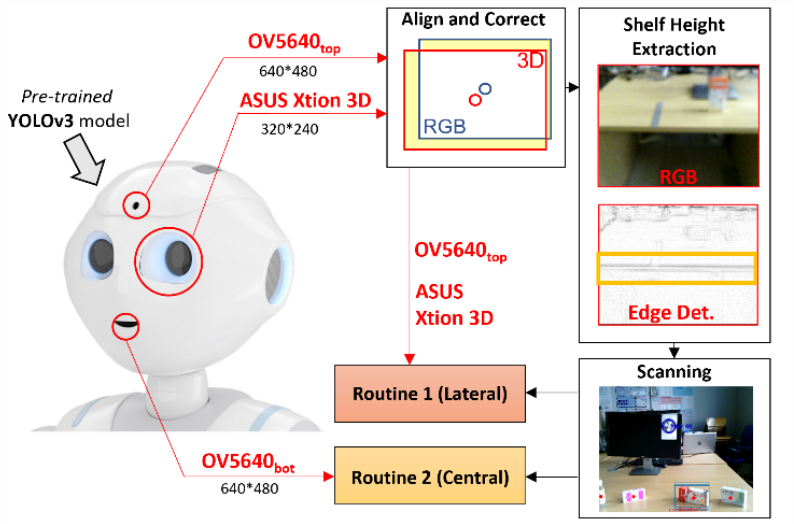

Figure 4.4 Overview of the Workflow of MONOCULAR framework

The software cropping operation constitutes the frame extraction step. Experimental results demonstrated that for Pepper Y20 cropping 14 pixels @ 320*240 resolution from the bottom of the 3D sensor's image, compensates for the mutual camera displacement in height while keeping into account the different vertical FOV. Also, 13 pixels @ 320*240 resolution crop on the left 3D sensor image covers the lateral displacement of the left eye with respect to the zero on the x-axis, where the forehead camera is placed. The different horizontal and vertical FOVs also require cropping operations on the RGB images. This results in the removal of 25 pixels @ 640*480 resolution on the top of the image and 13 pixels @ 640*480 resolution on the right.

2nd step: Shelf Height Extraction. To properly manage any type of manipulation routine, the proposed system must be able to estimate as precisely as possible the height of the shelf on which the goods are placed. For this purpose, MONOCULAR embeds a dedicated routine that exploits the top RGB camera and the 3D sensor.

Fig. 4.5a shows the height estimation setup. Specifically, the robot acquires the distance between the 3D camera and the center point of an NAO mark placed on a frontal surface via depth map data. This distance has been named $d_{T,NM}$ in Fig. 4.5b. Once $d_{T,NM}$ is extracted, the head pitch is increased up to α_{HP} =36° (max allowed: 36.5°). Next, MONOCULAR queries the robot for a top RGB camera photo, which is analyzed by an embedded and low-complexity Laplacian edge detection algorithm with a 3x3 convolution mask [155]. If the central area of the image is occupied for >70% by strong edges pixels, it means that the shelf edge is centered with the FOV of the camera. If <70% is detected, the robot starts moving toward the shelf with preset steps. Optimal positioning is reported in Fig. 4.5a and 4.5b, where the center of the RGB camera frame coincides with the shelf edge. Fig. 4.5c shows the corresponding RGB camera frame, while Fig. 4.5d displays the Laplacian edge detection outcome. Since the distance between the NAO mark and the shelf edge is known *a priori* (d_1 in Fig. 4.5b), the height of the shelf can be estimated as per the following equation:

$$h_{sh} = h_{ct} - \left(d_{T,NM} - d_1\right) \cdot \mathrm{tg}(\alpha_{HP}) \quad (10)$$

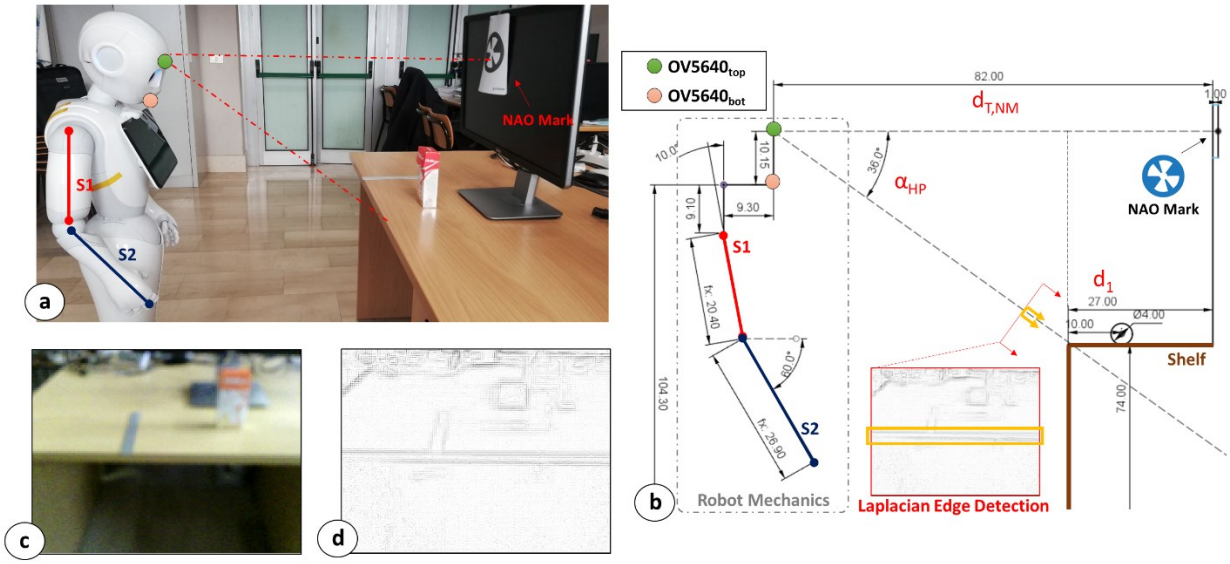

Figure 4.5 Shelf Height Extraction procedure. (a) Snapshot of the experimental setup; (b) robot mechanics explanation for the procedure; (c) RGB camera snapshot; (d) Laplacian Edge detector outcome for frames in (c).

where h_{sh} is the shelf height, h_{ct} is the height of the top RGB camera (i.e., 114.45 cm), $d_{T,NM}$is the distance between the camera and the NAO mark as per Fig. 4.5b. d_1 is a preset distance between the NAO mark and the shelf edge, while α_{HP} is the head pitch angle.

3rd Step: Scanning. Once the shelf height is estimated, the cropped RGB acquisition is then sent to an object detection routine based on the YOLOv3 method [153, 154]. The YOLOv3 model is offline pretrained to recognize specific objects. It oversees extracting a number of bounding boxes returning, for each of these, the probability that the box contains an object and the probability that the object belongs to a specific class. Once the object to be manipulated is recognized via static scanning of the shelf, the MONOCULAR framework decides the manipulation routine be carried out. It can be Routine #1 (Fig. 4.4) if the objects are placed laterally from the robot's point of view, or Routine #2 if the object is placed centrally.

4.3.2 Object Detection Algorithm

The MONOCULAR object detection engine has been entrusted to a portable version of YOLOv3, the Mini-YOLOv3 by [154]. The algorithm has been implemented as Python scripts on NAOqi OS. The object detection model pre-training has been carried out offline and consisted of data collection and preparation, model training, inference testing, and model extraction. During the data collection, images were captured both via the forehead and mouth camera, with different operative angles (repeating every time the approach routine).

For a proof of concept, four classes of pharmaceutical packages, two drinks, and four food envelopes have been considered. Overall, 3982 images were collected for a total of 15928 annotations (4 per image). The labeling has been manually realized via LabelImg software. As per the guidelines in [152], during the labeling, the bounding boxes were

drawn including the entirety of the object and a small amount of space between the object and the bounding box.
The images were then prepared to verify if none of the annotations are out-of-frame, checking the EXIF orientation correctness, resizing the image to match the final model request, and correcting the color to improve the model performance (grayscale and contrast). For the purpose, Roboflow preprocessing steps have been included in the data preparation. They consisted of auto-orienting, data augmentation to protect the model against overfitting, and YOLO model match. Roboflow has been selected because it allows downloading the preprocessed images and annotation locally on a Jupiter Notebook including Google Colab. Specifically, this latter online notebook has been used because it provides free GPU usage for complex model training. For the here proposed problem, a Keras YOLOv3 implementation and a repetitive fine-tuning approach (by freezing and unfreezing the YOLOv3 Darknet body) have been used. Training started from the available pre-fit weights [152], then, new data have been uploaded via the Roboflow setting a batch size of 32, epochs at 500, and Adam optimizer with a learning rate of 10^{-3} with Darknet body full-freeze and 10^{-4} for unfreezing and fine-tuning. The fitted model, with the custom weights, has been exported and implemented as a Python script.

4.3.3 Routine #1: Laterally placed objects

4th Step: Robot-Object Adjustment. Since all the objects are placed at specific coordinates on the shelf (as per Sec. 2.5) to facilitate the grabbing procedure, if the YOLO tag has been correctly extracted according to step 3 described in Sec. 4.3.1, MONOCULAR starts a robot-object alignment procedure compatible with the position of the latter. The maneuver uses as a positioning reference the NAO mark placed centrally on the shelf in accordance with Fig. 4.5a. This postural adjustment consists of small lateral movements of the robot according to the distance from the center of the torso framework and arm selected for the grasp. The shoulder joints are positioned at ± ~255 mm w.r.t. to the 0 x-axes of the robot.

5th Step: Object Approach with Scanning. After the postural adjustment, the robot starts moving toward the shelf by keeping the NAO mark in the center of the forehead RGB camera frame. The distance is assessed on the 3D sensor outcome every 200 ms. These approaching movements are carried out slowly to avoid omnidirectional wheels slipping.
During this movement, the arm selected for the grab is set in position #1 as per Fig. 4.6a. The robot movements towards the shelf, are stopped at a specific distance which ensures that the arm distension coincides with the reference point of the envelope to be grasped according to the preset object positioning grid.
Once the grabbing position is reached, MONOCULAR triggers the robot to scan again the shelf, selecting the YOLO tag of interest. Yaw and pitch degrees of the neck servomotor are progressively adjusted to keep the tag in the center of the frame (Fig. 4.6a – Step 1). According to Fig.4.6a, the RGB image and the fixed YOLO tag are resized, coherently

with the 3D image resolution, by applying a bilinear interpolation method. Only the area covered by the tag is -thus- selected for further analysis. The YOLO central anchor point coordinates are used to extract a reference distance in the blob vision from the 3D sensor.

6th Step: Object-Hand Coverage. This distance is used to fix a specific color in the colormap (i.e., red in Fig.3.b), by setting the depth threshold for the segmentation is set to 5 mm, and the colormap limits are restricted in the close proximity of the reference value in order to emphasize the selected object area. Fig. 4.6b shows the initial blob vision from Position #1. The area that includes the red blob constitutes the object to be grabbed. Once the number of "red pixels" is set, determining the *100% of coverage*, the robot hand moves on the selected object, partially covering it, from the 3D sensor point of view.
As a first target trajectory, the hand positions on the hypothesized object coordinates. This step is shown as Position #2 in Fig. 4.6a. Then, MONOCULAR counts the number of pixels that remained unaltered from the first red blob. This approximation is possible because of the robot's focus. Indeed, the presence of the hand sets to white and gray parts of the 3D blob from the depth map, according to Fig. 4.6a. If the *remaining blob* still occupies *more than ~ 40 – 45 %* of the initial coverage, the hand is moved ipsilaterally with the object side with the majority contribution (i.e., left side in Fig. 4.6c). In this process, the percentage is periodically monitored.
When it achieves the 40 - 45% of the initial number of "red pixels", the hand stops moving.

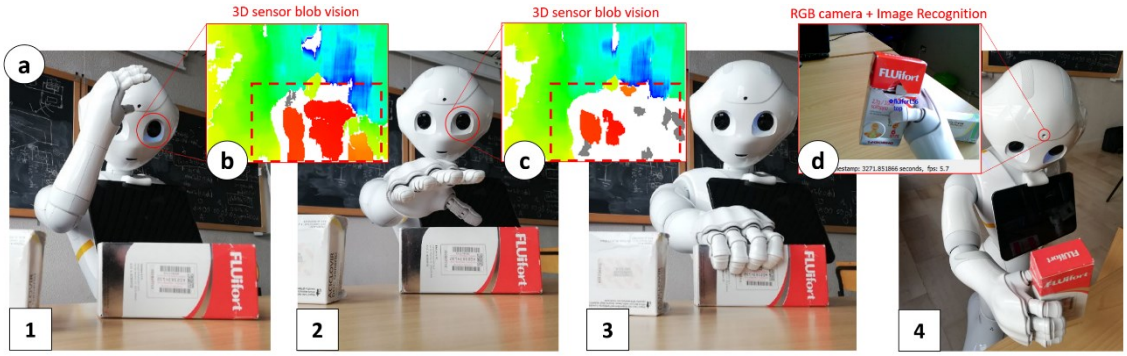

Figure 4.6 MONOCULAR Routine #1: Grabbing routine workflow. (a) Four-frame demo of the grasping routine; (b) Depth map by the 3D sensor before the hand positioning; (c) Depth map by 3D for hand-vs-object covering analysis; (d) Built-in image recognition system to validate the grabbed object.

7th Step: Grabbing and Scan. The robot's hand is now in the desired position. The shoulder pitch is progressively adapted to press against the shelf plane in order to develop a good grabbing force on the object.
The shoulder servomotor actual degrees are compared with the desired ones. When the current degrees stop to decrease, even if a request occurs, it means that the object has been blocked and no more pressure is needed. This step translates into hand closing.
Three sequential opening and closing procedures were demonstrated to strongly increase the success rate in object grabbing.
Once the object is firmly locked, it is lifted and rotated in order to position it stably near the front camera (See Fig 4.6 – Position #4). In this position, the object is analyzed through

the built-in image recognition engine to confirm the request matching. The robot is now ready to properly deliver the goods to the user who requested it.

4.3.4 Routine #2: Centrally placed objects

4th Step: Target Height Extraction. This step starts with the assessment of the shelf height parameter (according to eq. 10), h_{sh}, and the height of the object to be grabbed, namely h_{obj} in the following. This latter parameter is stored in the object recognition dictionary and can be accessed by Pepper through a memory call.
As a first step, MONOCULAR estimates the maximum height that the whole arm, given by the combination of S1 and S2 segments with proper orientation, should have. For this purpose, a target height is extracted according to the equation:

$$h_{tg} = h_{sjnt} - (h_{sh} + \frac{h_{obj}}{a}) \quad (11)$$

where h_{tg} is the target height defined as the maximum height that the arm (S1+S2) should have to properly approach the object, h_{sjnt} is the shoulder joint height (i.e., 95.2 cm in this application), while a is a coefficient related to the height of the object, h_{obj}.
The coefficient a is provided by the dictionary of objects stored in the Pepper memory and is related to the mass distribution of the package. Experimental analysis showed that a general height of approach of $\frac{h_{obj}}{2}$ (i.e., with a=2) allows the robot to achieve high grabbing accuracy with most of the involved packages.

5th Step: Arm Segments Movement Planning. Once the h_{tg} has been estimated, MONOCULAR compares h_{tg} with the length of the segment S1 (l_{S1}= 20.4 cm in our application).
If $h_{tg} \leq l_{S1}$, according to Fig. 4.7, MONOCULAR computes the angle that the S1 segment should keep with respect to the torso axis (α_{S1} in Fig. 4.7), according to the following equation:

$$\alpha_{S1} = \cos^{-1}\left(\frac{h_{tg}}{l_{S1}}\right) \quad (12)$$

In this configuration, the second segment, S2, remains perpendicular to the torso axis as per α_{S2} in Fig. 4.7.
Nevertheless, if $h_{tg} > l_{S1}$, the second segment S2 must be moved, by increasing α_{S2} in Fig. 4.7 to compensate for the difference between h_{tg} and l_{S1}, according to the equation:

$$\alpha_{S2} = \sin^{-1}\left(\frac{h_{tg} - l_{S1}}{l_{S2}}\right) \quad (13)$$

where l_{S2} represents the length of the second arm segment S2.
The combination of α_{S1} and α_{S2} is mirrored on the opposite arm.

6th Step: Object Approach. Once Pepper assumed the correct posture in a safe area (with no collision), the MONOCULAR framework starts driving the robot toward the object, keeping the achieved grabbing position as per Fig. 4.7. The movement considers

the distance between the robot and the NAO mark (through 3D sensor) and the distance between the object and the shelf edge (preset).

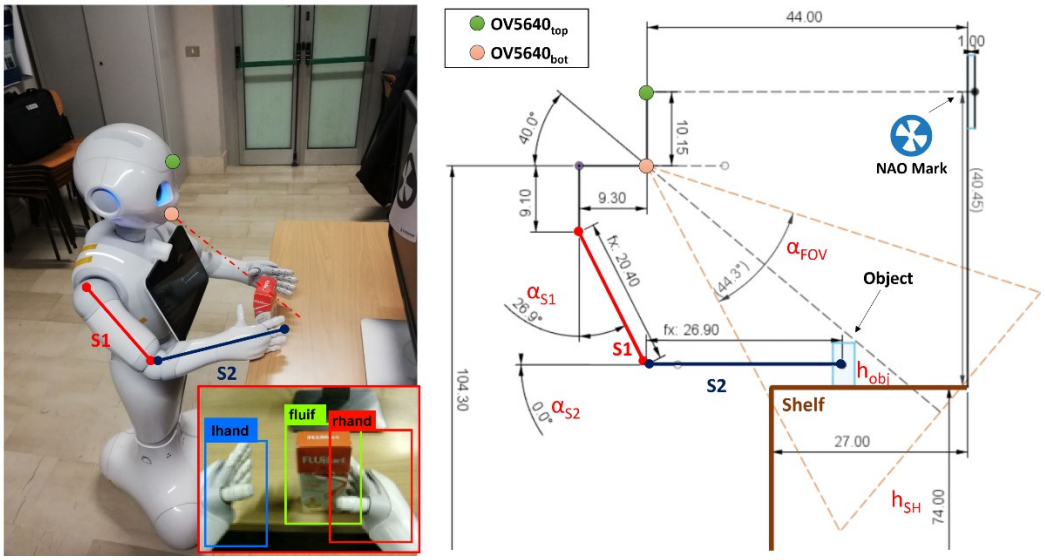

Figure 4.7 Arm Positioning and Object Approaching phases of MONOCULAR Routine #2 with robot mechanics explanation

All those approaching movements are carried out slowly to avoid omnidirectional wheels slipping. When in the grabbing position, MONOCULAR runs an object position check step. It consists of taking a frame of the bottom RGB camera perspective that is analyzed by a YOLO detector for stable labeling. The object position check procedure supervises the correct robot-object alignment.

As a first step, the proposed system extracts the distance between the camera and the object position, calculating d as follows:

$$\mathrm{d} = (l_{S1} \cdot \sin(\alpha_{S1}) + l_{S2} \cdot \cos(\alpha_{S2})) - 9.3 \tag{14}$$

where 9.3 cm corresponds to the horizontal distance between the bottom camera and the shoulder joint position according to robot mechanics in Fig. 4.7. The projection of d on the bottom camera axis is given by

$$\mathrm{x} = \frac{\mathrm{d}}{\cos(40°)} \tag{15}$$

where 40° is the mutual angle of OV5640 cameras (bottom camera axis). Starting from the above-presented parameter, x, it is possible to define the vertical field of view of the camera, assessed in correspondence with the object to be grabbed. This vertical field of view projection is defined as per:

$$\mathrm{VFoV}_{\mathrm{obj}} = 2 \cdot \mathrm{x} \cdot \mathrm{tg}\left(\frac{\mathrm{VFoV}_{\mathrm{bot,cam}}}{2}\right) \tag{16}$$

where $\mathrm{VFoV}_{\mathrm{obj}}$ is the projection of the vertical field of view of the camera ($\mathrm{VFoV}_{\mathrm{bot,cam}}$) in correspondence of the object. The extraction of the $\mathrm{VFoV}_{\mathrm{obj}}$ parameter allows MONOCULAR to assess the correct robot-object alignment. For this purpose, the framework considers only the central vertical column of the taken frame. Then, it extracts

the minimum number of pixels that should belong to the object in a specific zone of the frame. These zones depend on object height, h_{obj}, shelf height, h_{SH}, and robot arms positioning.

Specifically, considering an object placed on the shelf, it is possible to derive the position of its base as:

$$Ob_{low} = h_{bot,cam} - x \cdot \cos(50°) - h_{SH} \tag{17}$$

where $h_{bot,cam}$ is the height of the bottom camera (i.e. 104.3 cm in this application), while 50° is derived by 90°-40° with 40° the mutual angle of OV5640 cameras.

The Ob_{low} can be expressed in pixels considering the following relationship:

$$Ob_{low|pxl} = 240 - \left\lceil \frac{Ob_{low}}{\left(\frac{VFoV_{obj}}{480}\right)} \right\rceil \tag{18}$$

MONOCULAR also extracts the upper limit starting from $Ob_{low|pxl}$, as per:

$$Ob_{high|pxl} = Ob_{low|pxl} + \left\lceil \frac{h_{obj}}{\left(\frac{VFoV_{obj}}{480}\right)} \right\rceil \tag{19}$$

If the desired object is recognized, at least, in the vertical range $[Ob_{low|pxl}, Ob_{high|pxl}]$ as defined by eq. (18) and (19), the grabbing can start, otherwise, the robot should adjust its position to fit, at least, the range.

7th Step: Side Selection and Mutual Coverage. When the object position check is completed (Fig. 4.8 – step 1), the MONOCULAR framework starts assessing the overlapping degree among hands bounding boxes by YOLOv3 and the recognized object to be grabbed.

The hand with the bounding box mostly overlapped with the object is selected for the grabbing procedure, while the other hand is raised to avoid interference in lateral adjusting movements (Fig. 4.8 – step 2).

If the overlap involves less than 25% of the hand box, the robot is laterally moved to cover at least this limit. Once the alignment is over, MONOCULAR commands the hand closing (Fig. 4.8 – step 3). The routine ends with scanning that is carried out by means of a built-in image recognition system, following the guidelines proposed in [152].

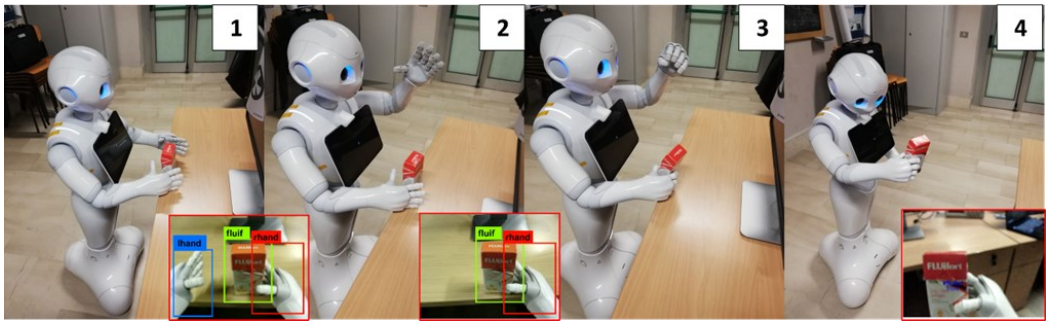

Figure 4.8 MONOCULAR Routine #2: four-frame demonstrative sequence for grabbing and scanning the centrally placed object

Experimental Results

This section will analyze in terms of performance the innovations introduced in the context of the present thesis. Formally, the section will be divided into two macro areas: one related to the Brain-Robot interface covered in Chapter 3, and one related to the routines implemented on Pepper and covered in Chapter 4. The first part of the chapter will be dedicated to the definition of datasets useful for the extraction of experimental results in the two involved macro areas. The second section will define the performance of BCIs used in the context of the AMICO infrastructure, analyzing them with standard metrics of accuracy and speed. The fourth section will analyze the cognitive state extraction system based on t-RIDE, while the fifth section will outline the performance of the emotion recognition system. Finally, a series of analyses for the optimization of the manipulation routines introduced on AMICO's robotic agent, Pepper, will be provided.

5.1 Datasets

5.1.1 Brain-Robot Interface

5.1.1.1 MRP-based BCI

MRP-based BCI (Sec 3.3) training/testing datasets have been acquired from seven subjects at the Design of Electronic Integrated Systems Lab, Politecnico di Bari, Italy. The mean age of involved subjects is 30 ± 4 years old, while the Male:Female ratio is 7:0. Participants have no BCI illiteracy.

Acquisition Timing Distribution. The training and testing sessions have been carried out on different days and at different moments of the day (early morning, late morning, before/after lunch, and afternoon).

Subject Setup. The experiment was conducted in a laboratory environment where the subjects sat in a comfortable chair approximately 70 cm - 1 m away from a 15″ monitor screen. No specific posture was required for subjects. No information about posture has been collected. The subject wears the EEG headset.

Environmental Conditions. The environmental conditions were not controlled allowing persons in the surroundings to discuss, talk, move, and so on, recreating a more suitable real-life scenario. However, no environmental disturbances were recorded during the acquisitions. Different levels of lighting have been adopted during the tests.
Table 5.1 summarizes the characteristics of the proposed dataset.

Dataset Acquisition. During the training sessions, the subject, which sits in front of the screen with the EEG headset worn, is subjected to the same protocol provided for MRP-based BCI in Sec. 3.1.3.2. The difference is that the training protocol suggests the side to be selected after the acoustic sound, by coloring it in blue.
Each session (regardless of training/testing phase) is composed of ~7 min of LH and RH finger movements (except for Subject 7). Movements are equally distributed between LH and RH.

Table 5.1 – Training/Testing Dataset Characteristics for MRP-based BCI

Sub.	Training Observations RH+LH	Testing Observations RH+LH	Num. Sessions Tr | Te	Note
1..6	1000 (equally distributed RH, LH)	200 (equally distributed RH, LH)	10 | 2	**Tr:** 10 sessions * 100 equally distributed finger movements. **Te:** 2 sessions * 100 equally distributed finger movements.
7	1000 (equally distributed RH, LH)	200 (equally distributed RH, LH)	20 | 2	**Tr:** 20 sessions * 50 equally distributed finger movements. **Te:** 2 sessions * 100 equally distributed finger movements.

5.1.1.2 P300-based BCI and Cognitive Status Extraction

Datasets used for the P300-based BCI (detailed in Sec. 3.4) training and testing have been acquired from four volunteers at g.tec medical engineering GmbH site in Schiedlberg, Austria and two volunteers from Politecnico di Bari, Italy. The mean age of involved subjects is 32 ± 3 years old, while the Male:Female ratio is 4:2. Participants have no BCI illiteracy.

The training and testing sessions have been carried out with the same Acquisition Timing Distribution features of the MRP-based BCI at Sec. 5.1.1.1. Similarly, the Subject Setup and Environmental Conditions also follow the guidelines imposed by the previous section. Concerning the Subject Setup, during the recording, 4 subjects adopted a relaxed position similar to a lying position, while 2 subjects maintained a composed posture on the chair with an upright back position.

Concerning the Environmental Conditions, several disturbances were recorded during the acquisitions. Most of these were auditory and were monitored according to their audio loudness expressed in *loudness units relative to full scale* (LUFS). All the training/testing sessions in which <90% of the duration was spoiled by sounds higher than -54 LUFS have been included in the here-proposed results to provide a conservative metric for accuracy. Only 1 session has been excluded with the 95% of disturbed observation (sound >-54 LUFS).

Table 5.2 summarizes the characteristics of the proposed dataset.

Dataset Acquisition. During the training sessions, the subject, which sits in front of the screen with the EEG headset worn, is subjected to the same protocol provided for P300-based BCI in Sec. 3.1.3.2. During this session the subject must select a target element through the dedicated input bar. Next, the paradigm starts flashing rows and columns for 30-50 runs, repeating the procedure for at least 1 session.

In the training context, the *run* is defined as the sequence of 7 flashes in which all the rows and columns of the 3x4 matrix are lighted.

Table 5.2 – Training/Testing Dataset Characteristics for P300-based BCI

Sub.	Training Observations		Testing Observations		Num. Sessions	Note
	Targ	NTarg	Targ	NTarg	Tr | Te	
1	3840	9600	400	1000	4 | 1	**Tr:** n.2 sessions * 50 runs, n.2 sessions * 30 runs *covering all 12 choices* **Te:** 20 sessions * 10 runs *randomly selected elements*
2,3,4,5	3600	9000	400	1000	4 | 1	**Tr:** n.3 sessions * 40 runs, n.1 session * 30 runs *covering all 12 choices* **Te:** 20 sessions * 10 runs *randomly selected elements*
6	3360	8400	480	1200	3 | 2	**Tr:** n.3 sessions * 40 runs, n.1 session * 30 runs *covering all 12 choices* **Te:** 2* n.12 sessions * 10 runs *covering all 12 choices*

In this context, the *session* is defined as the measurement procedure that leads to the coverage of all 12 possible choices. Each training session is composed of N runs*2 target flashes*12 choices + N runs*5 not-target flashes*12 choices.
In the testing context, the *run* is defined as the sequence of 7 flashes in which all the rows and columns of the 3x4 matrix are lighted, while the *session* is defined as the total of the N runs. In test environment at least 20 choices with 10 runs have been assessed.
Each testing session is composed of N runs*2 target flashes*M choices + N runs*5 not-target flashes*12 choices.
In the proposed application, to address the problem of Target-Not Target imbalanced datasets, a mixed approach undersampling/oversampling has been adopted. First, data from both datasets have been undersampled via NearMiss approach from imbleam Python library. Next, the minority class has been oversampled through the RandomOverSampler method from the same library.

Cognitive status extraction. The datasets and acquisition methods provided by the P300-based BCI can be extended - directly - to the cognitive status extraction engine. In this context, each acquired observation concurred with the extraction of the user's cognitive status. Indeed, no training is needed for the t-RIDE method presented in Sec. 3.5. Currently, for every 16 targets observations (8 runs) a P300 characterization for clinical support is provided.

5.1.1.3 EEG-based Emotion Recognition

The training/testing datasets related to the emotion recognition inference system (Sec. 3.6) are from seven subjects, who underwent the related protocol at Design of Electronic Integrated Systems Lab, Politecnico di Bari, Italy (mean age: 30 ± 4, M:F ratio 7:0).

Acquisition Timing Distribution. The training and testing sessions have been carried out on different days and at different moments of the day (early morning, late morning, before/after lunch, and afternoon).

Subject Setup. The experiment was conducted in a laboratory environment where the subjects sat in a comfortable chair placed in a not prescribed area of the room. The subject wears the EEG headset.

Environmental Conditions. The environmental conditions were not controlled allowing persons in the surroundings to discuss, talk, move, and so on, recreating a more suitable real-life scenario. However, no environmental disturbances were recorded during the acquisitions. Different levels of lighting have been adopted during the tests.
Table 5.3 summarizes the characteristics of the proposed dataset.

Table 5.3 – Training/Testing Dataset Characteristics for the emotion recognition system

Sub.	Training Observations	Testing Observations	Num. Sessions Tr | Te	Note
1..7	100 (mixed video and conversational tasks)	30 (mixed video and conversational tasks)	5 | 3	**Tr:** 5 sessions * 20 emotion stimulations **Te:** 3 sessions * 10 emotion stimulations.

Dataset Acquisition. As presented in Sec. 3.6, the proposed emotion recognition system is preceded by a training phase called in the calibration section. This calibration consists of submitting short video clips or conversational tasks to the subject, who wears the EEG headset. These videos or conversations are aimed at eliciting specific emotions in the subject involved in the evaluation. The chosen video clips are part of the database provided for this purpose by [135] and consist of ~30 s clips of music videos or movie scenes. Conversational tasks, on the other hand, are short sentences related to weather, happiness, hopes, disappointment, etc. The impact on users of such conversational tasks has been studied in the context of the AMICO project but is beyond the remit of this thesis.
During calibration, the robot randomly stimulates the subject, approaching him/her and starting an interaction either conversational or via tablet (i.e., video). EEG data are then collected in subsets of 33 s per stimulation as prescribed by Sec. 3.6.1.1. Once the stimulation is over, a self-assessment manikin is presented on the Pepper tablet. It consists of expressing a rating from 1 to 9 (integer) of the three main parameters of emotions' characterization: arousal, valence, and dominance. Future works will concern the introduction of a dedicated BCI for the self-assessment manikin completion.

5.1.2 Robot Routines

5.1.2.1 Distance for Object Approaching

The first analysis conducted in the assessment framework for robot routines concerns the identification of the maximum distance to ensure stable object detection via the YOLO detector. This analysis permits the definition of the *approach distance.*

The approach distance is here defined as the maximum distance needed to recognize "from afar" the needed object. In this situation, the robot system builds a first bounding box useful for the scanning procedure shared among the two manipulation routines embedded in the MONOCULAR framework as per Sec. 4.3. This distance defines the so-called approach area introduced in Sec. 2.2.1.

Specifically, seven approaching distances have been evaluated in the range 1 m - 2.5 m (step 0.25 m), because above 2.5 m no stable measurements were recorded, and because choosing distances below 1 m invalidates the approach's usefulness.

Ten runs per distance step contributed to the dedicated dataset, which contains data from a total of 70 approaching operations.

5.1.2.2 Shelf Height Range

The second analysis concerns the range of shelf heights for a safety manipulation approach. For this purpose, a pharmaceutical package is placed in its nominal position on a shelf with a variable height parameter.

The shelf height is chosen to be compatible with the target height, h_{tg}, to be tested.

In this respect, fourteen different h_{sh}, ranging from 64 cm to 90 cm with a step of 2 cm, have been set.

For every height value, 10 grabbing attempts per routine have been carried out. Next, the arm positioning phase is assessed in terms of grabbing accuracy. This parameter is defined as the success rate in number of successfully picking up and scanning operations over the 10 runs (e.g., 1 grab over 10 corresponds to a grabbing accuracy of 10%).

Approaching and grabbing routines start from a nominal approach distance of 1 m.

The dedicated dataset contains data from a total of 280 manipulation operations (140 per analyzed routine) with different shelf heights.

5.1.2.3 Routine #1: Hand-vs-Object Coverage

To define an optimal hand-vs-object coverage parameter for Routine #1 (according to Sec. 4.3.3) it is necessary to carry out a dedicated analysis.

As previously stated, this parameter refers to the blob pixels that remain uncovered after the hand positioning with respect to the initial blob. The uncovered blob corresponds to a remaining hand-vs-object coverage of 100 %.

For the purpose, fifty runs were recorded per each analyzed remaining coverage percentage starting from 20% (the object blob is covered by the hand for 80%) up to 70 % (covered for the remaining 30% of the initial blob). The selected step was 2 %. The grabbing accuracy refers to the number of successful grabbings over the 50 runs.

The approaching and grabbing routine starts from a nominal approach distance of 1 m with an h_{sh} of 76 cm.
The dedicated dataset contains data from a total of 1300 grabbing operations considering different hand-vs-object coverage values.

5.1.2.4 Routine #2: Bounding Boxes Overlap Percentage

To characterize the best bounding boxes' overlap percentage for Routine 2 (according to Sec. 4.3.4), a dedicated analysis has been carried out.
As stated, the bounding boxes overlap consists of counting the number of pixels shared among the bounding box of the hand selected for the grab and the object box during the manipulation procedure.
Different bounding boxes' overlap percentages have been assessed in terms of grabbing accuracy. For the specific case, 11 different values of percentage have been considered, starting from the 5%, up to 55%. A step of 5% has been chosen due to the difficulty in achieving a precise percentage value below this limit. For values above 55% in most of the cases, the package fell or started shifting solidly with the hand, making the assessment useless.
For testing purposes an h_{sh} of 76 cm has been chosen. The arm segments were always set with the following angles: α_{S1}=37.7° and α_{S2}=0°.
For every bounding box overlap percentage, fifty identical runs have been carried out.
The dedicated dataset contains data from a total of 550 grabbing operations considering different bounding boxes overlap percentage values.

5.2 BCI Performance

5.2.1 Performance metrics

The performance of the BCIs presented in the context of this thesis work will be presented in the following by means of the four metrics of discrimination accuracy:

$$Precision\ (\%) = 100 \cdot \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{20}$$

$$Recall\ (\%) = 100 \cdot \frac{TP}{TP + \mathrm{TN}} \tag{21}$$

$$Accuracy\ (\%) = 100 \cdot \frac{TP + \mathrm{TN}}{TP + TN + \mathrm{FP} + \mathrm{FN}} \tag{22}$$

$$F1 - score\ (\%) = 100 \cdot \frac{2TP}{2TP + \mathrm{FP} + \mathrm{FN}} \tag{23}$$

where TP is the number of correctly classified trials belonging to the positive class (e.g., P300 trials), TN is the number of correctly classified trials belonging to the negative class (e.g. not-P300 trials), FP is the number of trials belonging to the negative class that is misclassified as the positive class (e.g., not-P300 trials classified as P300 trials) and, similarly, FN is the number of trials belonging the positive class, misclassified.

Another metric used to characterize a BCI system in terms of speed is the information translate rate or ITR. This parameter provides an idea of the number of bits/minutes that the BCI architecture can provide, taking into account the classification accuracy. ITR can be defined as:

$$ITR = \frac{60\left(\mathrm{P} * \log_2 \mathrm{P} + (1-\mathrm{P}) * \log_2 \frac{1-\mathrm{P}}{\mathrm{N}-1} + \log_2 \mathrm{N}\right)}{\mathrm{T}} \tag{24}$$

where N represents the number of recognizable classes, P represents the accuracy rate in choice recognition and T is the time needed for choice recognition.

5.2.2 MRP-based BCI Performance

5.2.2.1 MRP-based BCI Discrimination Accuracy

To characterize the of MRP-based BCI performance, let us consider the test dataset presented in Sec. 5.1.1.1 and the performance metrics provided by the eqs. (20)-(24).
To define the above metrics, let us consider the movement of the left hand (LH) as the positive class and the movement of the right hand (RH) as the negative class.
Fig. 5.1 reports a histogram representation including the four considered discrimination metrics versus involved subjects.
Reported data shows an overall accuracy of 84.07 % ± 3.23 %, precision of 85.6 % ± 5.14 %, recall 82.43 % ± 6.24 % and F1-score of 83.77 % ± 3.41 %.
The evaluated discrimination metrics are in line with the state of the art of the main 2-choice BCIs on a single trial [157]. However, as already introduced in Sec. 3.3, this is a first-of-a-kind MRP-based BCI application for service selection, and this does not allow a direct comparison with other BCIs with similar characteristics.

5.2.2.2 MRP-based BCI ITR

To provide a platform-independent assessment of the BCI speed in terms of *commands/minute*, it is possible to analyze the ITR parameter as defined by eq. (24).

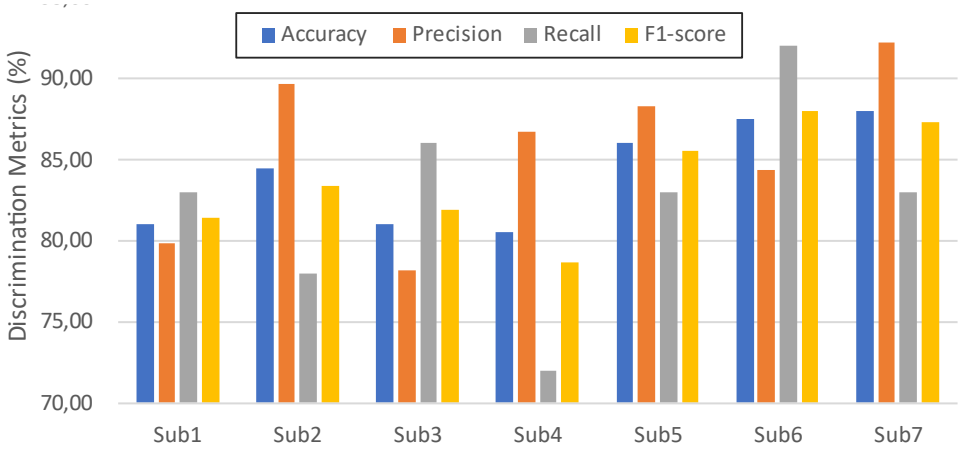

Figure 5.1 Discrimination metrics for MRP-based BCI (Test dataset in Sec. 5.1.1.1) versus Subject

For this purpose, it must be considered that for this BCI application the parameter N (the number of possible choices) is limited to 2. The T parameter (i.e., classification period) has been set to 2 s for the testing phase, without prejudice to the possibility of reducing this value to 1 s between one choice and another, improving the ITR.
Table 5.4 summarizes the ITR considering both T=1 s and T=2 s.
Results in Table 5.4 show that the MRP-based BCI can provide, on average, 11.17 ±4.68 commands/min at T=2 s and 22.34 ± 2.34 commands/min at T=1 s. The ability to nest about 11 choices per minute opens several formalizable choices of 2^{11}, corroborating the fluid capabilities as selective BCI of MRP-based neural interfaces.
In this context, it is useful to specify that even BCIs based on P300 able to provide N = 36 choices on a single selection screen are not capable to overcome an ITR of 7 commands/min [106]. For P300-based BCI, this parameter can grow up to about 14 commands/min with the accumulation of at least 4 trials [107] with interstimulus intervals < 300 ms (not applicable for cognitively impaired people).

Table 5.4 – MRP-based BCI ITR

ITR (comm./min)	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7
T=2 s*	8.95	11.33	8.95	8.64	12.47	13.69	14.11
T=1 s	17.91	22.67	17.91	17.29	24.95	27.39	28.24

* default.

5.2.3 P300-based BCI Performance

5.2.2.1 P300-based BCI Discrimination Accuracy

To characterize the P300-based BCI performance, let us consider the test dataset presented in Sec. 5.1.1.2 and the performance metrics provided by the eqs. (20)-(24).

Single Trial. Specifically, the above metrics are useful to define a single-trial response of the proposed BCI, i.e., the ability to respond to a single stimulation run (single flash of each row and column according to Sec. 5.1.1.2). Although this is not the default operating mode of the proposed P300-based BCI, which must accumulate at least 3 runs to provide an inference on the chosen element, this section provides a useful single trial overview of the realized NN.
To define the above-proposed metrics, let us consider the P300 trial as the positive class and the not-P300 trial as the negative class.
Considering the *single-trial* context (i.e., each trial returns binary inference P300/not-P300), the BCI approach proposed in this thesis returns the metrics in Table 5.5. Table 5.5 shows accuracy, precision, recall, F1-score for each subject involved in the test. It also reports the NN topology chosen by the topology selector according to Sec. 3.4.1. The nomenclature adopted for NN topology shall be understood as follows:

$$\mathrm{FC}\ \mathrm{N_{U0}}\ \mathrm{x}\ \mathrm{U_0(AF_{U0})} -\ \mathrm{N_{U1}}\ \mathrm{x}\ \mathrm{U_1(AF_{U1})} - \cdots \tag{25}$$

where FC identifies a fully connected topology (currently the only type of layer available in the Tuner), U_i are the number of units per hidden layer, while i depends on the position in the NN chain, AF_{Ui} is the activation function for the specific hidden layer (it can be R for ReLU, or T for Tanh). Finally, N_{Ui} is the number of hidden layers with the same characteristics. For example, FC 128(T) - 2x32(R) is a fully connected NN composed of a first hidden layer with 128 units and a Tanh activation function, followed by the cascade of 2 hidden layers with 32 units and ReLU activation function. It is important to remember that according to Sec. 3.4.1 the NN has an input and batch normalization layer before the above-reported topology and a single unit layer with Sigmoid activation as the output layer.

Table 5.5 – Discrimination Metrics for P300-based BCI in a single trial context

Sub #	Accuracy	Precision	Recall	F1-score	NN topology
Sub1	74.29	54.59	59.50	56.94	FC 3x32(R)
Sub2	77.00	58.23	69.00	63.16	FC 128(T) - 2x32(R)
Sub3	75.07	55.70	62.25	58.80	FC 2x32(R)
Sub4	75.00	55.98	58.50	57.21	FC 2x64(R)
Sub5	76.57	57.32	70.50	63.23	FC 3x32(R)
Sub6	76.55	57.26	70.63	63.25	FC 2x64(R) - 32(R)

The values of the extracted discrimination metrics are not directly comparable with state-of-the-art solutions with similar characteristics such as CNN-1 [108], BN3[107], and ConvLSTM [158] already introduced in the summary of solutions provided in Sec. 3.4.
The main reason lies in the input data necessary for the operation of the NN listed above[5] and the used stimulation protocol (4x3 speller matrix versus 6x6 one, with different interstimulus intervals). Indeed, all the above-reported topologies require EEG signals from 64 channels, while the here proposed approach analyzes data from 8 EEG channels. Nevertheless, the values reported in Table 5.5 show a promising trend able to overcome the state-of-the-art counterparts.
For comparative purposes, the proposed method has been applied to the BCI competition III: Dataset II (6x6 speller matrix) by [109] dataset and specifically to Subject B, which returns the best performance for all classifiers considered [107,108,158]. The training dataset provides 2550 target trials and 12750 not-target ones [109][6].
Table 5.6 shows a comparison of the discrimination parameters for the various state-of-the-art classifiers applied to the proposed test dataset which consists of 3000 target observations and 15000 not-targets.
The proposed approach has allowed the definition of an NN-based classifier able to improve the recall parameter by +10% and the F1-score by +6% if compared to the best state-of-the-art approach considered for the present analysis. Accuracy and precision are consistent with other proposed solutions.

[5] Different input data make it impossible to reconstruct an inference on the proposed data by reproducing methods [107,108,158].
[6] https://www.bbci.de/competition/iii/desc_II.pdf

Table 5.6 – Discrimination Metrics for P300-based BCI in a single trial context applied to BCI competition III: Dataset II

Sub #	Accuracy	Precision	Recall	F1-score
This work	**82.95**	**49.14**	**66.07**	**56.36**
CNN-1	81.03	44.67	57.83	50.41
BN3	80.14	41.50	46.70	43.95
ConvLSTM	**82.94**	**48.36**	34.50	40.27

Matrix Element Recognition. The nominal operation of this BCI requires the accumulation of at least 3 runs before predicting the element of the 4x3 speller matrix proposed in Sec. 3.1.3.2. In the context of the P300 spellers, it is necessary to introduce an additional parameter of accuracy which is the character recognition accuracy, or in this case, the *element recognition accuracy*.

This parameter returns the BCI's ability to recognize an element of the speller matrix varying the number of runs.

This parameter is extracted, according to Sec. 5.1.1.2 by administrating, during the test, to the subject 10 runs with a specified (randomly selected) target. The procedure is repeated for at least 20 sessions (of 10 runs).

Considering the i-th session, the correct recognition at the relative run is identified as 1, otherwise 0. The percentage of correct recognition, on at least 20 sessions, defines the *element recognition accuracy.*

Fig. 5.2 shows the mean element recognition accuracy, calculated on all the involved subjects. Error bars, on the other hand, represent the standard deviations calculated on the same set. Table 5.7 reports the element recognition accuracy for each involved subject.

The experimental results show that the accuracy provided by the introduced method remains stably above 80% already after 4 runs (~14 s).

As observed in the literature [106], exceeding this threshold is normally expected around 6-8 running the best cases (Subject B BCI III dataset) [107]. This convergence speed has positive implications for the BCI information translate rate.

5.2.2.2 P300-based BCI ITR

To provide a platform-independent assessment of the BCI speed in terms of *commands/minute*, the ITR parameter as defined by eq. (24) is analyzed in the following. For this purpose, it must be considered that for this BCI application the parameter N (number of possible choices) is 12. The T parameter (i.e., classification period) has been defined as:

$$T = T_{ISI} * N_{flash} * i_{run} \tag{26}$$

where T_{ISI} is the interstimulus interval (T_{ISI} = 500 ms), N_{flash} is the number of flashes necessary to complete a single run (N_{flash} = 7), and i_{run}=1,…,10 is incremented according to the number of runs collected to provide the inference.

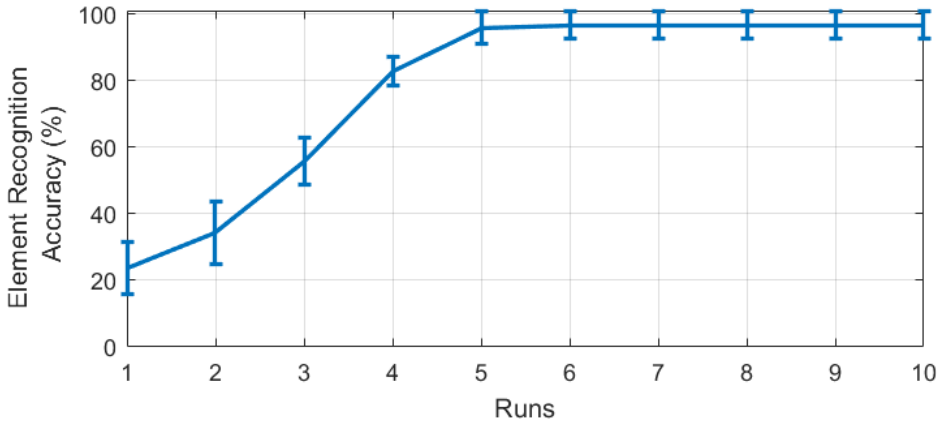

Figure 5.2 Mean Element recognition accuracy of proposed P300-based BCI versus run. Error bar: standard deviation on all the involved subjects

Table 5.7 – Element recognition accuracy of proposed P300-based BCI per subject

Sub #	Runs									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sub1	20	30	50	85	95	95	95	95	95	95
Sub2	30	35	60	80	100	100	100	100	100	100
Sub3	15	25	55	85	90	90	90	90	90	90
Sub4	15	25	45	75	90	95	95	95	95	95
Sub5	30	45	60	85	100	100	100	100	100	100
Sub6	31,82	45,45	63,64	86,36	100	100	100	100	100	100

Fig. 5.3 reports the ITR calculated on the mean element recognition accuracy versus runs collected to provide the inference. Table 5.8 shows ITR values subject by subject.
On average, with 4 runs, the proposed system can guarantee 15.31 comm./min. Despite the number of choices limited to 12 (against 36) and an interstimulus of 500 ms, higher than similar state-of-the-art solutions (300 ms), the ITR parameter calculated on the average of the accuracies allows +1 comm./min if compared with the fastest solution proposed by the authors of BN3[106,107].

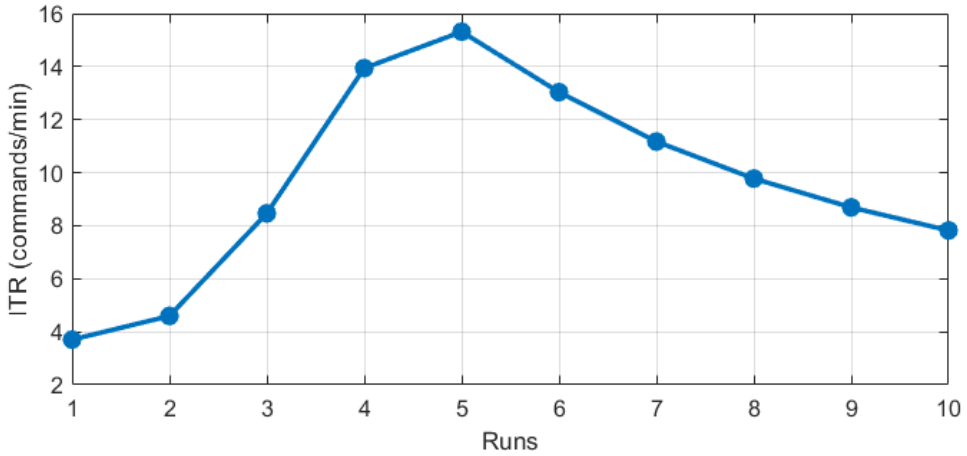

Figure 5.3 ITR calculated on the mean element recognition accuracy versus number of run.

Table 5.8 – P300-based BCI ITR per subject

Sub #	Runs									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sub1	2.29	3.38	6.84	14.74	15.00	12.50	10.72	9.38	8.33	7.50
Sub2	6.77	4.83	9.84	13.03	17.21	14.34	12.29	10.75	9.56	8.60
Sub3	0.83	1.27	3.63	4.00	4.43	5.32	6.20	7.09	7.98	8.86
Sub4	0.83	0.89	1.68	3.52	3.86	5.23	6.10	6.97	7.85	8.72
Sub5	6.77	8.27	9.84	14.74	17.21	14.34	12.29	10.75	9.56	8.60
Sub6	7.77	8.45	11.05	15.23	17.21	14.34	12.29	10.75	9.56	8.60

5.3 P300-based Cognitive Status Extraction

To characterize the P300-based cognitive status extraction system performance, the test dataset presented in Sec. 5.1.1.2 should be considered.

The system is based on the *t-RIDE* algorithm, which has been widely investigated in terms of performance in our previous works [120, 123].

The advantages of this method in the ERP waveform reconstruction are briefly summarized in the following, referring to the details of the associated search [120, 123].

According to [123], comparing t-RIDE, RIDE, ICA, and Grand Average (GA) in a P300 amplitude reconstruction (starting from a known clean P300 waveform spoiled in EEG background noise):

- t-RIDE and ICA converge to the same P300 amplitude results with an error of 0.1%.
- t-RIDE is, on average, +12.3% more accurate than a RIDE (without tuning);
- The P300 amplitude values extracted through GA value were 67% lower than the t-RIDE one.
- t-RIDE is more accurate than the other methods. t-RIDE showed the highest accuracy using 25 target trials: t-RIDE = 96.05%; RIDE = 78%; ICA = 75.9%; GA = 51.6%.
- t-RIDE needs fewer targets to reach 90% amplitude accuracy if compared with competitors. The number of target stimuli to reach 90% accuracy are: t-RIDE = 16; RIDE = 38; ICA = 30; GA = n.d. (GA never reaches 90% accuracy).
- To extract the P300 amplitude with a 90% accuracy, the time duration of the task has to be: t-RIDE = 28 s; ICA = 52.5s; RIDE = 66.5s considering an interstimulus interval T_{ISI} = 500 ms.

In terms of P300 latency reconstruction (starting from a known clean P300 waveform spoiled in EEG background noise), the following results are reported in [120, 123]

- t-RIDE is, on average, +1% more accurate than RIDE and +3% more accurate than ICA, although they exhibit the same convergence behavior.
- The GA converged value was 12% lower than the t-RIDE one.

- t-RIDE needs the same number of targets to reach 90% accuracy when compared to ICA and RIDE: t-RIDE = 6; RIDE = 6; ICA = 6. The number of targets to reach 90% accuracy for latency calculation with GA is 18.

For the above-mentioned reasons, for every 16 target observations (8 runs) a P300 characterization for clinical support is provided by the cognitive status extraction system as per Sec. 5.1.1.2.

Considering the test session, formally 400 target trials and 1000 not-target trials should be evaluated for subjects 1 to 5 and 480 target and 1200 not-target trials for Sub. 6.

However, since this section works downstream the BCI, this latter will provide the classification between target and non-target trials and redirect the weighted mean procedure by t-RIDE (oriented to P300-containing trials).

In this context, it should be specified that the single-trial accuracy of the P300-based BCI is essential as it supervises the labeling of the analyzed trials. Fig. 5.4 shows the confusion matrices associated with the 6 subjects involved, to extract the datasets used by the cognitive status extraction system. The resulting datasets are distributed according to Table 5.9. Table 5.9 reports the total trials labeled as targets and the corresponding part of TP and FP. Results in Table 5.9 show that it is necessary to create classifiers that maximize TP and reduce FP, i.e., that have as their design objective a high F1-score (maximizing the harmonic mean between precision and recall). This is possible, for example, by introducing weights in the classification of the class in minority (P300). The last column of Table 5.9 returns the number of extractions provided during the trial (i.e., the total number of target trials / 16 trials)

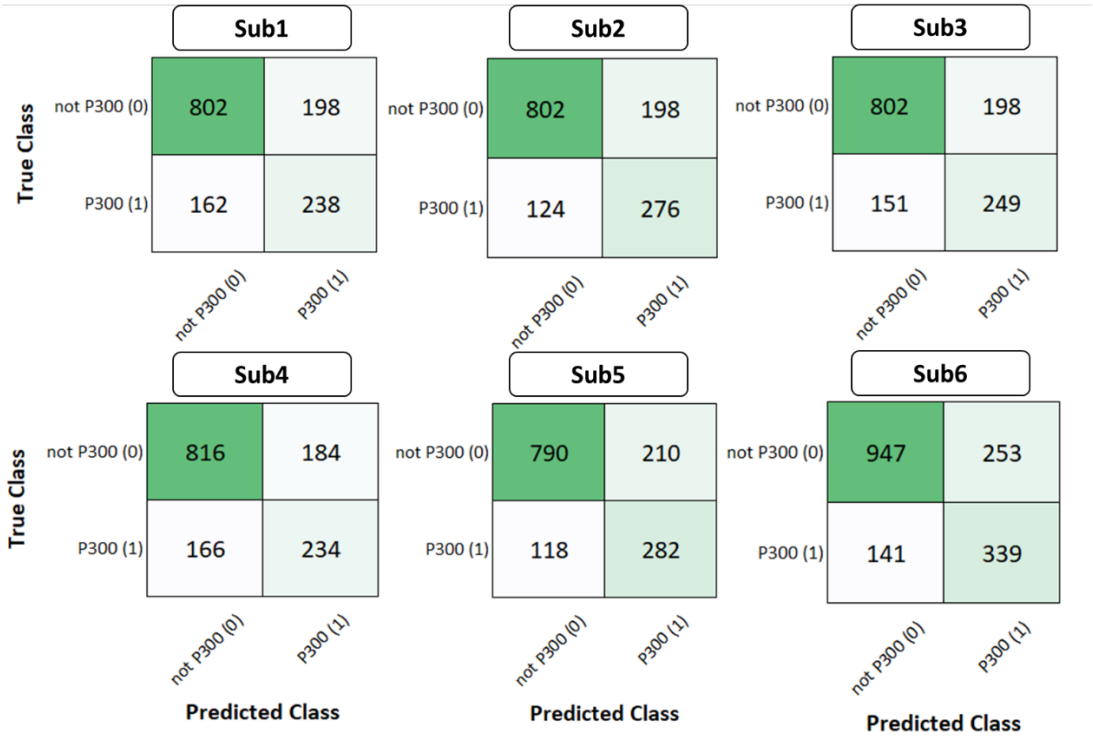

Figure 5.4 Confusion Matrices of P300-based BCI for single trial labelling employed by the Cognitive status extraction system

Table 5.9 – Dataset employed by the Cognitive status extraction system based on P300-based BCI labelling

Sub #	Tot Targ	TP	FP	Provided Cognitive status
Sub1	436	238	198	27
Sub2	474	276	198	29
Sub3	447	249	198	27
Sub4	418	234	184	26
Sub5	492	282	210	30
Sub6	592	339	253	37

On the dataset described in Table 5.9 has been carried out the extraction of the parameters for the spatio-temporal characterization of the P300 component: amplitude, latency and FoM (eq. (3)).

Fig. 5.5a shows the {P300 Amplitude, P300 Latency} plan, reporting for each evaluated subject the average value of amplitude and latency of the P300 component. Each amplitude-latency coordinate relative to the subject is supported by errobars that indicate the latency and amplitude ranges extracted from the various evaluations provided during the test. In addition, Fig. 5.5a shows the risk thresholds introduced by Frodl et al. [121] to discriminate based on the P300 component, HCI and MCI (see Sec. 3.5.1).

Fig. 5.5b shows the trend of FoM versus the involved subjects. This subplot also embeds the risk areas.

The plots in Fig. 5.5 show that in most cases (93% of the evaluations), the extracted parameters (i.e., latency, amplitude, FoM) fall within the physiological sections relating to healthy subjects, demonstrating even in the presence of datasets that are not perfectly balanced. Application of the methodology for people with MCI and HCI is detailed in [120, 123]

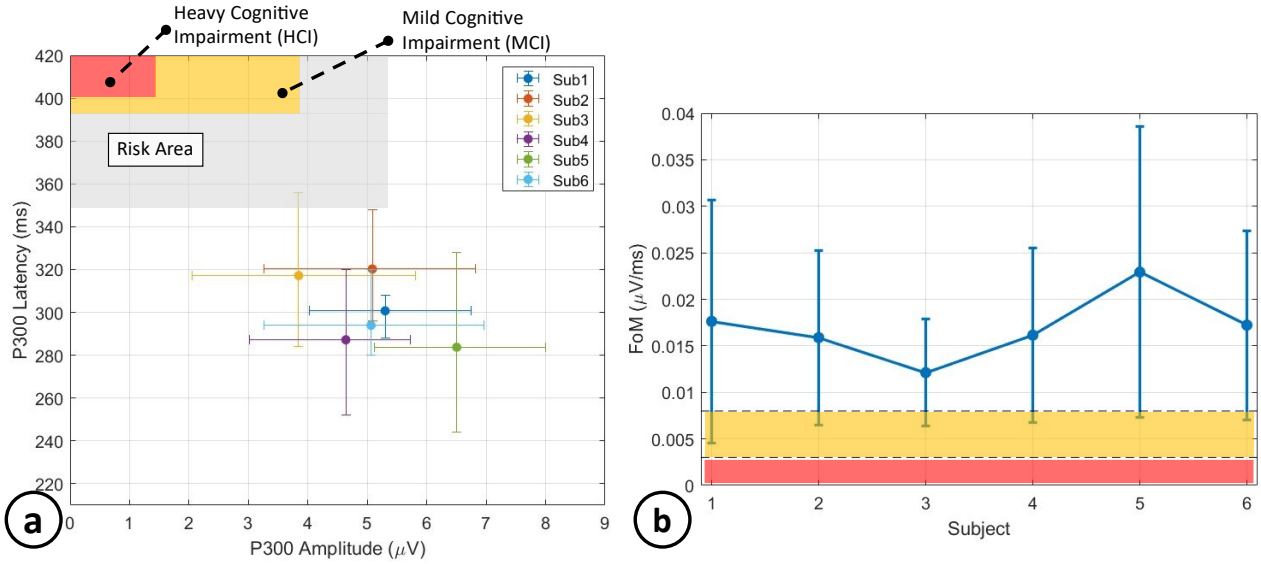

Figure 5.5 Cognitive status extraction system. (a) P300 amplitude – latency plan with highlighted risk areas. Errobars: latency and amplitude ranges. (b) FoM versus subjects. Errobars: worst/best FoM.

5.4 EEG-based Emotion Recognition

To characterize the of EEG-based Emotion Recognition system, the test dataset presented in Sec. 5.1.1.3 has been considered. It consists of 700 training and 210 test observations from 7 subjects.

5.4.1 Multi-Optimization based Calibration: A Case Study

As proof of concept, let us consider the same memory and resource configuration that Pepper robot reserves to the built-in task for the mood recognition [29]. The current built-in task can discriminate among 3 different moods: positive, negative, and unknown. For the purpose, Pepper employs external indicators such as facial expression reconstruction via RGB camera and voice analyzer. The chosen target leads to the following user-defined setup: 512 kB (i.e., RAM+ROM) of memory usage, 200 LoC (max) per routine, and cyclomatic complexity of the functions lower than 300.
The subject selected for the case study is the subject #2 (randomly selected).
Acquired data undergo the computation pipeline described in Sec. 3.6 and summarized in the following commented pseudocode.

Pseudocode 3.1: Calibration Routine

```
%% ----------- Initialization -----------
load user_cstr := (thr1,2 … 4)% set the design constraints for the system and each software metric
load FE_DB := (C1, C2 … C8191)% contains the MATLAB Coder generated script (execution efficiency)
load EEG_trials_TrainingSet := Trials Matrix % contains the 33s EEG trials composing the training set
load EEG_trials_TestingSet := Trials Matrix % contains the 33s EEG trials composing the testing set
load OnlineEEG_in := Value % set the fixed dimension of EEG input in real-time application (see Sec. III.C)
load Pred_out := Value % set the fixed dimension of output prediction
%% ----------- Routine -----------
run [par1, par2, par3, par4]= SoftwareMetricExtraction(FE DB) %returns the software metrics (par1..4) for each combination Ci in FE DB
run [sComb]=DesignConstraints_Check(par1, par2, par3, par4, user_cstr) %returns the FE techniques combinations that singularly respect the design constraints on software metrics (sComb).
run [Score] = ScoreExtraction(sComb) %returns the Score value for each suitable FE technique combination.
run [Acc] = AccuracyExtraction(sComb, EEG trials TrainingSet) %returns the cross-validated accuracy value for each suitable FE technique combination.
compose SA_points = [Score, Acc] % define a point for each combination in the 2D space {Score, Accuracy}
run optC = MultiOptimization(SA_points) %returns the optimal FE technique combination.
run [F] = FeatureSelection(optC, EEG trials TrainingSet) % embeds the chosen feature selection technique (e.g., NCA) to reduce the classification problem dimensionality.
run memCL=MemoryCount(OnlineEEG_in, F, Pred_out, user_cstr) % returns the memory available for classifiers implementation
run CL=CommitteeSel(EEG trials TrainingSet, memCL) % returns the best classifiers committee based on the accuracy values and available memory
transmit →Robot (EEG_trials_TestingSet, F, CL)
```

Based on the FE DB content (see Table 3.1 in Sec. 3.6.1.2), the multi-optimization routine step resulted in an optimal combination of 7 types of time and frequency domain features, of which 5 features exploit the time domain, while 2 features are based on the frequency domain.
For the sake of completeness, Table 5.10 summarizes the first eleven combinations that passed the complexity assessment and the design constraints check. Table 5.10 reports the combinations identifiers (given by the composition of IDs from Table 3.1), the Score before the feature selection step, and the related achieved accuracy.
Specifically, the selected combination (i.e., FIJKLM) consists of seven time-domain features that can be divided into three types: statistical features, Hjorth features, and HOCs ones according to the definition in [142] and [143], respectively.
Concerning the first type, the system extracts five known statistical features: mean, median, standard deviation, maximum, and minimum. This statistics-based FE procedure is applied to the original signals over the EEG groups and their 1st and 2nd derivatives. Features in the frequency domain concern the EEG spectrum on the bands falling within the range of 4 Hz - 35 Hz. The EEG rhythms intervening in this range are θ (4-8 Hz), slow α (8-10 Hz), α (8-12 Hz), β (12-30 Hz), and γ (>30 Hz). In the emotion recognition context, for each band, the Power Spectral Density (PSD), along the considered channels, has been extracted. Thus, for each analyzed band, five statistical features of PSD data have been evaluated: mean, median, standard deviation, maximum, and minimum. Finally, the mean and the maximum of the PSD logarithm have been also included in the grid search analysis.

Table 5.10 – Multi-optimization routine analysis

Comb. ID	Score	Acc (%)*
FIJKLM	0.31	79.69
BCIK	0.25	74.38
FGHI	0.24	74.27
EJ	0.36	73.96
FJKL	0.31	71.98
ABJ	0.42	71.77
A	0.17	61.98
B	0.08	61.56
DH	0.26	60.52
EI	0.15	58.33
GKL	0.02	54.69

* Cross-validated accuracy based on k-fold validation with k=5

The above-described FE method combination all over the grouped EEG waveforms results in a total of 270 features. The selected combination of FE techniques requires, on average, 159.3 kB over the 512 kB constraints. No single functions overcome 150 lines of code, and the set cyclotomic complexity (i.e., 300).
The 270 features from the complexity and accuracy assessment blocks are, thus, sent to the feature selection algorithm, i.e., the NCA.

Fig. 5.6 shows an experimental example of the NCA-based weighting procedure. The memory consumption related to the feature extraction stage recorded a reduction of 83%, resulting in memory usage of 27.01 kB and a total of 19 selected features all over the AVD model parameters.

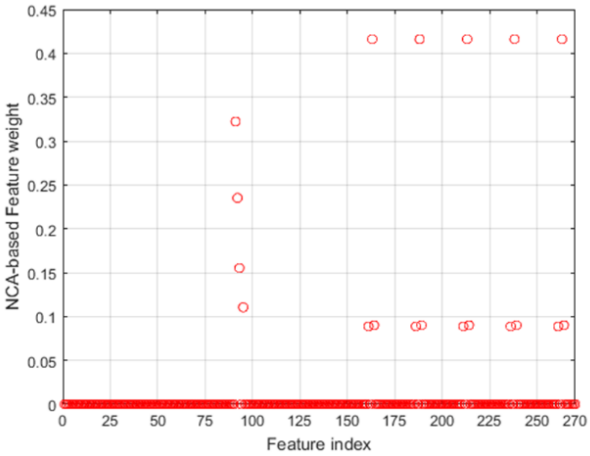

Figure 5.6 NCA-based feature weights for all the features of the selected subject on the Valence parameter

The system trained all 16 classifiers composing the last step of the calibration procedure. Specifically, the considered classifiers are trained by means of a set of MATLAB® scripts generating compact model structures. These structures are then codified in C language through the MATLAB® Coder App with a focus on memory efficiency.
The resulting structs are analyzed to find the best tradeoff between accuracy and memory usage, setting the memory constraint for the classifier to 310 kB.
Table 5.11 summarizes the accuracy achieved in cross-validation for each analyzed classifier. Specifically, an accuracy value per each AVD model parameter has been reported to favor the classifiers' committee selection comprehension. Table 5.11 also shows the size in kB extracted via `whos.Bytes()` method. This metric concerns all the classifiers trained (one per each AVD model parameter).
The last column shows those classifiers category marked as Out of Memory (OoM) and, thus, excluded for the online implementation.

For the specific application, the selected committee is composed of a QSVM for both valence and dominance and an MGSVM for arousal parameters. Overall the used memory is about 245.55 kB.

5.4.2 Online Emotion Recognition: Architecture Performance

The described architecture has been oriented to be suitable for AVD model parameters discrimination because the use of this 3D model octants permits to classify of up to 8 emotions: anger, joy, surprise, fear, unconcern, satisfaction, calm, and sadness as stated in Sec. 3.6.1.3. To obtain the multiclass prediction over 8 emotions, all the classifiers composing the committee must operate in parallel providing binary discriminations that will determine the arousal, valence, and dominance levels (i.e., high or low).

Table 5.11 – Classifiers committee selection

Class.	Accuracy (%)			Size (kB)			Note
	A	V	D	A	V	D	
GNB	60.71	67.14	64.29	905.15	894.42	912.32	OoM
KNB	64.29	71.60	57.14	1138.01	1113.21	987.44	OoM
LSVM	53.57	57.14	46.43	39.31	37.15	38.65	
QSVM	53.57	**82.14**	**79.57**	79.18	**81.31**	**80.11**	
CSVM	71.43	75.00	74.43	80.17	80.17	81.91	
FGSVM	75.00	57.14	64.29	88.59	74.21	94.25	
MGSVM	**85.71**	67.86	73.57	**84.13**	81.24	80.36	
CGSVM	61.32	61.14	57.14	81.64	78.62	83.29	
FKNN	63.57	64.29	60.71	241.58	241.58	241.58	
MKNN	74.29	60.84	64.29	241.58	241.58	241.58	
CGKNN	68.14	69.43	67.14	236.14	236.14	236.14	
CosKNN	60.71	61.14	53.57	222.47	219.36	194.25	
CKNN	67.86	66.71	57.14	241.6	241.6	241.6	
WKNN	68.29	74.11	65.71	245.19	251.62	238.18	
SKNN	67.86	**83.22**	75.00	4288.93	**4312.88**	4201.23	OoM
DNN*	**86.43**	78.29	**83.44**	**5211.30**	4982.41	**5381.24**	OoM

*n.2 dense hidden layers with ReLU activation with 16 hidden units per layer and a dense layer for the output with sigmoid activation. RMSprop optimizer and binary cross-entropy loss function have been used.

Nevertheless, the system can be also employed on a subset of the AVD model, e.g., the Arousal-Valence (AV) one. In this case, only two classifiers from the classifier committee operate in parallel, aiming to discriminate four different emotions: happiness, boredom, anger, and calm. Specifically, exploiting the 210 observations composing the test dataset (30 observations per subject), the optimal processing chains (NCA-selected features and classifier committees) all over the subjects have been tested.

In an 8-class discrimination context (i.e., AVD model), the proposed architecture showed an overall accuracy of 75.71 ± 4.6 %. Specifically, in 2/7 cases (28.57%) the accuracy achieved 80% or above. In 5 cases (71.43 %) an accuracy between 70% and 80% has been recorded.

Reducing the complexity to a 4-class discrimination context (i.e., AV model), the architecture demonstrated to be able in classifying four emotions with an accuracy of 79.52 ± 4.48 %. In 4/7 cases (57.14%) the accuracy achieved 80% or above. While in the remaining cases (3/7 – 42.86%) the architecture showed an accuracy of >73.33%.

Finally, evaluating each model parameter – singularly - the proposed architecture achieved, on average, an accuracy of 85.24% on the dominance parameter, 85.24 %, and 87.14 % for valence and arousal, respectively.

Fig. 5.7a shows the overall architecture accuracy versus involved subjects when 8 emotions are discriminated (orange solid line). While the gray solid lines show the accuracy trend over 4 emotion discrimination problem all over the subjects.

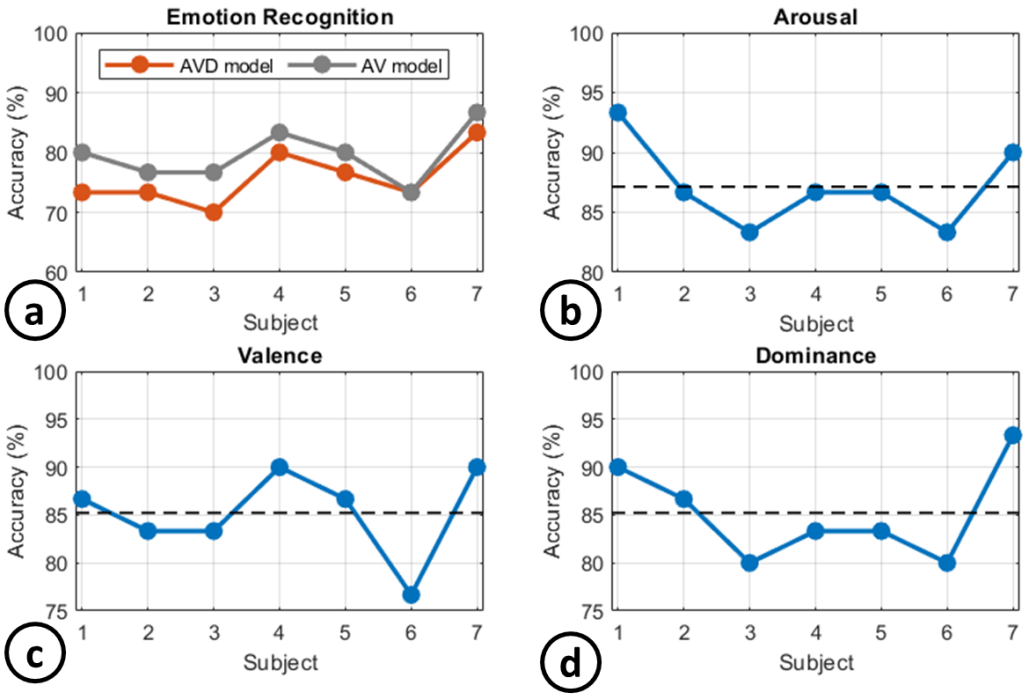

Figure 5.7 EEG-based Emotion Recognition system. (a) Accuracy on AVD and AV model; (b) Arousal discrimination accuracy; (c) Valence discrimination accuracy; (d) Dominance discrimination accuracy.

The accuracy achieved provides promising results when compared with most of the state-of-the-art solutions proposed in Sec. 3.6, although the reference dataset is not directly comparable.

To provide a consistent comparison with state-of-the-art solutions, the proposed framework was used to classify data with the same pre-processing but coming from the DEAP dataset [135]. DEAP dataset includes physiological recordings and videos of face reactions from Ns=32 subjects aged, on average, 26.9 years old (range: 19-37 years old). The recordings are from an experimental protocol in which the subjects are asked to watch at Nv=40 one-minute-long music videos capable of eliciting different arousal, valence, and dominance levels. DEAP dataset also includes the subjects' emotional ratings in terms of arousal, valence, dominance, and like/dislike, carried out by using the self-assessment manikin rating scale.

Table 5.12 shows a comparison between the previously described state-of-the-art studies and the here presented work applied to the DEAP dataset. Table 5.12 focuses on the number of classified emotions and the accuracies achieved in this multiclass context. From the data in Table 5.12, it is possible to notice that considering the AV model (i.e, 4 emotions), the accuracy achieved by the proposed method overcomes the ones proposed in [130], [128], [133], [129] in terms of the number of recognizable emotions.

On average, an improvement of +15% (ranging from +5.92% to +21.77%) has been recorded when an AV model is used to discriminate the emotions.

Table 5.12 – EEG-based Emotion Recognition System: Comparison with the state of the art on DEAP dataset.

Ref	Classified Emotions	Accuracy (%)
[132]	Happy, pleased, relaxed, excited, neutral, calm, distressed, miserable, depressed	A: 82.35 V: 79.95 D: 71.14 AVD: 65 (80% max)
[130]	Excited, relaxed, negative	A: 94.1 (max) V: 58.8 (max)
[131]	Fear, sad, frustration, happy, pleasant, satisfied	AV: 55.58
[128]	Calm, anger, happiness	AV: 60
[133]	Two binary classifiers (singularly or together): HV/LV and HA/LA	A: 79.95 V: ~80 AV: 71.43
[129]	HA/LA	A: 62.58
This work	Singularly: HV/LV, HA/LA, HD/LD AVD: Anger, joy, surprise, fear, unconcern, satisfaction, calm, sadness AV: happiness, boredom, anger, calm	A: 81.18 V: 77.31 D: 78.02 AVD: 75.12 AV: 77.35

Acronyms: H(L)V - High(Low) Valence; H(L)A - High Low Arousal; H(L)D – High(Low) Dominance

Considering the AVD models, the performance singularly achieved on each AVD model parameter, by the proposed architecture, on 8 emotions is comparable with the one proposed in [132] over 8 emotions and neutral. In this respect, even if the two platforms are not strictly comparable due to the different number of considered classes, the present architecture recorded an improvement of about +10% on the average accuracy.

5.5 Robot Routines

5.5.1 Distance for Object Approaching

As introduced in Sec. 5.1.2.1, the assessment of the proper distance for the object approach is based on a dataset composed of 10 runs per 7 distances ranging from 1 m to 2.5 m (step 0.25 m).

To assess the approaching distance, a dedicated parameter has been defined: stable object recognition accuracy (SORA).

It is defined as the number of times the YOLO bounding boxes appear over the 10 analyzed frames (~ 2 s) at a specific distance. 10 frames out of 10 with YOLO bounding box leads to a SORA= 100%, 9/10 SORA=90% and so on).

The boxplot representation in Fig. 5.8a shows that with an *approaching distance* <1.25 m the median set to values higher than the 80% of SORA, with an average of 84±12.64% of accuracy.

For enabling the blob tracking procedure required by Routine #1, a few stable bounding box repetitions can be enough, allowing - de facto- an approaching distance that ranges between 1 m and 1.75 m with no strict constraints. It ensures good flexibility to potential odometry errors in navigation.
In the same context, another distance related to Routine #1 has been extracted: *the tag extraction distance*. This distance is defined as the optimal distance to stability extract the bounding box for the hand-vs-object coverage routine.
In this case, six-step distances have been evaluated in the range of 35 - 40 cm (step 1 cm as per the odometry control resolution). This range derives from positioning constraints and can move in a little range compatible with odometry errors. Also for this case, ten runs per distance step contributed to the boxplot representation in Fig. 5.8b for a total of 60 approaching routines.
The boxplot shows that above 37 cm, the SORA starts strongly decrease. The reason lies in the data collection procedure for the YOLO training. Most of the collected images were acquired with blob tracking on and 35 ± 4 cm tag extraction distance. It reduced the classifier accuracy in the nominal operation.
Finally, an approaching distance of 1.5 m has been selected for the final implementation of Routine #1 and #2, while a tag extraction distance (with feedback control) of 35 cm is selected to proceed with Routine #1.

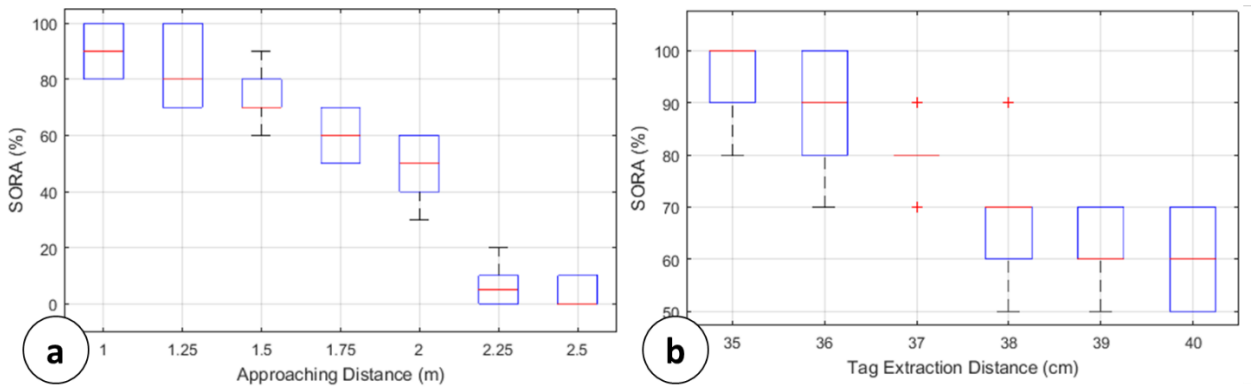

Figure 5.8 Boxplot representation of (a) stable object recognition accuracy (SORA) parameter versus the approaching distance on 10 runs per distance step; (b) SORA versus the Tag extraction distance on 10 runs per distance step.

5.5.2 Shelf Height Influence on Grabbing Accuracy

The present section aims to characterize the maximum and the minimum shelf height that ensures a good grabbing success rate for the MONOCULAR framework.
The reference dataset has been presented in Sec. 5.1.2.2 and consists of 14 h_{sh} values, ranging from 64 cm to 90 cm with a step of 2 cm. Ten grabbing per h_{sh} have been carried out to assess the procedure in terms of grabbing success rate.
Fig. 5.9 summarizes the grabbing accuracy versus the shelf height.

Results show that with shelf height included in the range that goes from 70 cm to 80 cm, the grabbing accuracy is constantly above 90%. Height values above 80 cm see a drastic decrement in accuracy. This behavior is related to the increment of α_{S1}above 68°.
In this respect, experimental analysis on α_{S1} parameter demonstrated that the brush DC coreless SE24PCTCA of Pepper is not able to keep a reliable angle value above this limit. It limits MONOCULAR routine use above 84 cm. Another limitation of the routine above 84 cm concerns the object check routine because the object tends to disappear from the bottom camera field of view.
Below 70 cm the main problems are related to the increment of α_{S2}. Increasing α_{S2} above ~10° the grabbing is not perpendicular and tends to be less effective with the considered package.
Fig. 5.10 reports the robot mechanics in three main cases: (a) minimum shelf height (i.e., 64 cm), (b) typical shelf height (i.e., 76 cm), and (c) high shelf (i.e., 84 cm).

5.5.3 Routine #1: Hand-vs-Object Coverage

To characterize the hand-vs-object coverage parameter, the dataset defined in Sec. 5.1.2.3 will be used. Briefly, 50 runs per coverage percentage value are assessed starting from 20% up to 70 % with a step of 2 %.
Fig. 5.11 shows the grabbing accuracy for the right (blue solid line) and left hand (red solid line), versus the remaining hands-vs-object coverage. Experimental results showed that with a remaining coverage that ranges between 40 % and 45 % the grab realized by means of the robot's right hand is able to achieve a success rate above 90%.

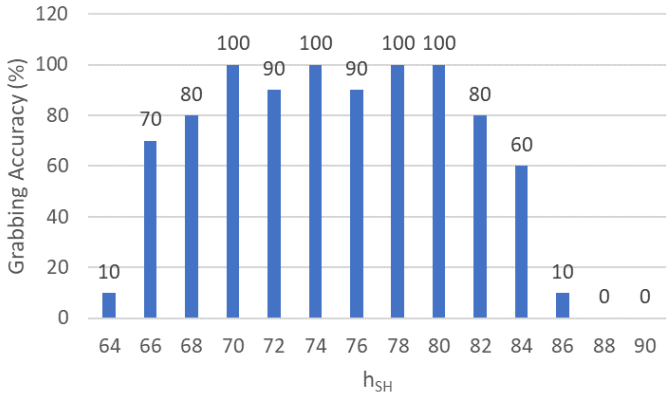

Figure 5.9 Grabbing accuracy versus shelf height

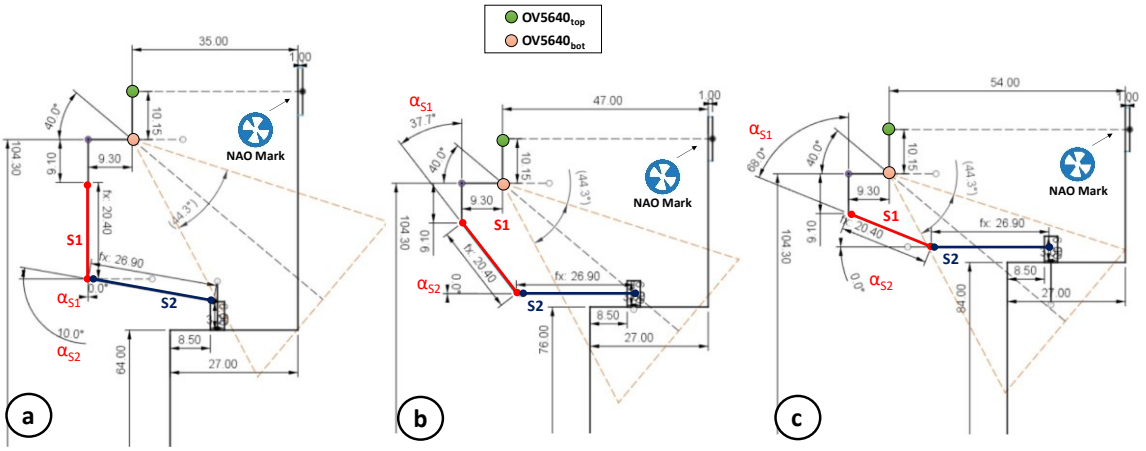

Figure 5.10 Robot mechanics in three main cases: (a) minimum shelf height, (b) typical shelf height and (c) high shelf.

Similarly, considering grabbing via left hand, the best interval falls within the remaining hands-vs-object coverage range of 25% – 32%.

The peak shift is mainly due to the 3D sensor perspective, because of its ipsilateral presence with the hand which carries out the grabbing. For the nominal operation of the current application, the above-mentioned ranges have been selected

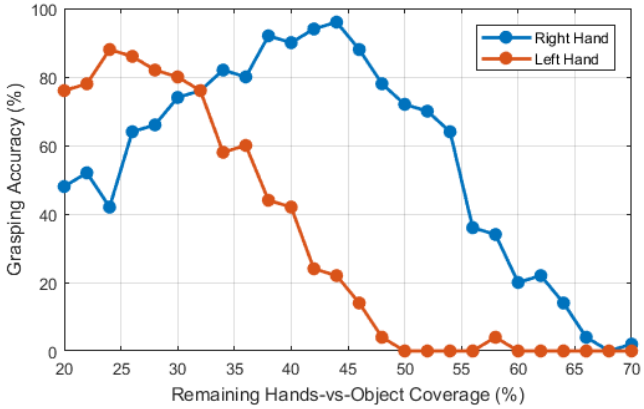

Figure 5.11 Grabbing accuracy versus Remaining Hands-vs-Object coverage percentage for grasps carried out via right and left hands.

5.5.4 Routine #2: Bounding Boxes Overlap

To characterize the optimal bounding boxes overlap percentage, the dataset presented in Sec. 5.1.2.4 will be referred. Briefly, 11 values of a percentage ranging from 5%, up to 55% with a step of 5% have been assessed in terms of grabbing accuracy. Experimental test results are shown in Fig. 5.12. Data from Fig. 5.12 show that the best overlap range moves from 20% and the 40%, with a peak at 25%.

Above this limit, empty or semi-empty packages start falling or solidly shifting with the hand. It results in a reduction of grabbing accuracy with the increment of the overlap.

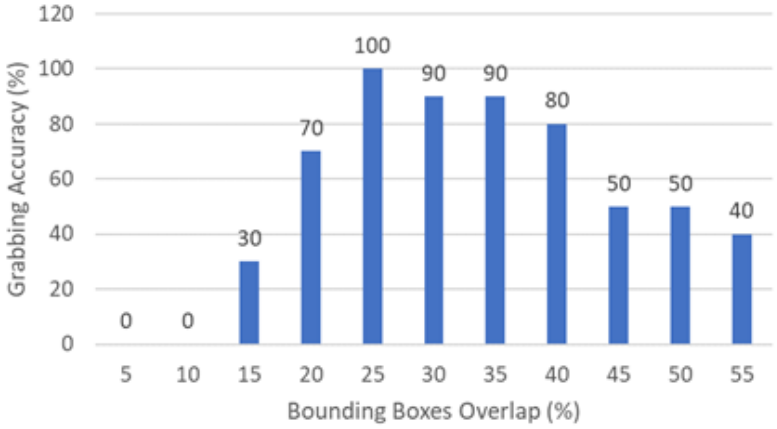

Figure 5.12 Grabbing accuracy versus bounding boxes overlap percentage

5.5.6 Real-life Scenario Application: Success Rate

To characterize the improved robot capabilities in the AMICO framework, 360 complete manipulation procedures have been carried out starting from BCI selection. This section analyzes the number of fails in the grabbing procedure analyzing and dividing them per step affecting the success rate.

Overall, 313 out of 360 (86.94%) grabbing operations have been successfully completed. While 47/360 fails have been recorded during test. Fig. 5.13 shows the occurrence rate of the failure causes, dividing them per manipulation steps.

Experimental results show that in 21.27% (10/47) of cases, the error is related to an erroneous shelf height estimation (see Sec. 4.3.1). The operations have been automatically stopped to avoid collisions between hand frame and the shelf.

The arm segments movement planning (see Sec. 4.3.4) for Routine #2 is the phase that showed the highest failure rate with 12/47 failures (25.53%). Poor management of the arm segments in most cases (9/12) led to a high and unstable grab, while in the remaining cases (3/12) to a collision with the shelf, ending the procedure prematurely.

The Object Approach step (both for Routine #1 – Sec. 4.3.3 and #2 - Sec. 4.3.4) constitutes the third main cause of failure with the 17.02% of cases.

Side selection (see Sec. 4.3.4) is the 4th cause with 7/47 fails (14.9%). The main cause of the error is the lateral adjustment to cover the 25% of bounding boxes overlap. Indeed, during adjustments, the robot moves lightweight envelopes, causing their fall.

The Grab and Check procedure for both routines caused failures in 6 out of 47 (12.76%) cases. The main cause of the failure lies in the unstable recognition of the check and envelopes that shifted from the grab during wrist rotation.

The YOLO-based Object Detection step constitutes the last cause of errors with 4/47 failed operations.

Figure 5.13 Error occurrence (%) per routine step during the system characterization

Conclusions

AAL infrastructures are playing a critical role in a world that is moving towards global aging. New robotics-empowered AAL infrastructures trend has demonstrated the ability of these systems to provide entertainment, reminders, social improvement, physical activity support (where possible), and cognitive testing, improving the elderly quality of life for active aging.

However, most of these systems: (i) are non-inclusive for a good portion of the 65+ population, who experience partial or total disabilities that prevent their practical use; (ii) lack capabilities for assessing the patient's cognitive status; (iii) are not always adaptive to the emotional state of users; (iv) are designed for the implementation of social interaction tasks but not for the provision of services.

In this context, the present thesis introduced new features that allow pre-existing AAL infrastructures with robotic agents to evolve towards a more user wellness-aware architecture. Specifically, to address the aforementioned problems, the proposed architecture has introduced an innovative user-robot-environment interface based on a BCI approach. This interface opens the possibility of formalizing requests regardless of the severity of the user's disability.

In this respect, two BCI s have been implemented on the robot agent: an MRP-based BCI and a P300 one. The first BCI (MRP-based BCI) provides a fluid interface to the user who through short paths based on nested choices can formalize a high amount of requests in a short time. The second (i.e., P300-based BCI) is, instead, a more static interface but able to guarantee greater choice recognition accuracy and simultaneous analysis of the user's cognitive state.

This last feature has been added to complete one of the most widespread services by AAL infrastructures: the administration of cognitive tests. This assessment feature can therefore provide quantitative monitoring of the patient's cognitive status, as well as a diagnostic support index in case of neurocognitive impairments, exploiting the biomarking capabilities of the P300.

To address the third point on the list, a user-tailored emotion recognition system has been presented here. It has been designed to discriminate up to 8 emotions by exploiting a search for feature extraction approaches based on two grid searches: one concerning the complexity of the algorithms and one concerning the accuracy of the inference.

The entire infrastructure was validated through in vivo tests on 13 volunteers, who in several sessions were subjected to the performance analysis of (i) MRP-based BCI; (ii) P300- based BCI; (iii) Cognitive Status Extraction, and (iv) Emotion Recognition systems.

The results show a promising accuracy of 84.07 % for the 2-choices MRP-based BCI. This is a first-of-a-kind BCI employing MRP for selection purposes. The ITR of the BCI achieves 11.17 commands/minute.

P300-based BCI achieved an accuracy on a single trial up to 76 %, a precision of 58.23%, a recall of 70.63%, and an F1-score 63.25 %, overcoming in almost all performance metrics the related state-of-the-art solution. In a choice recognition context, the P300-

based BCI demonstrated to be able in achieving an accuracy of 96.66 % after 21s of stimulation. It also ensures a maximum ITR of 17.21 comm./min.
The use of these neural interfaces, in the execution times, does not appear dissimilar from systems that use voice or mechanical commands for selection on tablets or generic video terminals. This similarity opens the possibility of using neural interfaces even for users not suffering from disabling pathologies. Indeed, through the collection of EEGs preventive monitoring, analysis of attention levels, and so on can be conducted even in potentially healthy patients. The approach could also be extended to secondary users such as children by exploiting the biomarking potential of P300 also in fields such as the assessment of hyperactivity or attention deficit.
Future perspectives concern the expansion of the range of choices related to MRP-based BCI by exploiting movements (or motor imagination) of body segments afferent to distant and distinct areas under the plane of the homunculus. Investigating this possibility by expanding the proposed BCI with just 2 choices would push ITR almost doubled, allowing a high operational fluidity.
As for the P300-based BCI, a characterization of robustness to noisy environments, like real-life cases, is still underway. The results of these surveys, while promising, are not self-conclusive, since a larger pool of users must be analyzed.
Continuing on the biolabeling potential of the P300, the experimental evaluations conducted on the cognitive status extraction system returned, as expected, spatio-temporal characteristics compatible with potentially healthy subjects.
It should be specified that this system operates on the basis of labeling against a P300-based BCI with a moderate F1-score. This leads to an analysis of latency and amplitude of the P300 all the more uncertain the lower the ratio between TP and FP. Future perspectives in this sense concern the design and integration into the overall system of a BCI dedicated only to labeling, which presents a high F1-score (with a reduction of FP).
Experimental tests on the emotion recognition system showed that in an 8-emotion discrimination problem (using the AVD model) it can achieve a multiclass accuracy of ~76 % (average) or considering a model that involves only 4 emotions (i.e., AV model), ~ 80 %.
The emotional state recognition system is connected, in the context of the AMICO project, to an adaptive feedback system for the user. This system, together with cognitive evaluation, creates an overall capacity of "Perception - Comprehension - Action" for the AAL infrastructure, addressing aspects that are currently weakly investigated.
The high accuracy values in the context of 8-emotion discrimination open up the possibility of creating fuzzy thresholds for the three descriptive parameters of the emotional spectrum (i.e., arousal, valence, dominance). Future perspectives will concern the creation of emotion recognition systems that exploit this type of thresholding for the definition of a higher number of emotions. Indeed, discriminating more finely and with appropriate accuracy, the spectrum of the emotional state would allow the creation of a more emotional-aware infrastructure able to evaluate more carefully the operations to be taken to bring emotions back to positive valences.

This could be supported by dedicated recommendation systems with the memory of the emotions felt in the presence of certain feedback.
In the context of this thesis, a low-complexity manipulation system capable of operating offline while preserving sensitive data stored in the robot's memory has also been implemented as a secondary requirement for the implementation of specific service requests.
This system, tested in real-life scenarios, returned a grabbing accuracy of ~87% for different shelf heights, demonstrating the employability of improved social robotics for daily-life assistance and ambulatorial contexts.
At present, this manipulation system is subservient to specific environmental conditions (objects on specific positioning grids, specific distances from shelf edges, etc.). The routines have been optimized to cover human positioning errors in terms of placement and rotation, however, although functional to the specific application, the routine is far to be considered optimal.
One of the main reasons concerns the lack of control of individual fingers on the chosen robot, as well as a feedback sensing system regarding the grip itself. Future perspectives will concern the inclusion of low-cost electronic skin systems and the expansion, by means of applicable external devices, of the number of servomotors for the management of the robot hand.

References

[1] World Health Organization (WHO) report on aging and health. Online (October 1, 2022):

[2] Organisation for Economic Co-operation and Development (OECD) Demography Section: , https://data.oecd.org/pop/old-age-dependency-ratio.htm#indicator-chart

[3] F. Schwiegelshohn et al., "A Holistic Approach for Advancing Robots in Ambient Assisted Living Environments," 2015 IEEE 13th International Conference on Embedded and Ubiquitous Computing, 2015, pp. 140-147, doi: 10.1109/EUC.2015.37.

[4] Ferrucci L, Giallauria F, Guralnik JM. Epidemiology of aging. Radiol Clin North Am. 2008 Jul;46(4):643-52, v. doi: 10.1016/j.rcl.2008.07.005. PMID: 18922285; PMCID: PMC2692491.

[5] De Carolis, B., Ferilli, S., and Macchiarulo, N. (2022, July). Ambient Assisted Living and Social Robots: Towards Learning Relations between User's Daily Routines and Mood. In Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (pp. 123-129).

[6] Blackman, S., Matlo, C., Bobrovitskiy, C., Waldoch, A., Fang, M., Jackson, P., Mihailidis, A., Nygård, L., Astell, A. & Sixsmith, A. (2016). Ambient Assisted Living Technologies for Aging Well: A Scoping Review. Journal of Intelligent Systems, 25(1), 55-69. https://doi.org/10.1515/jisys-2014-0136

[7] Muñoz, Andrés, et al. "Design and evaluation of an ambient assisted living system based on an argumentative multi-agent system." Personal and Ubiquitous Computing 15.4 (2011): 377-387.

[8] Oliveira, Juliana Damasio, et al. "Improving the design of ambient intelligence systems: Guidelines based on a systematic review." International Journal of Human–Computer Interaction 38.1 (2022): 19-27.

[9] Doughty, Kevin, Keith Cameron, and Paul Garner. "Three generations of telecare of the elderly." Journal of Telemedicine and Telecare 2.2 (1996): 71-80.

[10] Røhne, M., Boysen, E. S., & Ausen, D. (2017). Wearable and Mobile Technology for Safe and Active Living. In pHealth (pp. 133-139).

[11] De Venuto, D.; Mezzina, G. High-Specificity Digital Architecture for Real-Time Recognition of Loss of Balance Inducing Fall. Sensors 2020, 20, 769. https://doi.org/10.3390/s20030769

[12] Tanwar, Rohit, et al. "Pathway of trends and technologies in fall detection: a systematic review." Healthcare. Vol. 10. No. 1. MDPI, 2022.

[13] Herzog, David Josef. "Video Camera in the Ambient Assisted Living System. Health Versus Privacy." Blockchain and Other Emerging Technologies for Digital Business Strategies. Springer, Cham, 2022. 55-76.

[14] Cicirelli, Grazia, et al. "Ambient assisted living: A review of technologies, methodologies and future perspectives for healthy aging of population." Sensors 21.10 (2021): 3549.

[15] Ranieri, Caetano Mazzoni, et al. "Activity recognition for ambient assisted living with videos, inertial units and ambient sensors." Sensors 21.3 (2021): 768.

[16] Andò, Bruno, et al. "An Introduction to Patterns for the Internet of Robotic Things in the Ambient Assisted Living Scenario." Robotics 10.2 (2021): 56.

[17] Mahmood, Safdar, et al. "Prospects of robots in assisted living environment." Electronics 10.17 (2021): 2062.

[18] Feil-Seifer, David, and Maja J. Mataric. "Defining socially assistive robotics." 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.. IEEE, 2005.

[19] Bemelmans, Roger, et al. "Socially assistive robots in elderly care: a systematic review into effects and effectiveness." Journal of the American Medical Directors Association 13.2 (2012): 114-120.

[20] ISO 13482:2014 Robots and robotic devices — Safety requirements for personal care robots. Online: https://www.iso.org/standard/53820.html

[21] Jacobs, Theo, and Gurvinder Singh Virk. "ISO 13482-The new safety standard for personal care robots." ISR/Robotik 2014; 41st International Symposium on Robotics. VDE, 2014.

[22] International Federation of Robotics. Service Robots definition. Online: http://www.ifr.org/service-robots/

[23] AAL Joint Programme. Online: http://www.aal-europe.eu/

[24] I. Volosyak, O. Ivlev and A. Graser, "Rehabilitation robot FRIEND II - the general concept and current implementation," 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005., 2005, pp. 540-544, doi: 10.1109/ICORR.2005.1501160

[25] Jardón, Alberto, et al. "Usability assessment of ASIBOT: a portable robot to aid patients with spinal cord injury." Disability and Rehabilitation: Assistive Technology 6.4 (2011): 320-330.
[26] ASIBOT Lab page. Online: http://roboticslab.uc3m.es/roboticslab/content/about-us
[27] Zeng, Zhanjing, Po-Ju Chen, and Alan A. Lew. "From high-touch to high-tech: COVID-19 drives robotics adoption." Tourism geographies 22.3 (2020): 724-734.
[28] Wilson, Rachel, Imogen Keane, and Ray Jones. "Affective Responses of Older Adults to the Anthropomorphic GenieConnect Companion Robot During Lockdown of the COVID19 Pandemic." 2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 2022.
[29] Pandey, Amit Kumar, and Rodolphe Gelin. "A mass-produced sociable humanoid robot: Pepper: The first machine of its kind." IEEE Robotics & Automation Magazine 25.3 (2018): 40-48.
[30] Dag, Munir, Eric Svanelöv, and Christine Gustafsson. "Experiences of using Bestic, an eating aid for people with intellectual disabilities." *Journal of Intellectual Disabilities* 21.1 (2017): 87-98.
[31] Zlatintsi, Athanasia, et al. "I-Support: A robotic platform of an assistive bathing robot for the elderly population." Robotics and Autonomous Systems 126 (2020): 103451.
[32] A.L.O. Botlr. Online: https://techcrunch.com/2014/08/13/starwood-introduces-robotic-butlers-at-aloft-hotel-in-palo-alto/
[33] Chen, Shu-Chuan, Cindy Jones, and Wendy Moyle. "The Impact of Engagement with the PARO Therapeutic Robot on the Psychological Benefits of Older Adults with Dementia." Clinical Gerontologist (2022): 1-13.
[34] Wang, Xinxia, Jun Shen, and Qiu Chen. "How PARO can help older people in elderly care facilities: A systematic review of RCT." International Journal of Nursing Knowledge 33.1 (2022): 29-39.
[35] RP-Vita. Online: https://robots.ieee.org/robots/rpvita/
[36] RADIO project. Online: https://cordis.europa.eu/project/id/643892
[37] ENRICHME project. Online: https://cordis.europa.eu/project/id/643691
[38] GrowMeUp project. Online: https://cordis.europa.eu/project/id/643647
[39] MARIO project. Online: https://cordis.europa.eu/project/id/643808
[40] ACANTO project. Online: https://cordis.europa.eu/project/id/643644
[41] RAMCIP project. Online: http://cordis.europa.eu/project/rcn/194064
[42] I-SUPPORT project. Online: https://cordis.europa.eu/project/id/643666
[43] eNHANCE project. Online: https://cordis.europa.eu/project/id/644000
[44] M. Merten, A. Bley, C. Schroeter and H. Gross, "A mobile robot platform for socially assistive home-care applications," ROBOTIK 2012; 7th German Conference on Robotics, 2012, pp. 1-6.
[45] Coşar, S., Fernandez-Carmona, M., Agrigoroaie, R. et al. ENRICHME: Perception and Interaction of an Assistive Robot for the Elderly at Home. Int J of Soc Robotics 12, 779–805 (2020). https://doi.org/10.1007/s12369-019-00614-y
[46] Iosifidis, Alexandros, et al. "The MOBISERV-AIIA eating and drinking multi-view database for vision-based assisted living." Journal of Information Hiding and Multimedia Signal Processing 6.2 (2015): 254-273.
[47] Fasola, Juan, and Maja J. Matarić. "A socially assistive robot exercise coach for the elderly." Journal of Human-Robot Interaction 2.2 (2013): 3-32.
[48] Khosla, Rajiv, and Mei-Tai Chu. "Embodying care in Matilda: an affective communication robot for emotional wellbeing of older people in Australian residential care facilities." ACM Transactions on Management Information Systems (TMIS) 4.4 (2013): 1-33.
[49] Bishop, L., van Maris, A., Dogramadzi, S. & Zook, N. (2019). Social robots: The influence of human and robot characteristics on acceptance. Paladyn, Journal of Behavioral Robotics, 10(1), 346-358. https://doi.org/10.1515/pjbr-2019-0028
[50] V. Di Palma et al., "“Medical Assistance in Contextual awareness” (AMICO): a project for a better cardiopathic patients quality of care," 2019 IEEE 8th International Workshop on Advances in Sensors and Interfaces (IWASI), 2019, pp. 230-233, doi: 10.1109/IWASI.2019.8791308.
[51] AMICO project. Online: http://www.ponricerca.gov.it/comunicazione/example-projects/industrial-research-and-experimental-development-projects-in-the-12-specialization-areas/amico-medical-assistance-in-contextual-awareness/
[52] Massimino, G., et al. "Model order reduction for the analysis of large arrays of piezoelectric micromachined ultrasonic transducers in water." Applied Acoustics 182 (2021): 108231.
[53] ALMemory. Online: http://doc.aldebaran.com/2-4/naoqi/core/almemory.html

[54] Braun, M., Schubert, J., Pfleging, B., and Alt, F. Improving driver emotions with affective strategies. Multimodal Technologies and Interaction 2019, 3(1), 21.
[55] G.Nautilus Research - g.tec medical engineering GmbH. Online: https://www.gtec.at/product/gnautilus-research/
[56] Unicorn Hybrid Black - g.tec neurotechnology GmbH. Online: https://www.unicorn-bi.com/it/
[57] ENOBIO- Neuroelectrics . Online: https://www.neuroelectrics.com/solutions/enobio
[58] Homan, Richard W., John Herman, and Phillip Purdy. "Cerebral location of international 10–20 system electrode placement." Electroencephalography and clinical neurophysiology 66.4 (1987): 376-382.
[59] Kilicarslan, Atilla, Robert G. Grossman, and Jose Luis Contreras-Vidal. "A robust adaptive denoising framework for real-time artifact removal in scalp EEG measurements." Journal of neural engineering 13.2 (2016): 026013.
[60] NAO Marks. Online: http://doc.aldebaran.com/2-1/naoqi/vision/allandmarkdetection.html
[61] G. Mezzina, D. Ciccarese and D. De Venuto, "Preventing Food Waste by the Shelf-Life Dynamical Control in Smart Homes," in IEEE Sensors Letters, vol. 6, no. 8, pp. 1-4, Aug. 2022, Art no. 6002604, doi: 10.1109/LSENS.2022.3191410.
[62] FSR402. Online: https://www.interlinkelectronics.com/fsr-402
[63] Kæseler, Rasmus L., et al. "Feature and Classification Analysis for Detection and Classification of Tongue Movements From Single-Trial Pre-Movement EEG." IEEE Transactions on Neural Systems and Rehabilitation Engineering 30 (2022): 678-687.
[64] Kropotov, J., Ponomarev, V., Tereshchenko, E. P., Müller, A., and Jäncke, L. (2016). Effect of aging on ERP components of cognitive control. Frontiers in aging neuroscience, 8, 69.
[65] Mendoza-Montoya, O., Antelis, J. M., & Delijorge, J. (2022). P300-based brain–computer interface for communication and control. In Biosignal Processing and Classification Using Computational Learning and Intelligence (pp. 271-292). Academic Press.
[66] Shakeel, A., Navid, M. S., Anwar, M. N., Mazhar, S., Jochumsen, M., & Niazi, I. K. (2015). A review of techniques for detection of movement intention using movement-related cortical potentials. Computational and mathematical methods in medicine, 2015.
[67] Shibasaki H, Hallett M. What is the Bereitschaftspotential? Clin Neurophysiol. 2006 Nov;117(11):2341-56. doi: 10.1016/j.clinph.2006.04.025. Epub 2006 Jul 28. PMID: 16876476.
[68] Grodd W, Hülsmann E, Lotze M, Wildgruber D, Erb M. Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. Hum Brain Mapp. 2001 Jun;13(2):55-73.
[69] Sur S, Sinha VK. Event-related potential: An overview. Ind Psychiatry J. 2009 Jan;18(1):70-3. doi: 10.4103/0972-6748.57865. PMID: 21234168; PMCID: PMC3016705.
[70] Newsome, Rachel N., et al. "Neural correlates of cognitive decline in older adults at-risk for developing MCI: evidence from the CDA and P300." Cognitive neuroscience 4.3-4 (2013): 152-162.
[71] C.C. Duncan, et al. "Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300 and N400", Clinical Neurophys., v.120, n.11, pp. 1883-1908, 2009.
[72] D. De Venuto, V. F. Annese, G. Mezzina, M. Ruta and E. Di Sciascio, "Brain-computer interface using P300: a gaming approach for neurocognitive impairment diagnosis," 2016 IEEE International High Level Design Validation and Test Workshop (HLDVT), 2016, pp. 93-99, doi: 10.1109/HLDVT.2016.7748261.
[73] Janyalikit, Thapanan, and Chotirat Ann Ratanamahatana. "Time Series Shapelet-Based Movement Intention Detection Toward Asynchronous BCI for Stroke Rehabilitation." IEEE Access 10 (2022): 41693-41707.
[74] G. Mezzina and D. De Venuto, "Cortical Activity Digitization by Symbolization in Brain-Computer Interface Context," 2020 15th Design & Technology of Integrated Systems in Nanoscale Era (DTIS), 2020, pp. 1-6, doi: 10.1109/DTIS48698.2020.9081325.
[75] G. Mezzina and D. De Venuto, "Local Binary Patterning Approach for Movement Related Potentials based Brain Computer Interface," 2019 IEEE 8th International Workshop on Advances in Sensors and Interfaces (IWASI), 2019, pp. 239-244, doi: 10.1109/IWASI.2019.8791266.
[76] Liu, Tengjun, et al. "Reduce brain computer interface inefficiency by combining sensory motor rhythm and movement-related cortical potential features." Journal of Neural Engineering 17.3 (2020): 035003.
[77] Mascolini, Alessio, Imran Khan Niazi, and Luca Mesin. "Non-linear optimized spatial filter for single-trial identification of movement related cortical potential." Biocybernetics and Biomedical Engineering 42.1 (2022): 426-436.
[78] Robinson, Neethu, et al. "Emerging trends in BCI-robotics for motor control and rehabilitation." Current Opinion in Biomedical Engineering 20 (2021): 100354.

[79] Kæseler, Rasmus L., et al. "Feature and Classification Analysis for Detection and Classification of Tongue Movements From Single-Trial Pre-Movement EEG." IEEE Transactions on Neural Systems and Rehabilitation Engineering 30 (2022): 678-687.

[80] Müller-Putz, Gernot. "Toward Non-invasive BCI-Based Movement Decoding." Neuroprosthetics and Brain-Computer Interfaces in Spinal Cord Injury. Springer, Cham, 2021. 233-249.

[81] Aliakbaryhosseinabadi, Susan, et al. "Participant-specific classifier tuning increases the performance of hand movement detection from EEG in patients with amyotrophic lateral sclerosis." Journal of Neural Engineering 18.5 (2021): 056023.

[82] Castro-Aparicio, Juan C., Ruben I. Carino-Escobar, and Jessica Cantillo-Negrete. "Movement-Related Electroencephalography in Stroke Patients Across a Brain-Computer Interface-Based Intervention." Latin American Workshop on Computational Neuroscience. Springer, Cham, 2022.

[83] Korkmaz, Onur Erdem, et al. "An efficient 3D column-only P300 speller paradigm utilizing few numbers of electrodes and flashings for practical BCI implementation." PloS one 17.4 (2022): e0265904.

[84] Tang, Zhichuan, et al. "A BCI painting system using a hybrid control approach based on SSVEP and P300." Computers in Biology and Medicine (2022): 106118.

[85] Aygün, Abdullah Bilal, and Ahmet Reşit Kavsaoğlu. "An innovative P300 speller brain–computer interface design: Easy screen." Biomedical Signal Processing and Control 75 (2022): 103593.

[86] G. Mezzina and D. De Venuto, "Low-Complexity Unidimensional CNN based Brain Speller for Embedded Platforms," 2021 IEEE Nordic Circuits and Systems Conference (NorCAS), 2021, pp. 1-6, doi: 10.1109/NorCAS53631.2021.9599874.

[87] G. Mezzina and D. De Venuto, "Four-Wheel Vehicle Driving by using a Spatio-Temporal Characterization of the P300 Brain Potential," 2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), 2020, pp. 1-6, doi: 10.23919/AEITAUTOMOTIVE50086.2020.9307405.

[88] Fouad, Islam A. "A robust and reliable online P300-based BCI system using Emotiv EPOC+ headset." Journal of Medical Engineering & Technology 45.2 (2021): 94-114.

[89] Najjar, A. Benabid, et al. "Toward the Optimization of the Region-Based P300 Speller." CMC-COMPUTERS MATERIALS & CONTINUA 67.1 (2021): 1169-1189.

[90] Mussabayeva, Ayana, Prashant Kumar Jamwal, and Muhammad Tahir Akhtar. "Ensemble Voting-Based Multichannel EEG Classification in a Subject-Independent P300 Speller." Applied Sciences 11.23 (2021): 11252.

[91] Uma, M., et al. "Analysis of Effect of RSVP Speller BCI Paradigm Along with CNN to Analysis P300 Signals." International Conference on Human-Computer Interaction. Springer, Cham, 2021.

[92] De Venuto, D.; Mezzina, G. A Single-Trial P300 Detector Based on Symbolized EEG and Autoencoded-(1D)CNN to Improve ITR Performance in BCIs. Sensors 2021, 21, 3961. https://doi.org/10.3390/s21123961

[93] M. de Tommaso, E. Vecchio, K. Ricci, A. Montemurno, D. De Venuto and V. F. Annese, "Combined EEG/EMG evaluation during a novel dual task paradigm for gait analysis," 2015 6th International Workshop on Advances in Sensors and Interfaces (IWASI), Gallipoli, 2015, pp. 181-186, doi: 10.1109/IWASI.2015.7184949.

[94] Farwell L A and Donchin E, 'Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potential', Electroencephalogr and Clin Neurophy, 70:510-523, 1988.

[95] Blum, Sarah, et al. "A Riemannian modification of artifact subspace reconstruction for EEG artifact handling." Frontiers in human neuroscience 13 (2019): 141.

[96] Xu, Ren, et al. "Movement-related cortical potentials in paraplegic patients: abnormal patterns and considerations for BCI-rehabilitation." Frontiers in neuroengineering 7 (2014): 35.

[97] Jiang, Ning, et al. "A brain–computer interface for single-trial detection of gait initiation from movement related cortical potentials." Clinical Neurophysiology 126.1 (2015): 154-159.

[98] D. De Venuto, V. F. Annese, M. Ruta, E. Di Sciascio and A. L. Sangiovanni Vincentelli, "Designing a Cyber–Physical System for Fall Prevention by Cortico–Muscular Coupling Detection," in IEEE Design & Test, vol. 33, no. 3, pp. 66-76, June 2016, doi: 10.1109/MDAT.2015.2480707.

[99] D. De Venuto, V. F. Annese, G. Mezzina and G. Defazio, "FPGA-Based Embedded Cyber-Physical Platform to Assess Gait and Postural Stability in Parkinson's Disease," in IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 8, no. 7, pp. 1167-1179, July 2018, doi: 10.1109/TCPMT.2018.2810103.

[100] K. Schindler et al., "On seeing the trees and the forest: Single signal and multisignal analysis of peri ictal intracranial EEG", Epilepsia, 2012.
[101] G. Mezzina and D. De Venuto, "Local Binary Patterning Approach for Movement Related Potentials based Brain Computer Interface," 2019 IEEE 8th International Workshop on Advances in Sensors and Interfaces (IWASI), 2019, pp. 239-244, doi: 10.1109/IWASI.2019.8791266.
[102] Mezzina G. and De Venuto D., "Cortical Activity Digitization by Symbolization in Brain-Computer Interface Context," 2020 15th Design & Technology of Integrated Systems in Nanoscale Era (DTIS), Marrakech, Morocco, 2020, pp. 1-6, doi: 10.1109/DTIS48698.2020.9081325.
[103] G. Mezzina and D. De Venuto, "Smart Sensors HW/SW Interface based on Brain-actuated Personal Care Robot for Ambient Assisted Living," 2020 IEEE SENSORS, 2020, pp. 1-4, doi: 10.1109/SENSORS47125.2020.9278808.
[104] G. Mezzina and D. D. Venuto, "Semi-Autonomous Personal Care Robots Interface driven by EEG Signals Digitization," 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE), 2020, pp. 264-269, doi: 10.23919/DATE48585.2020.9116499.
[105] D. De Venuto and G. Mezzina, "Multisensing System for Parkinson's Disease Stage Assessment Based on FPGA-Embedded Serial SVM Classifier," in IEEE Design & Test, vol. 38, no. 4, pp. 44-51, Aug. 2021, doi: 10.1109/MDAT.2019.2951117.
[106] De Venuto D, Mezzina G. A Single-Trial P300 Detector Based on Symbolized EEG and Autoencoded-(1D)CNN to Improve ITR Performance in BCIs. Sensors. 2021; 21(12):3961.
[107] Liu, M.; Wu, W.; Gu, Z.; Yu, Z.; Qi, F.; Li, Y. Deep learning based on batch normalization for P300 signal detection. Neurocomputing 2018, 275, 288–297.
[108] Cecotti, H.; Graser, A. Convolutional neural networks for p300 detection with application to brain–computer interfaces. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 33, 433–445.
[109] Rakotomamonjy, A.; Guigue, V. BCI competition III: Dataset II-ensemble of SVMs for BCI P300 speller. IEEE Trans. Biomed. Eng. 2008, 55, 1147–1154.
[110] Carabez, E.; Miho, S.; Nambu, I.; Yasuhiro, W. Convolutional Neural Networks with 3D Input for P300 Identification in Auditory Brain-Computer Interfaces. Comput. Intell. Neurosci. 2017, 2017, 8163949.
[111] Maddula, R.; Stivers, J.; Mousavi, M.; Ravindran, S.; de Sa, V. Deep Recurrent Convolutional Neural Networks for Classifying P300 BCI signals. In Proceedings of the 7th Graz Brain-Computer Interface Conference, GBCIC 2017, Gratz, Austria, 18–22 September 2017.
[112] Li, F.; Li, X.; Wang, F.; Zhang, D.; Xia, Y.; He, F. A Novel P300 Classification Algorithm Based on a Principal Component Analysis-Convolutional Neural Network. Appl. Sci. 2020, 10, 1546.
[113] Chollet, F. & others, 2015. Keras. Available at: https://github.com/fchollet/keras
[114] O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., & others. (2019). Keras Tuner. Available at: https://keras.io/keras_tuner/
[115] Li, Lisha, and Kevin Jamieson. "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization." Journal of Machine Learning Research 18 (2018): 1-52
[116] Vanhoucke, V., Senior, A., & Mao, M. Z. (2011). Improving the speed of neural networks on CPUs. Google Research.
[117] Li, Xinru, Zelin Hu, and Xiaoping Huang. "Combine ReLU with tanh." 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). Vol. 1. IEEE, 2020.
[118] J. Polich, C. L. Ehlers, S. Otis, A. J. Mandell and F. E. Bloom, "P300 latency reflects the degree of cognitive decline in dementing illness", EEG Clin. Neurophysiol., vol. 63, no. 2, pp. 138-144, 1986.
[119] S. Makeig, "Blind separation of auditory event-related brain responses into independent components", Proc. Nat. Acad. Sci. USA, vol. 94, no. 20, pp. 10979-10984, 1997.
[120] D. De Venuto, V. F. Annese and G. Mezzina, "Remote Neuro-Cognitive Impairment Sensing Based on P300 Spatio-Temporal Monitoring," in IEEE Sensors Journal, vol. 16, no. 23, pp. 8348-8356, Dec.1, 2016, doi: 10.1109/JSEN.2016.2606553.
[121] T. Frodl et al., "Value of event-related P300 subcomponents in the clinical diagnosis of mild cognitive impairment and Alzheimer's disease", Psychophysiology, vol. 39, no. 2, pp. 175-181, 2002.
[122] Y. Takeda, M.-A. Sato, K. Yamanaka, D. Nozaki and Y. Yamamoto, "A generalized method to estimate waveforms common across trials from EEGs", NeuroImage, vol. 51, no. 2, pp. 629-641, 2010.
[123] Venuto, D. De, et al. "A mobile health system for neurocognitive impairment evaluation based on P300 detection." ACM Transactions on Cyber-Physical Systems 2.4 (2018): 1-21.

[124] De Venuto, Daniela, Valerio F. Annese, and Giovanni Mezzina. "An embedded system remotely driving mechanical devices by P300 brain activity." Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017. IEEE, 2017.

[125] Sinha N. (2021) Affective Computing and Emotion-Sensing Technology for Emotion Recognition in Mood Disorders. In: Marques G., Kumar Bhoi A., de la Torre Díez I., Garcia-Zapirain B. (eds) Enhanced Telemedicine and e-Health. Studies in Fuzziness and Soft Computing, vol 410. Springer, Cham. https://doi.org/10.1007/978-3-030-70111-6_16

[126] G. Mezzina and D. De Venuto, "Automatic Accuracy versus Complexity Characterization for Embedded Emotion-Sensing Platforms in Healthcare Applications," 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), 2022, pp. 676-683, doi: 10.1109/COMPSAC54236.2022.00116.

[127] Christos D. Katsis, George Ganiatsas and Dimitrios I. Fotiadis, "An integrated telemedicine platform for the assessment of affective physiological states", Diagnostic pathology, vol. 1.1, pp. 1-9, 2006.

[128] K. Barjinder, D. Singh and P. R. Pratim, "EEG Based Emotion Classification Mechanism in BCI", Procedia Computer Science, vol. 132, pp. 752-758, 2018

[129] T. D. Kusumaningrum, A. Faqih and B. Kusumoputro, "Emotion Recognition Based on DEAP Database using EEG Time-Frequency Features and Machine Learning Methods", IOPScience 2020 J. Phys.: Conf. Ser., vol. 1501, pp. 012020.

[130] G. Li, Z. Zhang and G. Wang, "Emotion recognition based on low-cost in-ear EEG", IEEE Biomedical Circuits and Systems Conference (BioCAS), pp. 1-4, 2017.

[131] M. Mohammadpour, S. M. R. Hashemi and N. Houshmand, "Classification of EEG-based emotion for BCI applications", Artificial Intelligence and Robotics (IRANOPEN) Qazvin, pp. 127-131, 2017.

[132] S. Gannouni et al., "Adaptive Emotion Detection Using the Valence-Arousal-Dominance Model and EEG Brain Rhythmic Activity Changes in Relevant Brain Lobes", IEEE Access, vol. 8, pp. 67444-67455, 2020.

[133] V. Gupta, M. D. Chopda and R. B. Pachori, "Cross-Subject Emotion Recognition Using Flexible Analytic Wavelet Transform From EEG Signals", IEEE Sensors Journal, vol. 19.6, pp. 2266-2274, 2019.

[134] M. Egger, M. Ley and S. Hanke, "Emotion recognition from physiological signal analysis: A review", Electronic Notes in Theoretical Computer Science, vol. 343, pp. 35-55, 2019

[135] Koelstra, S., et al. DEAP: A Database for Emotion Analysis; Using Physiological Signals. IEEE Transactions on Affective Computing 2012, 3.1: 18-31, Jan.-March 2012, doi: 10.1109/T-AFFC.2011.15.

[136] Xie, T., Cao, M., and Pan, Z. Applying Self-Assessment Manikin (SAM) to Evaluate the Affective Arousal Effects of VR Games. Proceedings of the 3rd International Conference on Image and Graphics Processing. 2020.

[137] Augello, Agnese, et al. "Extending affective capabilities for medical assistive robots." Cognitive Systems Research 73 (2022): 21-25.

[138] Stroud, R. O., Ertas, A., and Mengel, S. Application of Cyclomatic Complexity in Enterprise Architecture Frameworks. IEEE Systems Journal 2019, 13.3: 2166-2176.

[139] Zhu, J., Zheng, W., Peng, Y. Duan, R., and Lu, B. EEG-based Emotion Recognition Using Discriminative Graph Regularized Extreme Learning Machine. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC 2014, pp. 974-977, 2014.

[140] Hatamikia, S., Maghooli, K., and Nasrabadi, A.M. The emotion recognition system based on autoregressive model and sequential forward feature selection of electroencephalogram signals. J Med Signals Sens, 4.3: 194-201, 2014.

[141] Winkler, I., Jäger Vojkan Mihajlović M. and Tsoneva T. Frontal EEG Asymmetry based classification of emotional valence using common spatial patterns. World Academy of Science Engineering and Technology 2010, 70:9, 373-378.

[142] Hjorth, B., Elema-Schönander, A.B. EEG analysis based on time domain properties. Electroencephalogr. Clin. Neurophysiol 1970, 29:306–310

[143] Patil, A., Deshmukh, C., and Panat, A. R. Feature extraction of EEG for emotion recognition using Hjorth features and higher order crossings. Conference on Advances in Signal Processing (CASP), Pune, India, 2016.

[144] Yang, W., Wang, K., Zuo, W. Neighborhood Component Feature Selection for High-Dimensional Data. Journal of Computers 2012, 7.1:161-168.

[145] NAOqi Framework. Online: https://developer.softbankrobotics.com/nao6/naoqi-developer-guide/former-naoqi-framework

[146] Self-Collision. Online: http://doc.aldebaran.com/2-5/naoqi/motion/reflexes-collision-avoidance.html#reflexes-collision-avoidance
[147] External Collision. Online: http://doc.aldebaran.com/2-5/naoqi/motion/reflexes-external-collision.html#reflexes-external-collision
[148] Pepper body frame. Online: http://doc.aldebaran.com/2-5/family/pepper_technical/masses_pep.html
[149] Harada, Koichi, and Eihachiro Nakamae. "Application of the Bézier curve to data interpolation." Computer-aided design 14.1 (1982): 55-59.
[150] ALVideoDevice method. Online: http://doc.aldebaran.com/2-4/naoqi/vision/alvideodevice-api.html#alvideodevice-api
[151] Khan, Zeashan Hameed et al. "Robotics Utilization for Healthcare Digitization in Global COVID-19 Management." International journal of environmental research and public health vol. 17,11 3819. 28 May. 2020, doi:10.3390/ijerph17113819
[152] Mezzina, G.; De Venuto, D. RGB and 3D-Segmentation Data Combination for the Autonomous Object Manipulation in Personal Care Robotics. In: 2021 16th International Conference on Design & Technology of Integrated Systems in Nanoscale Era (DTIS). IEEE, 2021. p. 1-6.
[153] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).
[154] Mao, Qi-Chao, et al. "Mini-YOLOv3: real-time object detector for embedded applications." IEEE Access 7 (2019): 133529-133538.
[155] Shrivakshan, G. T.; Chandrasekar, C. A comparison of various edge detection techniques used in image processing. International Journal of Computer Science Issues (IJCSI), 2012, 9.5: 269.
[156] YOLOv3 weights. https://pjreddie.com/media/files/yolov3.weights
[157] Jeong, Ji-Hoon, et al. "2020 International brain–computer interface competition: A review." Frontiers in human neuroscience 16 (2022): 898300.
[158] Joshi, R.; Goel, P.; Sur, M.; Murthy, H.A. Single Trial P300 Classification Using Convolutional LSTM and Deep Learning Ensembles Method. In Intelligent Human Computer Interaction. IHCI 2018. Lecture Notes in Computer Science; Tiwary, U., Ed.; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11278.