ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

MECCANICA E SCIENZE AVANZATE DELL'INGEGNERIA

Ciclo 37

Settore Concorsuale: 09/A2 - MECCANICA APPLICATA ALLE MACCHINE

Settore Scientifico Disciplinare: ING-IND/13 - MECCANICA APPLICATA ALLE MACCHINE

MODELING PERCEPTION OF HUMAN-ROBOT INTERACTION: TOWARD NATURAL AND SOCIAL HRI EXPERIENCES

Presentata da: Matteo Lavit Nicora

Coordinatore Dottorato

Lorenzo Donati

Supervisore

Rocco Vertechy

Esame finale anno 2025

"We spend a lot of time designing the bridge, but not enough time thinking about the people who are crossing it."

Dr. Prabhjot Singh

Abstract

Modeling Perception of Human-Robot Interaction: Toward Natural and Social HRI Experiences

Matteo LAVIT NICORA

The increasing integration of robotic systems into various domains such as industrial, medical, and social environments raises the need to put more and more emphasis on user-centered design to ensure that these technologies enhance human well-being and inclusivity. This thesis investigates the development of a generalized human-driven control architecture to facilitate natural and socially-aware human-robot interaction. The proposed framework incorporates insights from biomechanics, physiology, psychology, and social science to provide a comprehensive model of the user's experience. By leveraging this model, robotic systems can dynamically adapt their behavior, promoting interactions that are both personalized and empathetic.

The research focuses on two primary domains where the experience of interaction plays a vital role: collaborative industrial robotics and robotic neurorehabilitation. In industrial settings, cobots should aim to improve workplace ergonomics, efficiency, and operator well-being by complementing human flexibility and decision-making with robotic precision and consistency. In neurorehabilitation, instead, robotic systems should attempt to augment therapeutic practices, enhancing patient engagement and recovery outcomes through adaptive and socially-responsive behaviors. On these bases, the two use cases of interest also serve as testbeds for the implementation and validation of the proposed framework.

The methodology followed in this project encompasses three phases: a comprehensive review of the existing literature to identify the key factors influencing user experience in HRI; the design and implementation of experimental setups deployed for the aforementioned application domains; and their validation through empirical studies with diverse participant groups, including neurodivergent individuals. The approach incorporates ethical considerations and prioritizes non-invasive data collection methods, ensuring both usability in real-life scenarios and compliance with privacy standards.

Experimental results highlight the framework's ability to effectively integrate heterogeneous data, such as biomechanical, physiological, social and psychological signals, into actionable insights for real-time robotic adaptation. By facilitating smoother interactions and addressing the varied needs of users, the framework supports a user-centered approach to robotics. Building upon the lessons learned during the research activities, this thesis also outlines practical guidelines for replicating and extending the proposed architecture across different scenarios, emphasizing its potential to enhance both usability and social acceptability in HRI.

Overall, the findings underline the importance of interdisciplinary approaches in designing robotic systems that prioritize human experience. Future work will focus on refining the adaptive capabilities of the architecture and extending its application to broader contexts, contributing to the necessary constant improvement of human-robot collaboration.

Riassunto

Modeling Perception of Human-Robot Interaction: Toward Natural and Social HRI Experiences

Matteo LAVIT NICORA

La crescente integrazione di sistemi robotici in svariati ambiti, come quello industriale, medico e sociale, richiede una sempre piú pressante attenzione al design che deve essere centrato sull'utente per garantire che queste tecnologie migliorino il benessere degli utilizzatori e promuovano l'inclusività. Questa tesi presenta lo sviluppo di un'architettura di controllo generalizzata in cui l'esperienza della persona é posta al centro in modo da favorire interazioni uomo-robot sempre piú naturali e sociali. Il framework proposto integra soluzioni derivate dalla biomeccanica, fisiologia, psicologia e scienze sociali per offrire un modello completo dell'esperienza dell'utente. Sfruttando tale modello, i sistemi robotici possono adattare dinamicamente il loro comportamento, promuovendo interazioni personalizzate ed empatiche.

Le attivitá di ricerca presentate si concentrano su due ambiti principali in cui l'esperienza di interazione con i dispositivi robotici rappresenta un punto focale: la robotica collaborativa industriale e quella neuroriabilitativa. Nei contesti industriali, l'utilizzo dei cobot rappresenta un'importante opportunitá per promuovere ergonomia, efficienza e benessere per gli operatori, combinando la flessibilità e la capacità decisionale umana con la precisione e la ripetibilitá tipiche della robotica. Nell'ambito della neuroriabilitazione, invece, i sistemi robotici sono progettati al fine di potenziare le pratiche terapeutiche, migliorando il coinvolgimento dei pazienti e i risultati della riabilitazione attraverso comportamenti adattivi e socialmente responsivi. I due ambiti applicativi presentati servono anche come banchi di prova per l'implementazione e la validazione dell'architettura di controllo proposta.

La metodologia seguita nel progetto si articola in tre fasi: una revisione della letteratura esistente per identificare i fattori chiave che influenzano l'esperienza dell'utente nell'interazione con dispositivi robotici; la progettazione e l'implementazione di setup sperimentali basati sull'architettura proposta e declinati per i campi applicativi di interesse; e la loro validazione mediante studi empirici condotti con gruppi di partecipanti eterogenei, inclusi

soggetti neurodivergenti. L'approccio sperimentale scelto é basato su considerazioni etiche e privilegia metodi di raccolta dati non invasivi, garantendo sia l'usabilità in applicazioni reali sia il rispetto degli standard di privacy.

I risultati sperimentali evidenziano la capacità dell'architettura proposta di integrare efficacemente dati eterogenei, come segnali biomeccanici, fisiologici e sociali, traducendoli in informazioni utili per adattare in tempo reale il comportmento di sistemi robotici. Promuovendo interazioni più naturali e rispondendo alle diverse esigenze degli utenti, le soluzioni introdotte in questo progetto gettano le basi per un approccio al controllo robot centrato sull'utente eterogeneo e multidisciplinare. Grazie all'esperienza raccolta durante le fasi sperimentali, viene inoltre identificata e presentata una serie di linee guida pratiche per replicare ed estendere l'architettura proposta in diversi scenari, sottolineando il suo potenziale nel miglioramento sia dell'usabilità sia dell'accettabilità sociale di tecnologie in interazione con l'uomo.

In generale, i risultati ottenuti sottolineano l'importanza di approcci interdisciplinari nella progettazione di sistemi robotici che pongano al centro l'esperienza umana. Studi futuri si concentreranno sul perfezionamento delle capacità adattive dell'architettura e sull'ampliamento delle sue applicazioni a contesti più ampi, contribuendo al costante miglioramento delle collaborazioni tra uomo e robot.

Acknowledgements

No great things are ever achieved alone. The contents of this thesis are based on my interests and passions, but none of it would have ever been possible without the support of many people.

First of all, I want to thank my wife for always being on my side and encouraging me every step of the way. Her enthusiasm in striving for a positive impact has been a fundamental source of inspiration. Of course, I would have never got to this point without my parents, whose love and care have accompanied me every day. The same goes for my sister and her beautiful family, a constant example to look up to. Special thanks also to my grandparents, who have always been my biggest fans and whose memory always give me new energy. Family is not complete without friends: each one of them is part of these results in their own unique way.

Support at home provides motivation, but support at work gives you the tools to achieve your goals. Here, a long list of acknowledgments begins since the topics addressed in this thesis are varied and multidisciplinary. I am grateful for this opportunity as it allowed me to peek into fascinating fields of research and to meet a lot of competent and supporting people.

Starting from the team of the University of Bologna, I want to thank Prof. Rocco Vertechy, for believing in the project and tutoring me during the doctoral program. As my activities have been primarily based in the laboratories of STIIMA-CNR in Lecco, I also want to thank Prof. Lorenzo Donati and Prof. Marco Carricato, coordinators of the PhD program, for the availability and support they demonstrated whenever needed.

Experimental campaigns are always expensive in terms of time and resources. The only way to complete them is through the collaboration with a big and enthusiastic team. Therefore, I want to thank Matteo Malosio for his constant mentoring and guidance through the presented research activities. Moreover, a wide expertise have been involved in the project: the robotic team (Alessio Prini, Giovanni Tauro, Matteo Meregalli Falerni and Atul Chaudhary), the biomechanical experts (Alessandro Scano, Cristina Brambilla), the VR group (Marco Sacco, Sara Arlati, Vera Colombo), the UX expert (Marta Mondellini), the vision team (Tiziana D'Orazio, Roberto Marani and Laura Romeo), the physiology group (Giovanna Rizzo and Alfonso Mastropietro) and the supporting team (Tito Dinon, Chiara Tagliaferri and Rossella Scaioli) of STIIMA-CNR all contributed to the presented activities and deserve an acknowledgement for their support.

International partnerships have also been pursued, leading to fruitful and valuable collaborations that go beyond the sole goals of this project. Starting from the University of Augsburg, I want to express my gratitude to Prof. Elisabeth André for welcoming me during my research period abroad and for her constant proactivity in pushing towards better results. Special thanks also to Pooja Prajod for all the work that we have successfully carried out together and for sharing joys and sorrows of this academic achievement. I also want to acknowledge Stina Klein and Teresa Grunwald for all the organizational help they offered me during my stay in Augsburg. Fundamental expertise was also provided by DFKI, the German Research Center for Artificial Intelligence. In particular, I want to thank Prof. Patrick Gebhard, Fabrizio Nunnari, Dimitra Tsovaltzi, Sebastian Beyrodt and Rhythm Arora for believing in the project and pushing the limits of human behavior modeling.

The involvement of ASD participants in the experimental activities and the exploration of rehabilitation scenarios would have not been possible without a clinical partner. Invaluable support in this terms was provided by the whole team of the Medea La Nostra Famiglia association (Gianluigi Reni, Emilia Biffi, Fabio Alexander Storm, Mattia Chiappini, Carla Dei, Daniele Panzeri and Noemi Miscioscia). A final thank you goes to Simona Ravera and Ettore Micheletti, from Auticon, for their efforts in the complex task of recruiting neurodivergent volunteers and for assisting them during all the experimental procedures.

Contents

List of Figures

List of Tables

List of Abbreviations

AES	**A**dvanced **E**ncryption **S**tandard
ASD	**A**utism **S**pectrum **D**isorder
APL	**A**ctive **P**reference **L**earning
BASSF	**B**oredom, **A**nxiety, **S**elf-efficacy, **S**elf-compassion, **F**low
BVP	**B**lood **V**olume **P**ulse
CAD	**C**omputer **A**ided **D**esign
CAGR	**C**ompound **A**nnual **G**rowth **R**ate
CNR	**C**onsiglio **N**azionale delle **R**icerche
DBN	**D**ynamic **B**ayesian **N**etwork
DFKI	**D**eutsches **F**orschungszentrum für **K**ünstliche **I**ntelligenz
DPIA	**D**ata **P**rotection **I**mpact **A**ssessment
DPO	**D**ata **P**rotection **O**fficer
ECG	**E**lectro**C**ardio**G**ram
EDA	**E**lectro**D**ermal **A**ctivity
EEG	**E**lectro**E**ncephalo**G**ram
ELoC	**E**xperiential **Lo**cus **o**f **C**ontrol
EMG	**E**lectro**M**yo**G**raphy
ESM	**E**xperience **S**ampling **M**ethod
GPR	**G**aussian **P**rocess **R**egression
GSR	**G**alvanic **S**kin **R**esponse
GUI	**G**raphical **U**ser **I**nterface
HD	**H**igh **D**efinition
HHI	**H**uman **H**uman **I**nteraction
HRC	**H**uman **R**obot **C**ollaboration
HRV	**H**eart **R**ate **V**ariability
HRI	**H**uman **R**obot **I**nteraction
ICI	**I**nternal **C**ontrol **I**ndex
IRCCS	**I**stituto di **R**icovero e **C**ura a **C**arattere **S**cientifico
JSON	**J**ava**S**cript **O**bject **N**otation
LoC	**Lo**cus **o**f **C**ontrol
MARSSI	**M**odel of **A**ppraisal, **R**egulation and **S**ocial **S**ignal **I**nterpretation
MCU	**M**icro**C**ontroller **U**nit
NT	**N**euro**T**ypical
NT	**NO**n **V**erbal **A**nnotator
OS	**O**perating **S**ystem
PC	**P**ortable **C**omputer
PID	**P**roportional **I**ntegral **D**erivative

PRISMA	**P**referred **R**eporting **I**tems for **S**ystematic reviews and **M**eta-**A**nalysis
RGB	**R**ed **G**reen **B**lue
RGBD	**R**ed **G**reen **B**lue **D**epth
ROI	**R**egion **O**f **I**interest
ROS	**R**obot **O**perating **S**ystem
RULA	**R**apid **U**pper-**L**imb **A**ssessment
SAM	**S**elf-**A**ssessment **M**anikin
SSI	**S**ocial **S**ignal **I**nterpretation
STIIMA	**S**istemi e **T**ecnologie **I**ndustriali **I**ntelligenti per il **M**anifatturierio **A**vanzato
TCP/IP	**T**ransmission **C**ontrol **P**rotocol/**I**nternet **P**rotocol
UA	**U**niversity of **A**ugsburg
USM	**U**ser **S**ocket **M**essaging
VSM	**V**isual **S**cene-**M**aker
WHO	**W**orld **H**ealth **O**rganization
WOS	**W**eb **O**f **S**cience
WRMSD	**W**ork **R**elated **M**uscolo**S**keletal **D**isorder
YALLAH	**Y**et **A**nother **L**ow-**L**evel **A**gent **H**andler
YLDs	**Y**ears **L**ived with **D**isability

To my wife and my family.

Chapter 1

Introduction

1.1 Problem overview

The tendency that we are witnessing today seems to point to a reality in which very soon robotic devices will actively become part of our daily lives. In fact, thanks to rapid technological innovations and the constant reduction of the cost of automation, this kind of tools keep becoming more and more widespread, with a compound annual growth rate (CAGR) of around 15% (Benchmark International, 2024). We are now used to seeing robot installations for industrial manufacturing, for which the growth still remains exponential (see Figure 1.1), but the same adapted technology is finding new fertile ground in almost every sector of application (see Figure 1.2). Just to name a few, robots are becoming a relevant presence also in healthcare (Kyrarini et al., 2021), logistics (Tutam, 2022), agriculture (Cheng et al., 2023), space exploration (Bogue, 2012), education (Atman Uslu, Yavuz, and Koçak Usluel, 2023) and even social (Breazeal, Dautenhahn, and Kanda, 2016) and domestic (Bogue, 2017) sectors. Now, a question arises: in the great push that is driving the development of this technology, are we fully taking into account the people that will use it?

It is well known that robots are exceptional solutions when it comes to repetitive operations or physically demanding and dangerous tasks since they can substitute their human counterpart, freeing them from intense labor in favor of activities requiring problem-solving and flexibility skills. However, we cannot assume that those benefits are the only consequence to the rise of this technology. In fact, given the impact that the capillary diffusion of robots may have on our society, it is of utmost importance to study the effects that interacting with these devices may have on the users in order to understand them better and leverage this newly acquired knowledge to improve their

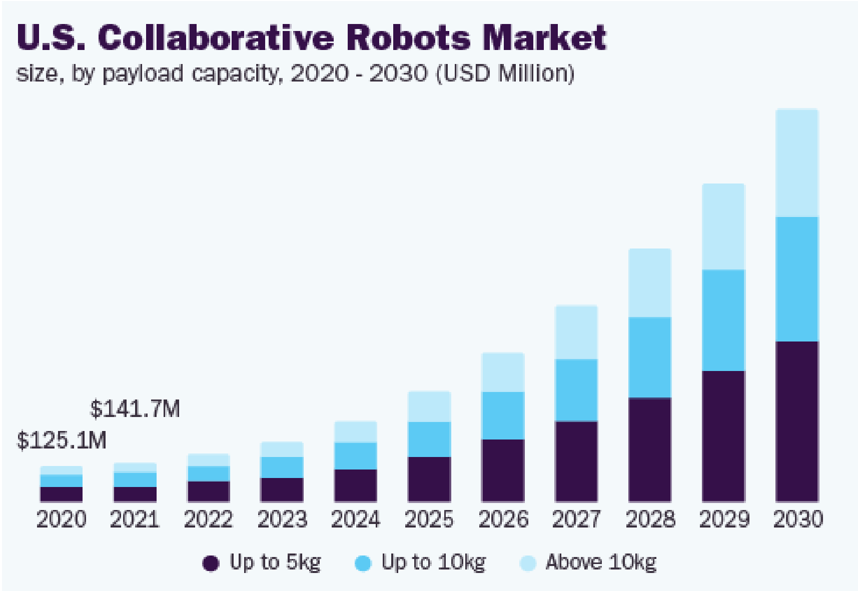


FIGURE 1.1: Historical and forecasted data (2020-2030) for industrial collaborative robots (Grand View Research, 2022).

effectiveness even further. Operto, 2019 presents the results of a survey involving 700 Italian citizens belonging to different social groups. Even though this study is not generalizable as it only pertains to Italy, it provides a glance on the perception that people have over this topic. If, on the one hand, the general public seems to value the growth of robotics, on the other a sense of fear towards the effects that it could generate on society emerges. Confirming results have also been found at the European level (Commission, 2017), where a similar survey shows how around 60% of the respondents have a positive view on robots but a majority of them also expresses society-related concerns.

The impact on employment (Carbonero, Ernst, and Weber, 2020) and inequality (Berg, Buffie, and Zanna, 2018) are of course one of the main sources of concern. Even though these topics are of utmost importance and require great discussions and further evaluations, they closely pertain to the socio-economic sphere and go out of the scope of the present project. Given the fact that this technology is already heavily present in our reality, the focus of this study is instead turned towards the experience that people have when interacting with robotic devices. In this sense, several other topics can be identified and are relevant for further exploration. First of all, not everyone has the same perception of technology: older adults are less likely to adopt new solutions and often experience frustration and difficulty of use (Heinz et al., 2013), while the younger population is more prone to adapt and feel

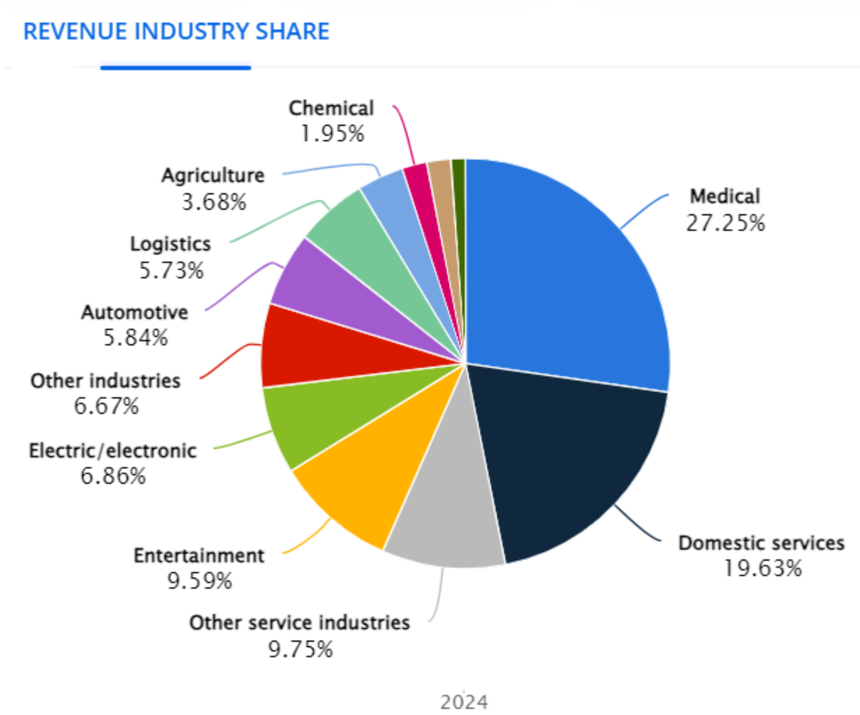


FIGURE 1.2: Robot industry share for the main fields of application (updated March 2024) (Forecast, 2024).

comfortable with such tools (Van Volkom, Stapley, and Malter, 2013). Moreover, Human-Robot Interaction (HRI) can be the cause of psychosocial stress and lead to experiences of isolation (Leso, Fontana, and Iavicoli, 2018). In fact, often robotic devices are designed to take care of duties previously carried out by men and women, with an intrinsic consequent risk of loosing most of those social nuances that are typical of human nature and sensibility. Therefore, in order to make sure that people's well-being is not negatively influenced, it is fundamental that a user-centered approach is leveraged in the design and implementation phases of robotic solutions. Ultimately, the goal should be the reconstruction of human-human interaction (HHI) experiences even within human-machine scenarios while guaranteeing safety and high performance. These aspects are particularly relevant when these technological advancements are applied to fields where close interaction is foreseen. As work represents a large portion of our day-to-day activities, the industrial sector is one that should definitely be prioritized when aiming to improve the well-being of operators directly working with collaborative robots (cobots). Similarly, the constantly growing robotic interventions in the service sector (see Figure 1.3), and especially in the medical and rehabilitation fields, are surely an example where physical and cognitive interaction with those tools plays a fundamental role.

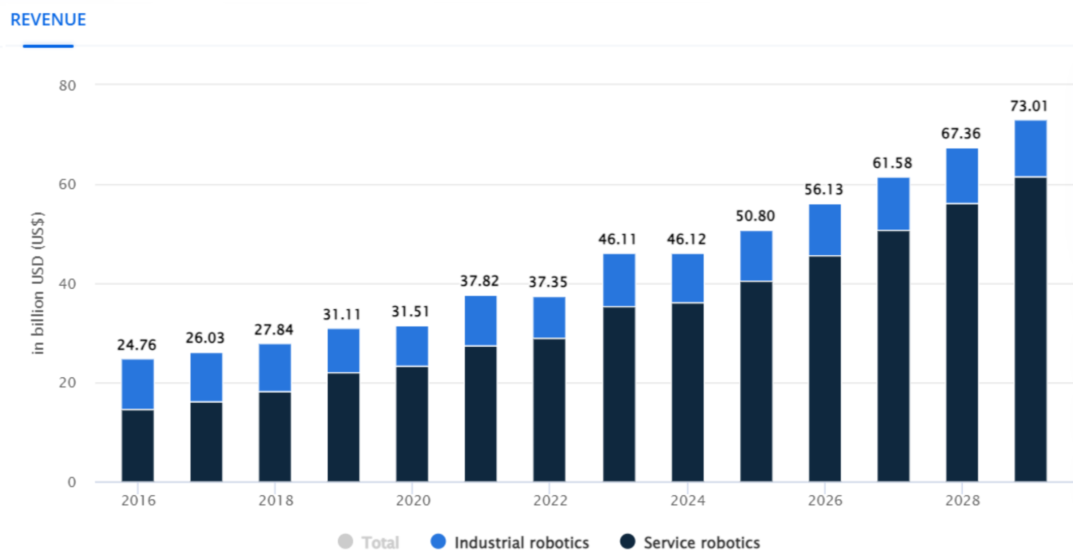


FIGURE 1.3: Historical and forecasted data (2016-2029) for industrial and service robots (Forecast, 2024).

Starting from these considerations, here a brief overview of the most relevant HRI aspects for the two mentioned application fields is reported, while more detailed information is provided in the following chapters.

The industrial sector

Thanks to the push of the fourth industrial revolution, the so-called Industry 4.0, the diffusion of automatic and digitalized solutions within the production chains of several industrial fields has risen to unprecedented levels (Lasi et al., 2014). In this context, extensive research has been carried out regarding the technical aspects of HRI, focusing especially on the topics of safety and productivity. One of the direct results of this push for innovation is the rise of the so-called cobots, which make it possible to eliminate the barrier separating the human worker from the automatic system and to put these two players closer, literally collaborating with each other (Matheson et al., 2019). Even though this approach aims to maximize the strengths of both the robots, perfect for repetitive operations, and the human workers, excellent for their flexibility and problem-solving capabilities, it does not automatically prove itself as beneficial for the operator (Liu et al., 2024; Weiss, Wortmeier, and Kubicek, 2021). In fact, the experience of a user in direct interaction with a robotic device is not simply limited to the forces exchanged between the two, but realized through a combination of vision, touch and hearing and has relevant effects also on a physiological, social and psychological level. This may be one of the reasons why another industrial revolution, the fifth one, is taking place at this very time (see Figure 1.4). The rising paradigm of Industry 5.0 puts the experience and well-being of the user at the center of focus. The European Commission defines the fifth industrial revolution as “an approach that aims beyond efficiency and productivity as the sole goals but places the well-being of the worker at the center of the production process” (Research and Innovation, 2022; Research and Innovation, 2021).

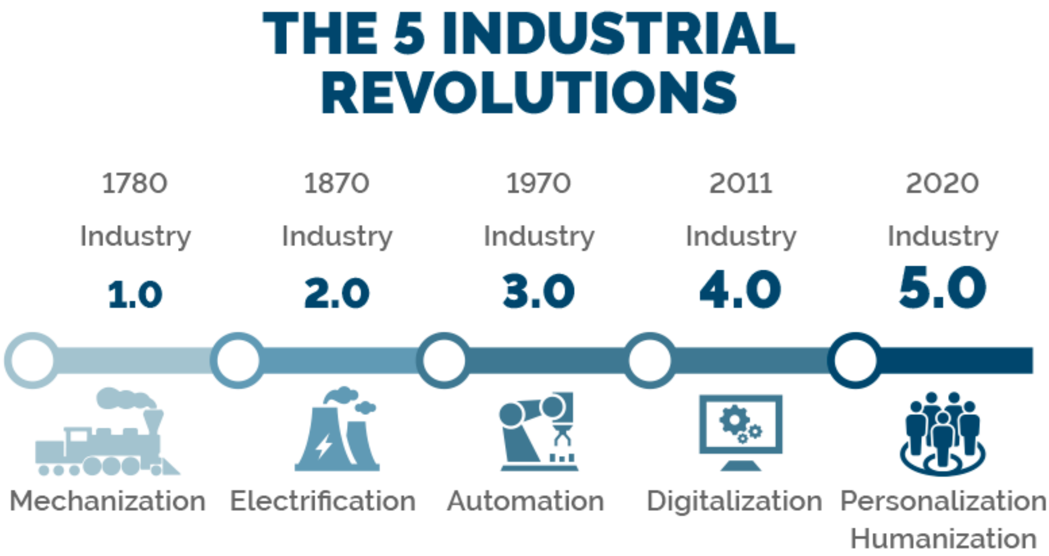


FIGURE 1.4: The five industrial revolutions (Proaction International, 2022).

In this context, some studies have already been published to understand how certain parameters can affect the user in terms of stress (Kato, Fujita, and Arai, 2010), trustworthiness (Müller et al., 2017) or dominance (Reinhardt et al., 2017). Moreover, research shows how direct contact can be effective regarding physical and psychological benefits (Thomas and Kim, 2021) and therefore ease the user’s interaction with the robotic device (Block et al., 2021). The same can also be said for the introduction of bidirectional human-robot communication. In fact, from a conceptual point of view, a virtual avatar could act as a mediator between the operator and the cobot with a subjective impression tailored on the specific user needs (Oosterhof and Todorov, 2008). These are just some of the concepts that should be addressed when aiming to implement a human-centered robotic framework capable of promoting well-being in HRI workplaces. In fact, the final goal of Industry 5.0 is to leverage the availability of all these new technologies to achieve social goals beyond employment and growth and to provide prosperity for the sustainable development of all humanity (Leng et al., 2022).

In this sense, the introduction of cobots in the working environment could also represent an opportunity of inclusion for vulnerable subjects. Among them, people with difficulties in social relationships, for instance characterized by the Autism Spectrum Disorder (ASD), are the ones that better fit into the purpose of this project (Hendricks, 2010). In fact, the fixed and predictable routine with precise task assignment that characterizes the collaborative work with a cobot may be beneficial in such a scenario. However, it is important to remember that the behavioral patterns elicited by neurotypical operators (NT) are expected to be different from the ones of operators characterized by ASD (Mondellini et al., 2023) and therefore particular attention needs to be devoted to tailoring the collaboration experience to the needs of each specific worker.

The medical sector

The industrial field is not the only one recording an exceptional rise in the rate of introduction of robotic devices. Similarly, the medical sector has seen an exponential growth in the number of installed robotic devices. As a notable example, forecasts show that the use of rehabilitation robots is rapidly speeding up (Grand View Research, 2021), probably due to the lacking number of available therapists than cannot keep up with the rate of population growth paired with a significant increase of the number of older adults. As shown in Figure 1.5, within the last 30 years the prevalence of conditions that would benefit from rehabilitation therapy has increased significantly, with the consequent rise of the number of Years Lived with Disability (YLDs). According to the last data released by the World Health Organization (WHO), about 2.4 billion people worldwide are currently living with one of those conditions, but the growing need for therapy is going largely unmet (World Helath Organization, 2024).

This is where the introduction of rehabilitation robots can be beneficial since they promise to relieve professionals from part of the physical and time-related burdens of the therapy while allowing for intense, precise and quantitative exercises (Qassim and Wan Hasan, 2020). However, the rehabilitation field is one where physical and cognitive interaction play a fundamental role leading to notable challenges in the application of robotic devices. Therefore, it is important to go beyond the direct benefits that this kind of devices provide to both patients and therapists and to explore the additional factors that come into play when a user is put in direct interaction with such tools.

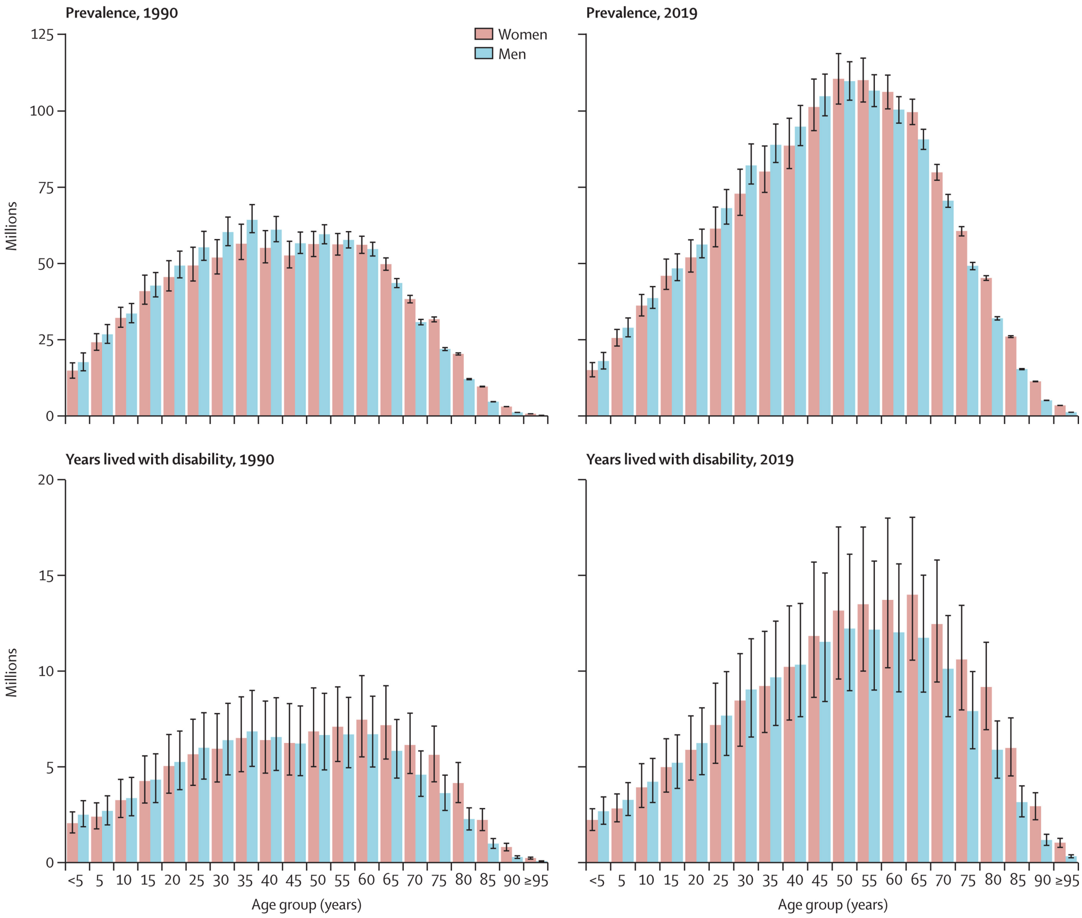


FIGURE 1.5: On the left the prevalence of conditions that would benefit from rehabilitation (top) and the consequent YLDs (bottom) in 1990. On the right, the same measures recorded in 2019 (Cieza et al., 2020).

The main research focus in this field is currently the development of devices and control algorithms optimized to provide therapy effectiveness in terms of "user's kinematic performance". However, one must consider that an effective recovery does not only come from intense and precise exercises but also from a positive active engagement of the patient (Blank et al., 2014). An effective rehabilitation robot should therefore be able to proactively adapt its behavior on the basis of the inferred overall status of the patient, in particular in terms of performance, fatigue, stress, attentiveness and engagement.

A lot of the studies carried out to provide more natural and social HRI experiences in the industrial field could, of course, be leveraged and adapted also for the robotic rehabilitation sector. Moreover, some work has already been carried out specifically for introducing affective capabilities into the control systems of this type of robots. For instance, Rivas et al., 2015 explored the possibility of detecting affective states such as tiredness, tension, pain and satisfaction for post-stroke virtual rehabilitation. Similarly, Liu et al., 2007 inferred affective cues from psychophysiological analysis to help children with ASD explore social interaction dynamics in a gradual and adaptive manner. Hoshina et al., 2020 performed preliminary experiments to achieve accurate detection of emotions such as depression and stress while Bonarini et al., 2008 presents a method for the recognition of stress from biological signals showing that it is possible to discriminate up to five levels with an accuracy up to 88.06%. In line with the contents of this project, their goal is in fact to leverage this information to successfully control socially-aware rehabilitation robots. The same can be said about the use of virtual coaches and for medical applications. For instance the Fit Track system (Bickmore, Gruber, and Picard, 2005) features the relational agent Laura, who serves as an exercise advisor. Laura engages with patients, motivating them to participate in physical activities and thereby fostering their rehabilitation progress. The SenseEmotion project (Velana et al., 2017) explored pain management strategies among the elderly, employing an avatar for crisis interventions to facilitate reassuring dialogues and support for older adults. Additionally, Giraud et al., 2021 proposed a tangible and virtual interactive system to train children with ASD in joint actions, demonstrating the broader potential of socially interactive agents in training social and motor skills relevant to neurorehabilitation.

1.2 Research objectives

Overall, the need for a more heterogeneous representation of the user interacting with a robotic device is clear. The goal of the present project is to leverage that knowledge to provide more natural and social interactions with robotic devices. In order to do so we need to study human-human interactions with a particular attention on the different needs of diverse groups of people. On this basis, there is a need to be able to quantitatively measure experience and reintroduce in the control logic the social components that are generally lost when interacting with a machine. All of this must fit into the design and validation of a generalized control framework providing the automation level that is necessary for a smooth interaction. A series of research questions therefore arise. Which are the relevant parameters influencing the experience of human-robot interaction? How can these measures be leveraged for the automatic adaptation of robotic systems? Can these logics be generalized for heterogeneous groups of people, such as neurotypical and

ASD users? Can the introduction of social-awareness capabilities and of interactive virtual characters render the experience of interaction more natural and social for the users? In order to start responding to these relevant and challenging questions, a series of research activities, grouped into three main phases, are foreseen for the project.

Phase 1: Preliminary activities

- An overview of the current state-of-the-art regarding human-in-the loop control approaches, with a particular focus on the use of social, physiological and psychological measures to enhance HRI experiences both in industrial and rehabilitation scenarios.
- The development of a generalized software architecture designed to introduce user-centered quantitative measures in the control logic driving the behavior of a socially aware robot-avatar system.
- Obtaining the ethical clearance from the ethical committee of the National Research Council of Italy, where the experimental activities will take place. In fact, due to the nature of the project, human participants need to be involved in the necessary data collection campaigns.

Phase 2: Experimental campaigns

- The setup and implementation of two lab-based use-case scenarios, one for exploring the proposed concepts in the robotic rehabilitation field and the second one to do the same in the industrial manufacturing sector.
- Leveraging the developed experimental setups to run a number of studies exploring the different research questions defined during the first phase of the project and confirming the feasibility of the proposed generalized control architecture.

Phase 3: Validations and guidelines

- The full integration of the two demonstrators and preliminary validation of their effectiveness in eliciting more natural and social HRI experiences.
- The extrapolation of useful guidelines that can be helpful in the reproduction and augmentation of user-centered robotic systems, including

the identification of the main technological and methodological limitations of the currently available tools. This step is crucial to provide a list of collateral research topics that need further efforts in order to effectively provide personalized and adaptive solutions also in out-of-the-lab use cases.

1.3 Thesis structure

The thesis is structured as represented in Figure 1.6. Chapter 2 is devoted to the analysis of the state of the art for all those features that will be part of the envisioned system: starting from the most common approaches for the inference of experience through signal interpretation and their implementation for control adaptation purposes and then moving to how robotics have been used up to now with respect to neurodivergent individuals. Chapter 3 introduces the concept of a generalized human-driven control architecture, from the requirements to the description of the specific modules making up the system. Chapter 4 shows how the mentioned generalized framework is put into practice for the development of the two lab-based scenarios that are leveraged as use-cases for the experimental activities of this project. Chapter 5 goes into the details of all the experimental campaigns carried out in order to address the several research questions that need an answer and to validate the proposed generalized framework. Finally, Chapter 6 is devoted to the discussion of the obtained results in order to draw a conclusion for the presented project, also highlighting limitations and future work.

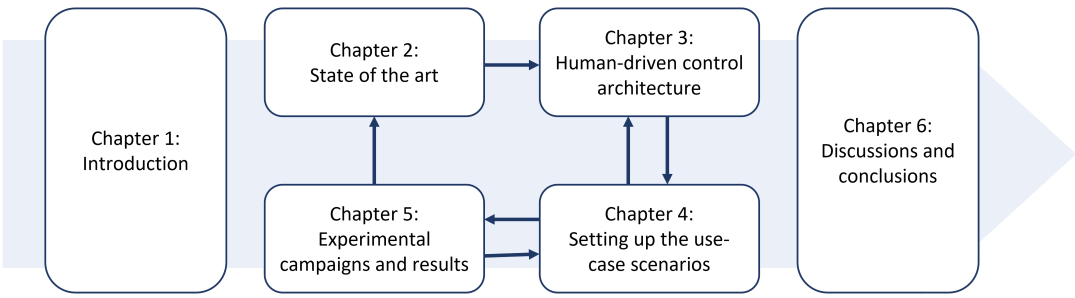


FIGURE 1.6: Schematic representation of the thesis structure showing how the chapters interact with each other.

Chapter 2

State of the art

To enable more natural and socially meaningful interactions between robots and humans, control algorithms must incorporate quantitative measures that reflect the user's experience during the interaction. These measures provide critical feedback, enabling the system to adapt dynamically to the user's needs, preferences, and emotional states. Existing work in this area can serve as a reference for designing a generalized solution, leveraging strengths and addressing weaknesses of current approaches. A comprehensive literature review is presented, starting from Section 2.1 with the common methods used for quantifying user experience and exploring their implementation in robotic control systems. Additionally, robotics offers significant inclusion opportunities for neurodivergent individuals, particularly those with ASD. To understand how this potential can be realized, it is essential to review prior uses of robotic technology with ASD subjects, as summarized in Section 2.2. Notice that, additional references specifically pertaining the experimental activities and use cases presented in Chapter 5 are reported at the beginning of each subsection.

2.1 Inferring experience for robot adaptation

The first step for the implementation of control strategies aiming to improve the interaction between a robot and a human is understanding how quantitative measures related to the user's experience can be extracted. Moreover, in order to have a complete overview of the user's status, these measures should provide information coming from different points of view, including biomechanics, physiology, psychology and social sciences. Here, the available knowledge on these multidisciplinary topics is addressed one by one.

2.1.1 Biomechanical assessments

The integration of biomechanical measures in human-robot interaction is crucial for enhancing the interaction quality and safety. These measures help in understanding the physical dynamics between humans and robots, which is essential for developing intuitive and effective control architectures.

Many studies in this field rely on the use of EMG data. For instance, Caporaso, Grazioso, and Gironimo, 2022 present a virtual reality system developed to evaluate ergonomics in human-robot cooperative workplaces. The proposed system uses surface electromyographic sensors and an accelerometer to analyze muscular activity and ergonomic status in real-time, providing workers with self-awareness of their physical conditions. Similarly, Rathi et al., 2017 introduce a flexible, metal-free electrode device used for capturing biomechanical movements in both humans and robots. The device described in the study provides qualitative and quantitative data on movements, offering applications both for the industrial and healthcare sectors. Always regarding the medical field, Kim, Vanloo, and Kim, 2021 propose a 3D origami sensing robot that uses EMG data to evaluate muscle functions. These robots demonstrate potential in providing empathetic adaptability and quality care in healthcare settings. Instead, Vera-Ortega et al., 2022 focus on a cooperative human-robot architecture for search and rescue missions, where bio-signal sensors are used to monitor stress, anxiety, and physical fatigue in responders. The proposed system facilitates remote control and communication between humans and robotic agents. With a different approach, Manjunatha, Jujjavarapu, and Esfahani, 2020 use EMG data to classify motor control difficulty in human-robot interaction, adjusting admittance control parameters accordingly. The study also employs Riemann geometry-based features and transfer learning to improve classification accuracy across sessions. EMG signals can also be used to control a robot movements in a more natural and intuitive way. Artemiadis and Kyriakopoulos, 2010 build upon this technique to control a robotic arm through a dimensionality-reduction approach that decodes muscle synergies into motion primitives. This method allows continuous control of a robot arm, demonstrating the potential of EMG for real-time robotic applications. Similarly, A hybrid approach combining pattern and non-pattern recognition strategies is used by Sbargoud et al., 2021 to process EMG signals for controlling a robotic hand. The study highlights the effectiveness of wavelet packet decomposition and artificial neural networks in decoding user movement intention.

Another common approach for biomechanical monitoring in HRI is the use of vision systems. Related studies can be split between the ones relying on reflective markers and the ones using marker-less solutions. Starting from marker-based systems, the BTS SMART device has been used to assess working ergonomics by evaluating the operational risk during tasks like object lifting and displacement. This system demonstrated high accuracy in computing risk multipliers, making it a reliable tool for ergonomic assessments (Patrizi, Pennestrì, and Valentini, 2016). Another study focused on the Vicon-460 system (Windolf, Götzen, and Morlock, 2008), which showed that careful configuration of camera setup, calibration volume, and marker size can significantly enhance accuracy and precision, achieving an overall accuracy of 63±5 micrometers under optimal conditions. In line with some of the EMG applications reported above, vision systems are also employed to improve teleoperation tasks. For instance, Minamoto et al., 2018 evaluated an interface that relies on head movements tracked by markers. The analyzed system allows control of robots with three degrees of freedom and demonstrated comparable operation times to gyro sensor-based systems.

While marker-based systems generally offer higher accuracy, marker-less systems provide sufficient precision for many applications at a lower cost, making them accessible for broader use in ergonomic assessments. The choice between these systems often depends on the specific requirements of the application, such as the need for high precision versus the practicality of deployment in various environments. In the case of the present project, the goal is to develop a generalized framework suitable for deployment also in real-life scenarios, where marker-less systems could fit well. For this purpose, some studies focus on solutions aiming to improve the performance of this kind of systems. Among all, Martini et al., 2024 attempted to improve marker-less human-robot interaction by addressing errors in human pose estimation and depth cameras. The proposed filtering pipeline refines 3D human poses using an RGB-D camera, reducing robot jittering and enhancing interaction smoothness. Similarly, Liang et al., 2019 developed a marker-less pose estimation system for construction robots using a deep convolutional network. The system estimates both 2D and 3D poses, demonstrating capabilities in proximity detection and object tracking, although occlusion remains a challenge. Another method is presented by Jatesiktat et al., 2024, using anatomical landmarks to improve 2D keypoint annotation accuracy in marker-less systems. The approach involves training a deep neural network with synchronized RGB cameras, achieving a mean Euclidean error of 13.23

mm, comparable to marker-based systems. Leveraging on similar data conditioning techniques, it is possible to employ marker-less devices in robotic systems reliably. For instance, Lagomarsino et al., 2022 follow an approach that monitors the operator's upper body kinematics on the basis of the input images of a low-cost stereo camera and artificial intelligence algorithms (i.e., head pose estimation and skeleton tracking) with satisfactory results if compared to state-of-the-art offline measurements. Similarly, Shafti et al., 2018 propose a solution where, by continuously observing a human user's posture, it is possible to invoke appropriate cooperative robot movements so that the user's posture is brought back to an ergonomic optimum.

2.1.2 Physiological signals

Physiological data provide objective and continuous insights into emotional states, cognitive load, and engagement levels that may not be fully captured through self-reports or observations. By integrating this information, robotic systems can adapt more effectively to users' needs, creating interactions that are more natural, responsive, and user-centered.

Several studies have focused on estimating the user's mental state during interactions with robots using physiological signals. For instance, heart rate variability has been used to assess mental fatigue, allowing systems to adapt robot behavior to reduce user workload (Villani et al., 2019). Similarly, HRV was used to monitor stress levels in operators during teleoperation tasks, with assistive technologies introduced to reduce mental workload and improve performance (Landi et al., 2018). Another study highlighted the use of HRV signals at different time scales to classify mental workload, showing that specific classifiers could achieve high accuracy in workload assessment (Shao et al., 2021). Electroencephalography is another critical tool for monitoring cognitive workload in HRI. In a tele-exploration scenario, EEG data was used to predict operator performance and situation awareness, with brain-based features proving effective in assessing mental workload and distraction (Memar and Esfahani, 2018). Additionally, EEG was employed to study cognitive workload's impact on error awareness in physical human-robot collaboration, revealing that increased workload diminished error awareness, which could compromise safety (John et al., 2024). Eye-tracking metrics have been shown to be reliable indicators of mental workload in HRI too. In a study involving physical human-robot collaboration, eye-tracking measures such as gaze entropy and pupil diameter were sensitive to task difficulty and could predict performance outcomes with reasonable accuracy (Upasani et al., 2023). Another study found that eye gaze was the best physiological indicator of cognitive workload, outperforming other signals like EEG and arterial blood pressure in a simulated driving study (Aygun et al., 2022).

Another relevant aspect to enhance the effectiveness of human-robot interactions is the ability to detect and react to the level of user engagement. With this goal, galvanic skin response (GSR) and skin temperature have been employed to estimate user engagement with an accuracy of 84.73% during interactions (Provost et al., 2007). Pruss et al., 2023, instead, used EEG signals to demonstrate that adaptively timed robot interventions, based on detected engagement lapses, are more effective in restoring user engagement compared to random interventions. Advanced engagement models have been developed to enhance the naturalness and comfort of HRI. The model proposed by Lu et al., 2024 utilize eye gaze, head pose, and action recognition to determine optimal interaction moments, addressing issues like eye contact anxiety. The system has been validated in real-world scenarios, such as retail environments, demonstrating its potential to improve user experience. Additionally, continuous engagement assessment models, using CNN and LSTM networks, have been proposed to compute engagement levels from video streams (Duchetto, Baxter, and Hanheide, 2020). These models have shown success in predicting engagement across different datasets and environments, providing a tool for measuring engagement in various HRI settings. Research by Rihet, Clodic, and Roy, 2024 has also examined the impact of robot-induced noise on physiological measurements. It was found that EEG and PPG signals were affected by such noise, whereas EDA was not. Adjusting preprocessing parameters improved the accuracy of EEG signal interpretation, underscoring the importance of signal selection and preprocessing in HRI. For this purpose there exist tools, like the HRI Physio Lib (Kothig et al., 2021), that are designed to facilitate the acquisition and analysis of physiological signals to create adaptive HRI scenarios. The cited library, for instance, allows for the synchronization and processing of signals to enable robots to respond to detected human states, such as engagement and stress, thereby enhancing interactive experiences.

Another significant role of physiological monitoring is enhancing human-robot interactions by detecting and responding to users' emotional states. For instance, in a manufacturing setting, Canete, Gonzalez-Sanchez, and Guerra-Silva, 2024 used consumer-grade EEG devices to monitor operators' brain signals to infer emotional and cognitive states. In the study, the robotic arm adapts its behavior in real-time based on the operator's stress and concentration levels, using RGB lighting to signal when stress levels are high. Also Swangnetr, 2010 used signals such as heart rate, galvanic skin response, and facial electromyography to classify emotional states in patient-robot interactions. These signals have been shown to correlate with valence and arousal, providing a basis for real-time adaptation of robot behaviors to enhance user experience. Another framework proposed by Singh et al., 2019 uses physiological data, facial expressions, and eye movements to estimate emotional states during robot teleoperation. The system classifies emotional states such as resting, stress, and workload, and dynamically updates the user interface in real-time. Since many approaches encompass the training of neural network models on physiological signals it is important to also explore the best strategies for the purpose. The research carried out by Gallardo et al., 2024 highlights that a general model fine-tuned with specific subject data performs better in predicting emotions. Often, overt behaviors do not align with emotional responses. It is therefore key to emphasize the importance of understanding internal states, as addressed by Staffa and Rossi, 2022. In fact, a systematic review highlighted the growing interest in using physiological monitoring to assess user experience during interactions (IRIARTE, ERLE, and Etxabe, 2021). EEG and GSR combined with ECG are among the most commonly used tools, indicating a trend towards integrating physiological data to improve experience evaluation.

2.1.3 Social cues

Also social cues can provide rich, real-time information about users' emotions, attention, and engagement. By leveraging this type of information, HRI systems can create interactions that feel more natural and attuned to human behavior, improving the effectiveness and acceptance of robotic systems in various social and collaborative settings.

For instance, social cues can play a crucial role in improving the fluency of human-robot collaboration. In a study involving a pick and place task, it was found that when robots used head movements or gestures to indicate non-reachability, humans responded more naturally and efficiently (Romat et al., 2016). Similarly, gaze cues have been shown to facilitate cooperation by improving human response times in task-oriented situations (Boucher et al., 2012). Effective gaze-based cueing requires overcoming in-attentional blindness. Lee et al., 2020 demonstrated that guiding a user's attention through eye contact before signaling cues significantly improved task performance, underscoring the need for sophisticated interaction designs to capture user attention effectively. With an interesting approach, Romeo et al., 2021b leveraged a deep learning architecture to predict apparent personality traits from body language cues, such as head pose and gestures. This approach allows robots to adapt to users by inferring personality traits, thus personalizing human-robot interactions. In general, body posture and head pose are significant social signals used to initiate and terminate interactions. Another study by Gaschler et al., 2012 shows that by training models like Hidden Markov Models on these cues, robots can recognize and respond to typical social behaviors with high accuracy, improving their ability to interact appropriately in social settings. The impact of social cues on decision-making has also been explored by Parenti, Belkaid, and Wykowska, 2023, revealing that incongruence between pre- and post-decision social signals from robots can significantly influence human task performance. This suggests that understanding and aligning social expectations is key for effective human-robot interactions. Also monitoring affective cues, such as anxiety, in real-time can enhance human-robot interaction. In a robot-based basketball game proposed by Liu, Rani, and Sarkar, 2006, adapting the game difficulty based on the participant's anxiety led to improved performance, highlighting the benefits of responsive interaction frameworks.

Robots can also leverage social cues to enhance their learning processes and overall abilities in social interactions. For example, the iCub robot uses mutual gaze, gaze following, and speech to learn new objects through interaction with a human teacher, demonstrating the potential of social cues in creating more natural and robust learning environments (Lombardi et al., 2022). Lee et al., 2023 used large language models to generate these cues and found that robots could engage in more context-aware and authentic interactions, emphasizing the role of both verbal and non-verbal cues in developing empathetic robots. Additionally, Fiore et al., 2013 showed that proxemic behavior, which involves the robot's use of space, affects how humans perceive the robot's social presence and emotional state. However, gaze behavior alone was not found to be significant in altering these perceptions, highlighting the importance of considering various social cues in designing robots to enhance their perceived social presence.

2.1.4 Psychological measures

Psychological measures, such as self-reports on emotions, stress levels, or engagement, provide subjective insights that complement physiological and behavioral data. By integrating psychological metrics, HRI systems can better assess user satisfaction, cognitive load, and emotional states, enabling robots to adapt their behavior in ways that align with human needs and expectations.

As a notable example, studies have developed scales like the Negative Attitudes Toward Robots Scale (NARS) (Nomura et al., 2008) and Robot Anxiety Scale (RAS) (Nomura et al., 2006a) to measure anxiety and negative attitudes, which are crucial for understanding communication avoidance behavior in human-robot interactions. In fact, trust is a critical factor in human-robot interaction. Li et al., 2024 showed that users' openness and robot reliability significantly affect trust levels. Users with low openness tend to exhibit lower trust and allocate more attention to monitoring the robot, suggesting that personality traits influence trust dynamics. Additionally, the propensity to trust and state anxiety mediate trust levels, affecting comfort distance and interaction behavior (Miller et al., 2021).

Emotional processes during human-robot interactions, such as those observed in cognitive testing, show no significant differences between human and robot administrators in terms of affective states and cognitive performance. However, non-verbal behaviors like gaze patterns differ, indicating unique interaction dynamics with robots (Desideri et al., 2019). Furthermore, perceived social intelligence of robots, measured through new scales, correlates with social competence and predicts positive feelings and interaction willingness (Barchard et al., 2020). The development of psychometrically validated questionnaires to measure social robot acceptability is still in its early stages. The questionnaires proposed by Krägeloh et al., 2019 assess various factors, including ethical issues, especially in therapeutic contexts with children. Continued psychometric work is necessary to enhance the reliability and validity of these measures. Additional interesting insights are provided by research on active physical human-robot interaction, aiming to quantify human physical and mental states during interactions. The findings presented by Hu et al., 2022 suggest that active robot actions can cause measurable changes in users' data, which relate to their perceptions and personalities. This understanding can inform the development of pHRI controllers that consider both physical and mental states.

Overall, the systematic review provided by Vagnetti et al., 2023 identified 27 instruments designed to assess psychological dimensions in relation to social and domestic robots. These instruments primarily focus on structural validity and internal consistency, but often neglect other psychometric properties such as measurement error and responsiveness. The review highlights the need for developing new instruments with rigorous methodologies to ensure reliable and valid measures for assessing psychological dimensions in human-robot interactions. Among possible alternatives, Huang et al., 2020 explored embedding psychological test questions into casual human-robot conversations to profile users. This method showed strong correlations with traditional written tests in young adults, suggesting its validity. However, the correlation was moderate in older adults, indicating potential limitations for certain populations.

2.2 Robotics and Autism Spectrum Disorder

As already mentioned, current advancements in the robotic field represent a great inclusion opportunity for neurodivergent individuals, particularly those characterized by the Autism Spectrum Disorder. However, before attempting to translate this opportunity into reality, it is important to understand how this kind of technology has been used for this purpose up to now.

Overall, the vast majority of the publications in the field focus on the use of robotics as an intervention tool for children with ASD. The primary aim is to enhance social interaction, communication, and educational outcomes for ASD subjects through robot-assisted therapies. Various aspects are explored in this sense, such as the effectiveness of social robots in therapy, the development of robotic systems for skill training, and the integration of robotics in educational settings for ASD children. For instance, many studies report that robots can significantly improve social interaction and communication skills in children with ASD. Children often perform better with robot partners than human partners, showing increased social behaviors and reduced repetitive actions during robotic sessions (Pennisi et al., 2016; Alghamdi, Alhakbani, and Al-Nafjan, 2023; Saleh, Hanapiah, and Hashim, 2020). Robotics has been used to teach various skills, including social, motor, and cognitive skills. Studies have shown that robotic interventions can effectively teach skills such as eye contact, facial emotion recognition, and even coding, with children maintaining these skills over time (Santos et al., 2021; Knight, Wright, and DeFreese, 2019). Robots have been found to be useful mediators in behavioral interventions, showing similar positive effects as human-facilitated interventions. They help in reducing behavioral and emotional symptoms and improving the ability to play and interact socially (Yun et al., 2017; DiPietro et al., 2019). The population involved in the above studies primarily includes children with ASD, typically ranging from ages 2 to 16 years and they often involve small sample sizes, with some including as few as 8 participants, while others review larger datasets from multiple studies. Additionally, only some of the studies also consider the inclusion of typically developing children for comparative purposes. It is clear that there is an unbalance in the way robotics have been applied up to now with people of the spectrum: the most common use-case is therapy with children which, even though remaining an highly promising and relevant application, does not realize the above mentioned inclusion opportunity that also adults could benefit from.

Only a few studies exist involving robotics and adults with ASD and, once again, they primarily focus on their use as therapy tools for enhancing independent living skills and social interactions. In these terms, social robots have been shown to improve social behaviors, reduce repetitive behaviors, and enhance spontaneous language during therapy sessions (Pennisi et al., 2016). Additionally, robotics interventions have demonstrated potential in enhancing independent living skills among young adults with ASD, although further research is needed to explore the generalization and maintenance of these skills (Sarri and Syriopoulou-Delli, 2021). The involvement of adults with ASD in designing interventions has also provided valuable insights into the practical implementation of robotics in therapy (Huijnen et al., 2017).

While these studies provide valuable insights, they largely overlook adults with ASD, despite the potential benefits robotics could offer this demographic. However, many robotic applications, such as collaborative robots in industrial settings, hold promise as powerful tools for promoting social and workplace inclusion for adults with ASD. The limited attention to this area leaves a critical knowledge gap, underscoring the need for further research to explore how robotics can support the unique needs and challenges of adults with ASD, ultimately fostering greater societal and economic inclusion for this vulnerable group.

Chapter 3

Human-driven control architecture

3.1 Schematization of the social mechanism

Without even noticing, whenever we interact with someone an innate social mechanism is activated and calibrates our actions and reactions depending on the behavior of the person in front of us. The sophisticated laws that rule this interaction are deeply rooted in one's culture, education and past experiences. Figure 3.1 tries to schematize said mechanism in order to understand it better and possibly reproduce it within the control system of automated machines.

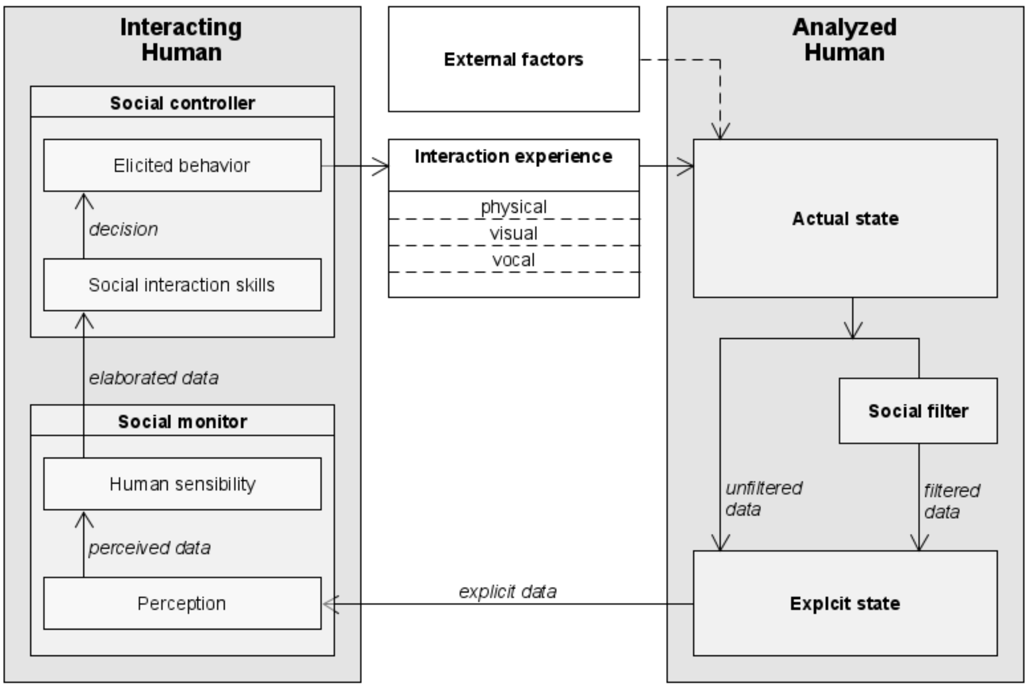


FIGURE 3.1: A schematic representation of the social mechanism behind human-human interactions.

Let's consider a dyadic setting involving a first *Interacting Human* actor who adapts his/her behavior in order to positively engage with a second *Analyzed Human* actor, whose overall state is affected by the interaction itself. Starting from the latter, the merge of physiological and psychological aspects make up an overall condition that is referred to as *Actual state* in the provided scheme. However, extensive research in social psychology has shown that the psychological, affective and emotional expressions visible from the outside are often different from what the person actually feels from the inside, since they get reshaped by a sort of *Social filter* (Ellwood, 1901; Deal, 2007; Hull, 1943). Instead, physiological reactions to one's actual state cannot be controlled but, on the other hand, they are often not usable within the context of human-human interactions. The sum of these filtered and unfiltered data make up a so-called *Explicit state*, which can be perceived and elaborated by the second interacting actor depending on their sensibility. Considering Figure 3.1, these innate human abilities are referred to as *Social monitor* and directly connected to a so-called *Social controller*. In fact, depending on the result of one's internal interpretation of the explicit state of the actor under analysis, a decision is taken on how to act/react with a consequent effect on the interaction experience from a physical (e.g., touch), visual (e.g., gestures, facial expressions) and vocal (e.g., utterances) point of view. Of course, this interaction is not the only factor producing an effect on the actual state of the analyzed actor. Several unknown *External factors*, such as personal issues, act as disturbance in the system and make the whole social mechanism even more complex and unpredictable.

As mentioned, the goal of the present project is to try and reproduce within the control system of technological solutions the social mechanism behind HHI just presented, aiming to render HRI a more natural and social experience. Considering the current technological advancements, computing power and available sensor devices make it possible to easily leverage those unfiltered physiological data that cannot be accessed in normal human interactions. On the other hand, the reconstruction of the presented *Social monitor* and *Social controller* capabilities poses an extremely complex challenge that will be extensively addressed by the present project.

In simplified terms, the main research question is to understand if and how it is possible to implement an "adaptive experience controller" such as the one reported in Figure 3.2. Let's now consider a second dyadic setting, involving a *Robotic system* and *Human user*. A technological version of the *Social monitor* is devoted to produce a measure of the interaction experience as close as possible to the user's actual one. The "error" between the measured and optimal experience then becomes the driver for a reconstructed *Social controller*, dispatching commands to the robotic system itself.

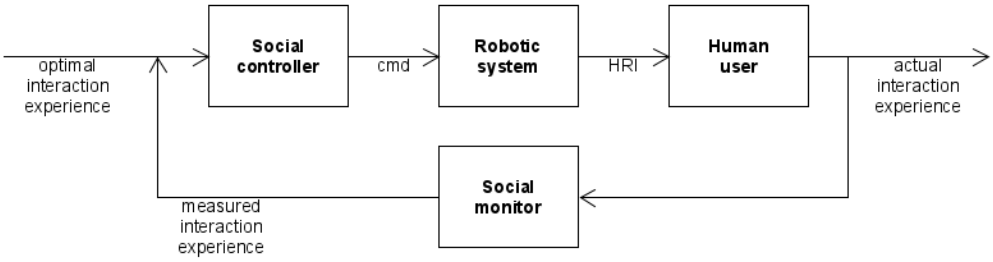


FIGURE 3.2: A schematic representation of the concept of an "adaptive experience controller".

Building upon our understanding of human-human interactions and past experiences in the development of human-driven control architectures, a series of aspects should be taken into particular consideration:

1. Human-robot interaction often only involves the physical sphere through direct or indirect contact between the two entities. However, when trying to reproduce the experience of a human-human interaction, visual and auditory aspects are also involved. These are fundamental to recreate the lost social nuances and should therefore be implemented into the system to enhance its interaction capabilities.

2. The user should be monitored in a heterogeneous way, both in terms of objective and subjective measures. Where possible, noninvasive sensors should be preferred to minimize the inconvenience of setup and promote a sense of natural interaction.

3. All raw data coming from sensors should be fused into a single and comprehensive representation of the user's experience. A model of optimal state should be produced and adapted to the needs of each single user in order to leverage the available quantitative information for automatic adaptation purposes, pushing towards a more social human-robot interaction.

The scheme depicted in Figure 3.2 is simply a general concept and only serves as a starting point for the actual implementation of the envisioned human-driven control architecture. Starting from said concept and the requirements listed above, the rest of this Chapter is devoted to a more concrete and detailed description of how said concept can be translated into practice.

3.2 Conceptual architecture

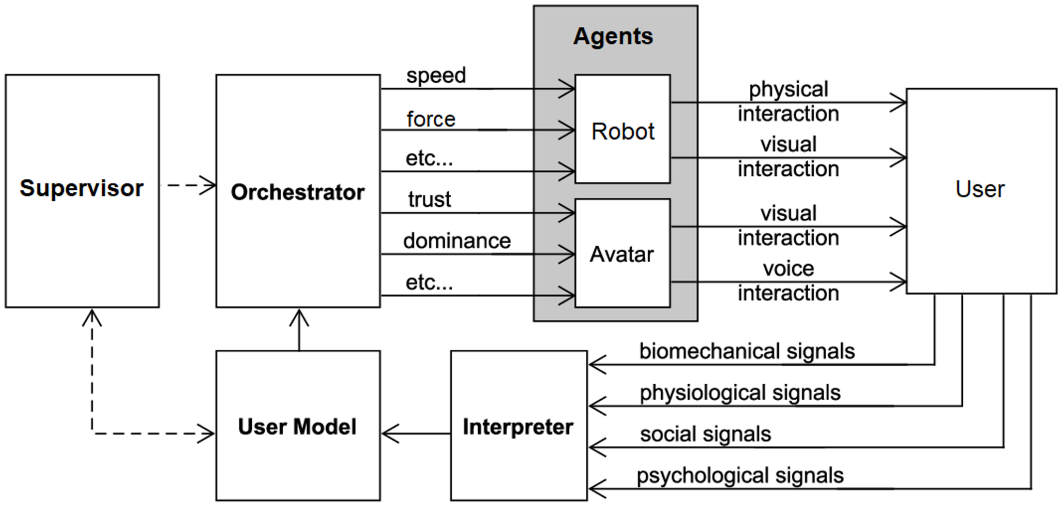


FIGURE 3.3: Schematic representation of the generalized human-driven control architecture.

In order to put into practice the conceptual mechanisms of social interaction presented in Section 3.1, a generalized human-driven control architecture is reported in Figure 3.3. As depicted, a series of fundamental modules are interconnected by solid arrows that represent either the stream of measurable/controllable parameters or a specific type of interaction. The collection of these elements allows for the definition of a closed control loop, running along with the execution of the human-robot task.

The purpose of this framework is to offer the *User* a natural HRI experience characterized by social and empathic aspects. To achieve this goal, the envisioned system includes both a generic *Robot*, and an interactive virtual *Avatar*. This additional feature allows to enrich the interaction with the User, adding gaze, gestures, and talk capabilities to the platform, with physical, visual and voice interaction modes. The system is aimed at achieving high levels of integration between the Robot and the Avatar, so that the latter can be considered a virtual representation of the intelligence of the robot in humanoid form. In these terms, the User can be said to interact with a unique entity, represented by the merge of the Robot and the Avatar.

For this purpose, the behaviors of the Robot and of the Avatar are coordinated by the *Orchestrator* module. This module has knowledge of the task to be carried out and is in charge of dispatching information to control both the Robot and the Avatar coherently and consistently. In doing that, the resulting behavior of the robotic system is tailored on the User's overall state in order to promote a positive and personalized interaction experience. A representative list of the main parameters available for adaptation have been identified through the analysis of the literature, presented in Chapter 2. Just to name a few, the speed, acceleration, distance from the User and force of interaction of the Robot and the traits of trust, dominance, gestures and vocal interventions of the Avatar can be tailored in a coordinated fashion on the basis of the interpreted signals.

The Orchestrator tunes the behavior of both the Robot and the Avatar on the basis of the heterogeneous representation provided by the *User Model*. This component requires general knowledge about the User (e.g., age, sex, height, weight and any other useful information) in order to access the right clustered information. Moreover, it allows both physical (e.g., fatigue thresholds) and cognitive (e.g., affective states, regulation strategies) modeling of the User, by processing the high-level measures that it is given as input in real-time during the execution of the task.

In order for the Orchestrator to work properly, a wide set of heterogeneous high-level information is required, spanning from physical and mental energy to psychological and social data. Closing the control loop, raw data is therefore collected from the User through sensors and questionnaires and then processed and elaborated by the *Interpreter* module. In particular, four main categories of signals, aiming to provide a comprehensive representation of the overall state of the User, are foreseen:

- <u>Biomechanical signals</u> are collected in a noninvasive way through markerless tracking systems. For instance, cameras can be used to track the main skeletal joints of the user in terms of position and velocity. This information is then used by the Interpreter module to infer higher-level quantities such as joint power, physical energy consumption and fatigue;
- <u>Physiological signals</u> are collected using wearable sensors. Measures such as ECG, HRV and EDA are collected in real-time during the execution of the task and fed to the Interpreter module in order to extract

information such as stress, frustration and mental energy consumption;

- Social signals are, once again, collected using non-invasive systems such as cameras. For instance, face-cropped videos of the user can be recorded and sent to the Interpreter module which is in charge of inferring social information such as valence, arousal and gaze. Additional labeling of the User's facial expression with the so-called Ekman's atlas of emotions (Ekman, Freisen, and Ancoli, 1980) could provide more detailed insights of the emotional sphere;
- Psychological signals are also considered thanks to both ad-hoc and standardized questionnaires. These information are useful as a subjective measure of the quality of experience and can be leveraged by the Interpreter module as validation for the objective, quantitative and elaborated measures mentioned above.

Always with reference to Figure 3.3, a second outer loop is represented using dashed arrows. In fact, the presence of a human *Supervisor* is still required to make sure that the system is set up properly and to intervene in case the adjustment of some parameters is required. Regarding the present project, this role is covered by the researcher that oversees the experimental sessions and makes sure that the system works properly with each single participant. However, considering a future application of this technology in real-world scenarios, the Supervisor could either be a physiotherapist or a production manager, depending on the field of application. Starting from the robotic rehabilitation scenario, a human physiotherapist still has a central role in the proposed approach. In fact, professional expertise is required for the patient's initial assessment (e.g., residual mobility, attention span), used to define the backbone of the User Model and a selection of suitable exercises. Moreover, an initial calibration process is useful in order to learn from the physiotherapist how to optimally balance the target execution performance for the exercise and the social experience for the specific patient. Moving to the industrial application, instead, a certain degree of freedom in the assignment of subtasks between the Robot and the User and in the disposition of all the components inside the workcell may be required to further improve the level of personalization. The mentioned outer loop is designed to serve exactly this purpose. The Supervisor analyzes all the data logged during the collaboration and proposes a reorganization of the task based on the balance between production requirements and the User's experience. The same can be done regarding the choice and positioning of all the components inside the workcell, especially important for Users characterized by ASD who are more sensible to the general organization of resources (Hayward, McVilly, and Stokes, 2019).

All software and hardware components will be included in the architecture as nodes in a Robot Operative System (ROS) framework (Quigley et al., 2009). ROS is a state-of-the-art open source tool that allows developers to build robotic applications in a modular and flexible way, enabling the deployment of generalized architectures ready to be plugged into different hardware solutions. This approach is perfectly suited for the present project as it is necessary that the same generalized human-driven architecture can be leveraged for different robots and sensors (e.g., industrial cobots, rehabilitation robots). In particular, regarding the interface of sensors with the rest of the framework, the Social Signals Interpretation tool (Wagner et al., 2013) will be used as it provides a flexible architecture to construct pipelines that handle data of different nature and with multiple modalities. Finally, even the definition of the task, the management and synchronization of all the sub-tasks and the coordination between the Robot and the Avatar foreseen for the system require a powerful and flexible high-level state machine. For this purpose, the Visual SceneMaker (VSM) tool (Gebhard, Mehlmann, and Kipp, 2012) is selected for its authoring, orchestrating, and executing capabilities even in complex scenarios. By building a bridge between these three tools it is therefore possible to actually implement the envisioned human-driven architecture feasible for deployment both in industrial and rehabilitation scenarios.

3.3 In-depth description of the modules

After the overview provided in Section 3.2, here a more detailed explanation of the fundamental software blocks is reported. In particular, an implementation plan for each of the modules is presented, explaining the tools that can be leveraged for the purpose and how they can be interfaced with the rest of the system to achieve seamless interaction between all the components of the generalized human-driven control architecture. Notice that, the order in which these elements were presented in Section 3.2 went through the control loop in an anti-clockwise direction in order to provide a clearer explanation of the concept. However, the logic for the implementation of those components starts from the lower-levels (basic functionalities) and then moves to the higher ones (building on those basic functionalities). For this reason, the order of presentation for the following sections is flipped backwards and follows the control loop in a clockwise direction, starting from the data collection and interpretation used to build the user model and then moving to the management of the interaction controlling the behavior of the robot-avatar system interacting with the user.

3.3.1 Interpreter

The Interpreter module is responsible for taking as input a series of raw signals (biomechanical, physiological, social and psychological) and using them to extract high-level indicators related to the experience of interaction between the User and the system (Robot+Avatar). Overall, four high-level indicators have been identified in order to have a complete and heterogeneous representation of the user during the interaction: mental energy, physical energy, social and affective state and psychological state. A detailed description of these indicators is reported below.

Mental energy

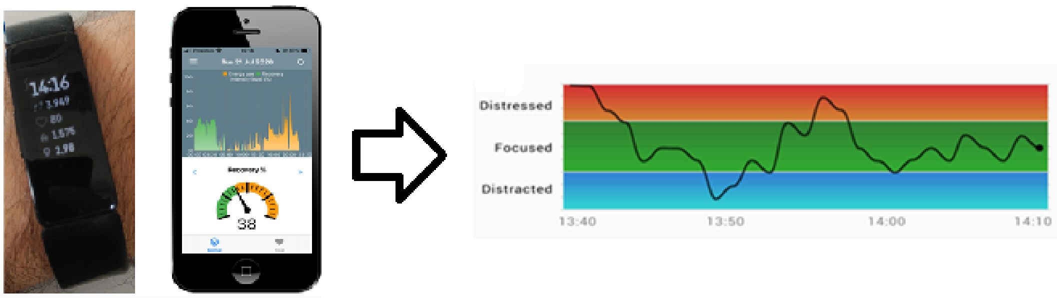


FIGURE 3.4: The combination of a FitBit activity tracker and the MindStretch app allow for continuous monitoring of the User's mental energy.

While the User is interacting with the system, his/her physiological responses (e.g., heart rate) and movement (e.g., steps) are monitored by means of wearables, such as the FitBit activity tracker. Using these physiological variables as input for the BioRICS' Mindstretch application (BioRICS, 2021), shown in Figure 3.4, it is possible to determine the metabolic energy use and/or recovery for mental tasks exhibited by the User in real-time. Mental energy use and recovery is a metric, expressed in the Mindstretch app as a percentage, which relates the mental energy wielded while performing a cognitive task to the mental energy baseline level defined for that individual. If the cognitive task is demanding, Mindstretch monitors the mental energy used by the individual to perform it. When there is no mental effort required to perform such task, Mindstretch monitors the mental energy recovery induced by that task (Smets et al., 2013; Norton et al., 2018).

The User needs to wear continuously (day and night) the FitBit, on average, for 3 days prior to ensure that the Mindstretch algorithm is fully adapted to the individual subject. Combining the mental energy monitored by Mindstretch, together with a performance metric, defined according to the specific task, allows defining the focus or Eustress zone of the User while performing such a task (Taelman et al., 2016). This mental focus zone is defined as the zone of mental energy use exhibited by the User when performing most efficiently the task (Selye, 1956). This focus zone is individually different per each User and will vary within the same day. Deviations of the mental energy exhibited by the User from this estimated focus zone can be used as an indication of distress or distraction, making the User go out of focus from the task and, thus, inducing a drop in attention and performance (Joosen, Exadaktylos, and Berckmans, 2015).

Physical energy

FIGURE 3.5: Two RGB-D cameras capture the User from two different points of view and the Interpreter uses the depth video to extract the skeletal joints position, useful to infer an index of physical energy consumption and fatigue.

In the proposed architecture, visual data is acquired using a set of RGBD cameras looking at the area shared between the User and the rest of the system from different points of view to contrast possible occlusions. In particular, the Microsoft Azure Kinect DK (Microsoft, 2021a) cameras have been selected, since they can provide high image resolution, up to 2560×1440 pixels at 30 Hz. Moreover, the Microsoft Azure Kinect Body Tracking Library is leveraged to track the User and estimate the 3D position of his/her skeletal joints with high accuracy and reliability, and low uncertainty (Romeo et al., 2021a).

For each camera, four streams are published to the ROS network delivering the compressed RGB image, the depth map, the depth map rectified in the color space geometry, and the skeletal data. Figure 3.5 shows an example of two synchronized cameras looking at the same area and tracking the skeletons of two Users. Note that recorded videos are encrypted via a 256-bit Advanced Encryption Standard (AES) to ensure the mandatory data security due to privacy reasons. The obtained information is then exploited to perform an online computation of the kinematics and the dynamics of the upper-limb (Scano, Molteni, and Molinari Tosatti, 2019) following the inverse dynamic approach (Dumas, Aissaoui, and Guise, 2004). Articular angles, velocities, accelerations and torques are used to provide an estimation of the exerted joint power and energy expenditure. This data represents the basis of the estimation of measures and parameters of effort and fatigue during the use of the envisioned system, including time-to-peak (Emery and Côté, 2012), range of motion alteration and effort related to energy expenditure. Furthermore, exploiting the NASA Anthropometric Tables (NASA, 2021) and tracked data, this module estimates the volumes occupied in space by the User and sends them through ROS to the robot controller, making them available for collision avoidance purposes or for other forms of interaction with the Robot or with the Avatar.

Social and affective state

To support a pleasant interaction experience, the envisioned system has to adjust the interaction in the case of suboptimal mental states, such as stress or boredom. To recognize such states from the the User's social and affective signals, the SSI tool (Wagner et al., 2013) is used to enable the recording, analysis, and recognition of human behaviors based on social and affective signals such as gestures, facial expressions and emotional speech or physiological signals such as Heart-Rate Variability (HRV). To this end, SSI allows to interface with and extract data from external sensors. Thanks to its modular architecture, the data processing in SSI is performed through pipelines consisting of a sequence of autonomous components that allow parallel and synchronized signal processing. Additionally, SSI supports machine learning pipelines for the execution of pre-trained models as well as on-device training of simple online learning classifiers, such as Naïve Bayes. This is especially useful for creating machine learning models that can be adapted to the individual User behavior over time.

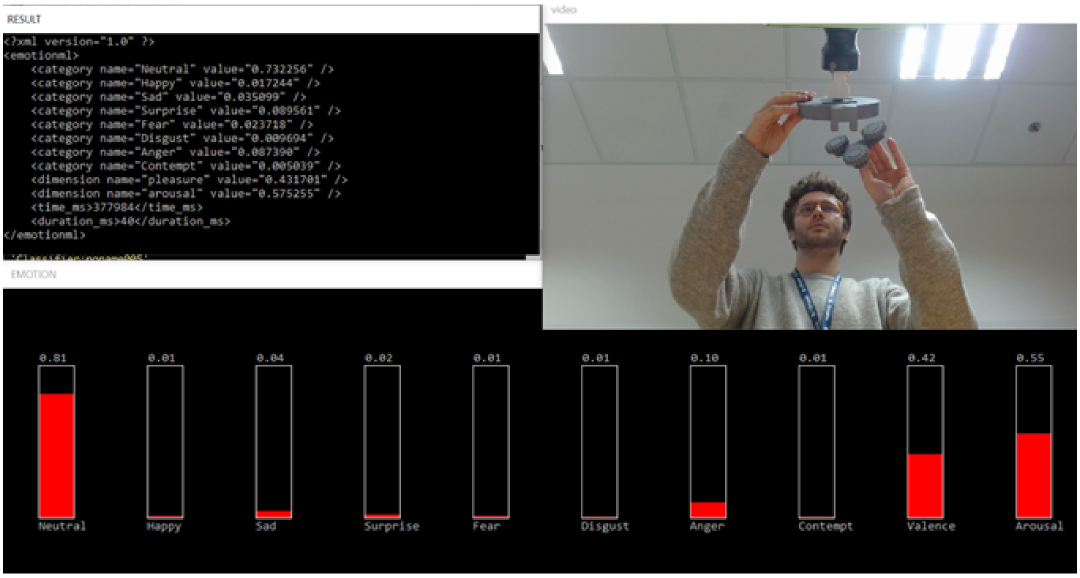


FIGURE 3.6: An RGB camera captures the User during the interaction with a robot and the Interpreter uses the video stream to infer a series of social and affective indexes.

Thanks to these functionalities, the envisioned system can be enforced with a number of classifiers developed to infer the affective and social state of the User in real-time. For instance, attention and distraction are a crucial point for the motivational strategy that can be put into action by the system. One way to discern these states is by using a gaze-based approach as presented by Prajod et al., 2023. User motivation and willingness to continue the task can also be affected by experiences of stress. In this sense, based on the observations presented by Prajod and André, 2022, hand-crafted HRV features can be leveraged to detect stress and react accordingly. Moreover, further information can be extracted from the User's facial expressions, detected using MediaPipe's Blaze face detection model (Bazarevsky et al., 2019) and classified into seven discrete emotion classes (see Figure 3.6): Neutral, Happy, Sad, Surprise, Fear, Disgust, and Anger, thanks to the AffectNet dataset (Mollahosseini, Hassani, and Mahoor, 2019). Additionally, the trained model can provide two continuous values (in the range [-1, 1]) for each image: Valence, indicating if the experienced emotion is positive or negative, and Arousal, representative of the level of intensity of the experienced emotion. Facial expressions can also be exploited to detect situations of physical pain. In fact, using deep learning (Wang et al., 2018; Hassan et al., 2019; Xiang et al., 2022) and transfer learning (Prajod et al., 2021) techniques, it is possible to link face-based pain detection triggers with the information provided by the "Physical energy" monitoring to cross-validate the estimations and obtain a more comprehensive and robust description of the User's overall state.

Of course, the rich information that can be inferred from social and affective signals, needs to find a way to enter the envisioned human-driven control loop. As the SSI framework allows for the implementation of Python-based custom plugins, a solution is provided by the use of the *rospy* (Ken Conley and Perron, 2012) and *rosbridge* (Mace, 2012) libraries allowing a direct translation between SSI pipelines and ROS topics. More details on this integration can be found in Chapter 4.

Psychological state

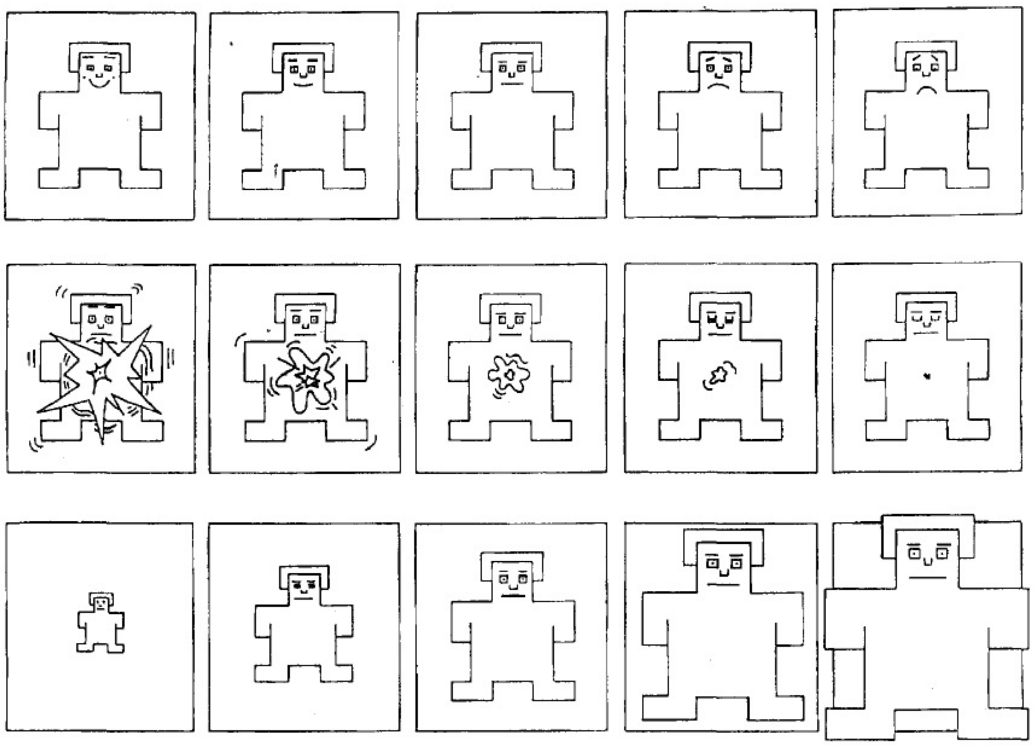


FIGURE 3.7: The SAM questionnaire used to collect information about the User's experience in terms of Valence, Arousal and Dominance.

In order to assess the experience associated with human-robot interactive activities, participants can be administered the Experience Sampling Method (ESM), a procedure developed to study behavior and the associated experience during their unfolding in real life, thus avoiding memory distortions (Csikszentmihalyi, Larson, and Prescott, 2014; Hektner, Schmidt, and Csikszentmihalyi, 2007). For this purpose, participants would be given a tablet reproducing an acoustic signal whenever a questionnaire needs to be completed.

Open-ended questions can be used to collect descriptions of the ongoing activity and related stake, location and social context. Moreover, a set of scales can be leveraged to assess the individual quality of experience associated to the ongoing task, by rating the level of cognitive, affective and motivational dimensions, including perceived activity-related challenges and personal skills in facing them. Some examples are:

- The Internal Control Index questionnaire (ICI) (Duttweiler, 1984), aiming to understand the internal locus of control (i.e., do you feel in control of the events or do you feel controlled by them?) of the participant.
- The Experiential Locus of Control questionnaire (ELoC) Jang et al., 2016, derived from the ICI to analyze the dimension of "feeling in control" not only as a personal attitude but also as a reaction to a specific activity.
- The Negative Attitude Towards Robots scale (Nomura et al., 2006b), exploring the feeling towards the practical, social and emotional interaction with robots.
- The Self-Assessment Manikin (SAM) Bradley and Lang, 1994, designed to infer measures of Valence, Arousal and Dominance through a non-verbal pictorial representation (see Figure 3.7).

One important aspect in the case of psychological reports is that they are collected asynchronously during the task and, therefore, cannot be a real-time input to the generalized human-driven control architecture. However, these measures still have an exceptional importance for the system as they allow the Supervisor to validate the other quantitative inferred indexes and to perform post-session analysis aimed at evaluating the effectiveness of the envisioned system in improving the experience of human-robot interaction.

3.3.2 User Model

The high-level information produced by the Interpreter is a good starting point to understand the overall current status of the User. However, the obtained indexes still need elaboration in order to fuse them in a single and heterogeneous representation of the actual experience of interaction: they are like pieces of a puzzle that need to be put together in order to reveal the full picture. This is without a doubt the most challenging aspect of the issue that the present project tries to tackle (even within human-human interactions we often find it difficult to understand each other) and the reason why a dedicated User Model is foreseen in the generalized architecture.

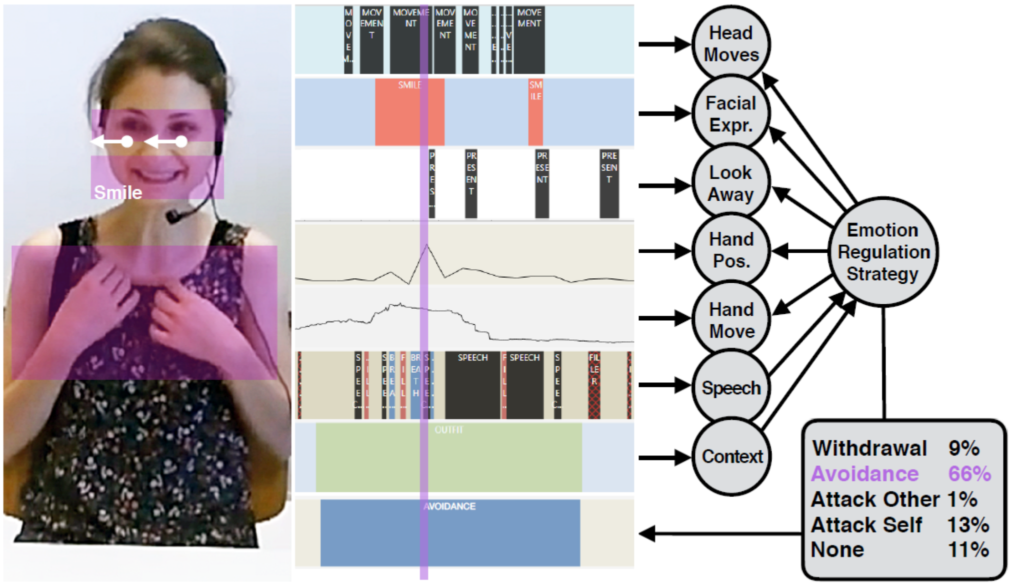


FIGURE 3.8: An example of the MARSSI model merging multiple social signals to estimate the confidence related to each modeled appraisal and regulation (Gebhard et al., 2018)

The first problem is that social signals classification alone is insufficient for understanding the real meaning behind emotional expressions. For instance, many communicative, emotional expressions are not directly related to internal emotional states: a smile could represent actual satisfaction, but also a mechanism used to hide shame and insecurity. One way to overcome this issue would be to enhance the social and affective high-level indexes produced by the Interpreter through a dedicated theory-based model capable of processing internal appraisal (Lewis, 2008) and regulation strategies (Gross, 2013). This is exactly one of the goals of the User Model block envisioned for the proposed human-centered control logic. A possible starting point for the implementation of this delicate module is the Model of Appraisal, Regulation, and Social Signal Interpretation (MARSSI) introduced by Gebhard et al., 2018. The proposed approach relies on Dynamic Bayesian Networks (DBNs) to fuse multiple social signals, as represented in Figure 3.8. Since DBNs support temporal representation, sequences for the interpretation of social signals can be learned. The User Model can therefore employ this DBN concept for real-time computation of a confidence value of possible modeled user affect, updating the possibilities of each modeled appraisal and regulation information. Another benefit is that the MARSSI model can be extended by different regulation strategies, useful for instance to deploy specialized versions of the model specifically designed for people characterized by ASD (Mazefsky et al., 2013).

Secondly, the present project proposes to complete the emotion-based information contained within social signals with indexes related to the mental and physical state/behavior of the User (see Section 3.3.1). The User Model should therefore make use of all these additional information to provide smarter pro-active responses of the system. To do so, rules derived from theory and from experimental behavioral analyses should be defined and integrated into the model. For instance, the detection of a state of physical fatigue obtained through biomechanical tracking could be merged with a possible expression of pain inferred by facial behavior analysis to trigger the correct rule-based reaction of the system. This reaction also depends on the specific field of application. In an industrial scenario, the production rhythm could be slowed down or a break could be suggested by the virtual character (Gatzounis et al., 2017). In a rehabilitation context, instead, the level of robotic assistance could be raised or the avatar could suggest a more ergonomic position to render the exercise execution more comfortable (Das and Mukhopadhyay, 2014). An additional example of the benefits of this rule-based approach can be found in the analysis of gaze behavior together with the emotional expressions mentioned above. Research shows that the direction of emotional expressions (i.e., understanding to whom or what that emotional information applies) is a crucial information to really decipher the User's intention (Hess and Fischer, 2013). In dyadic interactions, emotional expressions can be directed to the interaction partner, the situation, the dialog topic or at the person mentioned in the utterance. By linking the gaze or head movement while observing an emotional expression, its direction can be tracked (Bänninger-Huber and Steiner, 1992). The knowledge about an expression's direction could then be used for the automatic deduction of possible elicitors and the consequent selection of a reasonable system reaction.

Some of these rules and typical behaviors have been identified through the review of the available literature in Chapter 2, but a lot of aspects still need further evaluation. For this reason, Chapter 5 will be dedicated to a series of experimental campaigns aiming both at the identification of said rules and at the validation of the several components of the proposed architecture to obtain concrete answers in terms of feasibility and effectiveness.

3.3.3 Orchestrator

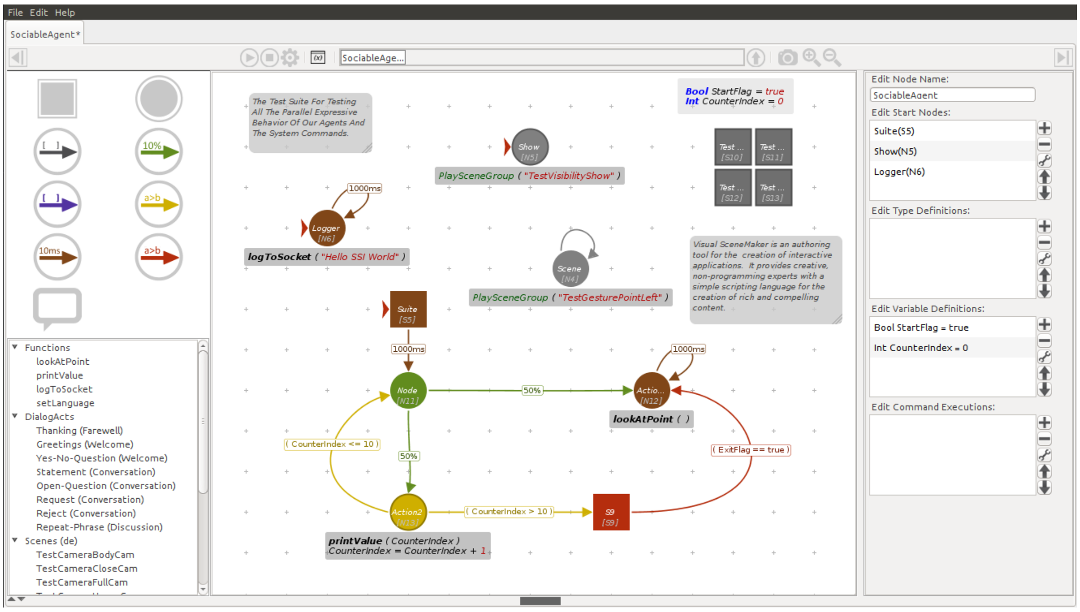


FIGURE 3.9: A screen capture of the VisualScene Maker Workspace where the high-level task definition and Robot-Avatar coordination can take place.

The Orchestrator is envisioned as a software framework for authoring, orchestrating, and executing scenario content with task specifications. In particular, relying on the content of the User Model, it is responsible for tailoring the actions of both the Robot and the the Avatar coherently and consistently in order to obtain a resulting behavior of the system adapted to the User's interaction experience. This component can be effectively implemented using Visual SceneMaker (VSM) (Gebhard, Mehlmann, and Kipp, 2012), which comes with an authoring tool for creating interactive presentations aimed at non-programming experts. It supports modeling verbal and non-verbal behavior of interactive agents and robots through a graphical interface and a simple scripting language that allows domain experts to create rich and compelling content. VSM's central authoring paradigm is the separation of content (e.g., the action to be performed by the Robot or by the Avatar) and logic (e.g., the system's reaction to user input). The content is organized as a collection of scenes which are specified in a multi-modal script with dialogue utterances and stage directions for controlling the Robot movements/parameters or the Avatar gestures, postures, and facial expressions. The logic of the interactive performance, instead is controlled by a scene flow as represented in Figure 3.9, implemented as a nested graph similar to Harel's statecharts (Harel, 1987). In this context the Supervisor, represented by an expert in the field of application but without the need for advanced programming knowledge, has the duty of setting up the task by using the pre-implemented building blocks. Of course, it is possible to prepare, save and play a number of scenarios depending on the specific task needs. Moreover, a list of manually tunable parameters is made available inside the VSM program so that the Supervisor can intervene for adaptation purposes without the need to stop and relaunch the on-going execution.

VSM is open-source and implemented in Java (SceneMaker, 2012). To achieve real-time communication with the rest of the architecture, VSM can be extended by a dedicated plugin pushing commands through ROS communication protocols. This can be done by simply building a bridge thanks to the functionalities of the *rosjava* (Damon Kohler, 2019) library (more details about this integration in Chapter 4). With this approach, the information provided by the User Model is directly available to the Orchestrator for task adaptation purposes. Similarly, the Orchestrator can dispatch its logic-dependent commands to the Robot-Avatar system and monitor their execution to correctly synchronize subsequent/parallel tasks or trigger necessary mitigation strategies.

3.3.4 Robot-Avatar

One of the main challenges for the envisioned system lays in the fact that hardware tools are not fully known before the solution is deployed for a specific application. As previously mentioned, this aspect is one of the main drivers for the choice of ROS, as it allows the implementation of a baseline software structure that can be interfaced with different devices afterwards by simply developing ad-hoc plugins. In particular, for the integration and control of the Robot, the *ros_control* (Chitta et al., 2017) packages allow to make controllers generic to any robot. With reference to Figure 3.10, a proper RobotHW implementation can render the developed control architecture suitable for a generic application. Moreover, relevant controller parameters can be modified online using the *dynamic_reconfigure* (Gassend, 2014) package, an extremely important feature as the ability to proactively tune the system to the needs of the User is one of the central goals for the present project. Finally, *moveit* (Chitta, 2016) can be leveraged to enforce the generalized system with powerful planning and replanning capabilities.

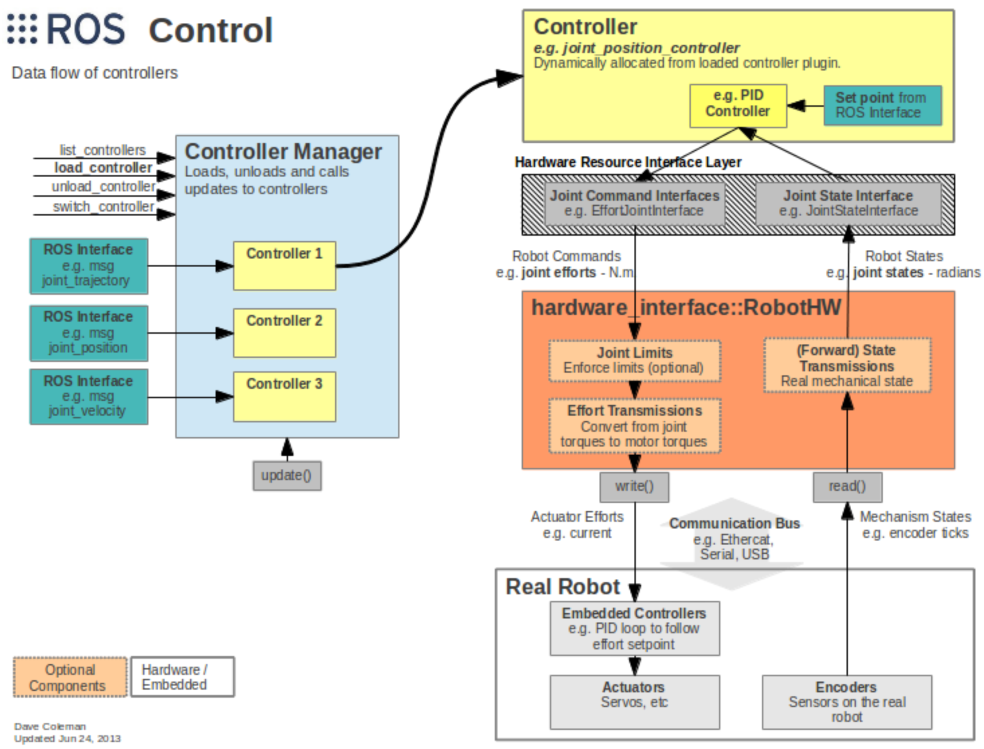


FIGURE 3.10: A diagram of the *ros_control* logic.

Leveraging all the tools mentioned above, the goal is to develop a low-level robot control module providing a number of ROS-based topics and services to the upper-levels. Topics consist in a continuous stream of formatted messages and can be used to share with the whole architecture the state of the robot movements and on-going actions (e.g., end-effector status). Services, instead, can be called by an external client to command the execution of a specific task (e.g., move along a certain trajectory, retune certain parameters).

As mentioned, this project also foresees the presence of an Avatar, seamlessly integrated with the Robot as if it was a visual representation of the intelligence of the system. As a project choice, the Avatar should be characterized by an androgynous aspect with the aim of limiting physical attraction/repulsion on a gender basis. The first design, produced by the DFKI institute in collaboration with a graphical artist, is reported in Figure 3.11. The represented design then needs to be converted into a digital 3D model and evaluated through off-line rendering, before transferring it to the Unity3D platform (Haas, 2014) for real-time rendering. The Avatar visualizer is developed using the YALLAH (Yet Another Low-Level Agent Handler) framework (Nunnari and Heloir, 2019) that allows customization with the Blender

3D (Blender, 2024) editor and deployment as a stand-alone Unity application (see Figure 3.12). The resulting player can therefore support animations, in-place rotation (e.g., to change the orientation of the avatar mimicking the movements of the robot to enhance their perception as a single entity), both Italian and English speech generation (to allow for both national and international experimental subjects) together with the built-in support for command-line control of networking options.

FIGURE 3.11: The conceptual design of the virtual character.

The Orchestrator can control the Avatar's behavior model thanks to another set of ROS-based services. By calling these services, it is possible to command the Avatar to perform several actions such direct the gaze, perform gestures, pointing at entities in space, and talk, depending on the needs of the specific application and situation. Role-wise, the Avatar acts as a mediator between the human and the Robot, filling the need for a visual counterpart to make interaction with robots more emphatic and acceptable. Considering interactive sessions with the User, behaviors should be modeled to provide a more natural and social interaction experience promoting motivation, easing self-regulation and helping the User to cope with emotions when abnormal stress or fatigue levels are detected.

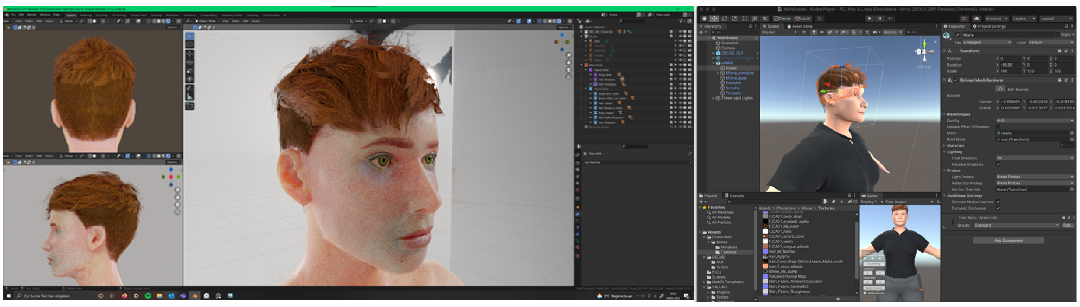


FIGURE 3.12: On the left the Avatar designed converted to a 3D model. On the right, the 3D model is imported into Unity3D and enriched with animation capabilities.

Chapter 4

Setting up the use-case scenarios

As anticipated in Chapter 1, the generalized human-driven control architecture presented in Chapter 3 is put into practice through two use-case scenarios that are of particular interest for the present project, as they require close interaction between the robotic device and the human user. Therefore, a description of the setups realized for MindBot, a mental-health friendly collaborative manufacturing workcell, and for the Empathetic Neurorehabilitation Trainer are reported in this Chapter.

4.1 MindBot: Mental-health friendly collaborative manufacturing

The constantly growing concept of Industry 4.0 is leading to completely new workspaces where automation machines cooperate with humans. However, the quality of experience and level of engagement of workers interacting with robots have become an active research topic only recently and still represent a largely unexplored domain. So far, industrial cobots have primarily been studied and designed addressing aspects related to the physical safety of the worker, aiming to optimize productivity performance by reducing uncertainty and instability in their cooperation with humans. While these topics still remain of great interest, new research branches must arise in order to explore the role that cobots could have in reducing the workers' psychological strain. The challenge lies in the fact that, differently from applications such as social robotics, the interaction between a human worker and a robot collaborator in an industrial scenario is bound to the specific task and production requirements. However, cobots have evolved to a point where many operations could be performed both by the manipulator and the worker, meaning that a certain degree of freedom in the assignment of subtasks between the two collaborators is possible. Moreover, a series of parameters characterizing human-robot collaboration, also identified within the previous works presented in Chapter 2, can be tailored with the aim of optimizing the worker's experience. It is clear that, in order to achieve such a goal, a multidisciplinary approach and a wide partnership contributing with several different fields of expertise are of utmost importance. In this regard, the MindBot project (Lavit Nicora et al., 2021), funded by Horizon2020, was launched with the aim of defining organizational and technical guidelines for the design of a "mental-health-friendly" cobot-based manufacturing workplace. Refer to Appendix A for the complete list of the partners making up the consortium and their role in the project.

4.1.1 Realization of the experimental workcell

The first step towards the goals of the MindBot project is the ideation and realization of a proper experimental setup. Taking inspiration from the most common industrial applications of collaborative robots in small and medium enterprises, the idea is to reproduce a realistic workcell in a laboratory environment where a human operator and a cobot can collaborate for the assembly of a certain product.

With reference to Figures 4.1 and 4.2, two tables placed in an L-shaped formation were positioned and constrained to each other thanks to an aluminum structure designed to guarantee that the several reference systems playing a role in the setup do not get mistakenly disaligned during the experimental activities, which would require time consuming recalibrations. Using the same structure, a Fanuc CRX10iA/l collaborative robot (Fanuc, 2021) was added to the workcell and fastened in a position allowing the cobot to reach most of the space available on the two tables. As represented, this configuration naturally identifies one area where the user can work on his/her independent tasks, another area assigned to the cobot and a third one, around where the two tables intersect, that can be used for collaborative activities. The cobot is equipped with a teach pendant tablet (mostly used for programming purposes and not during the experimental activities) which is typically kept somewhere easily reachable by the user in case the emergency button shutting down the system is needed. Moreover, a Robotiq Hand-e parallel gripper (Robotiq, 2021) is mounted on the end-effector interface of the cobot to enable pick and place operations.

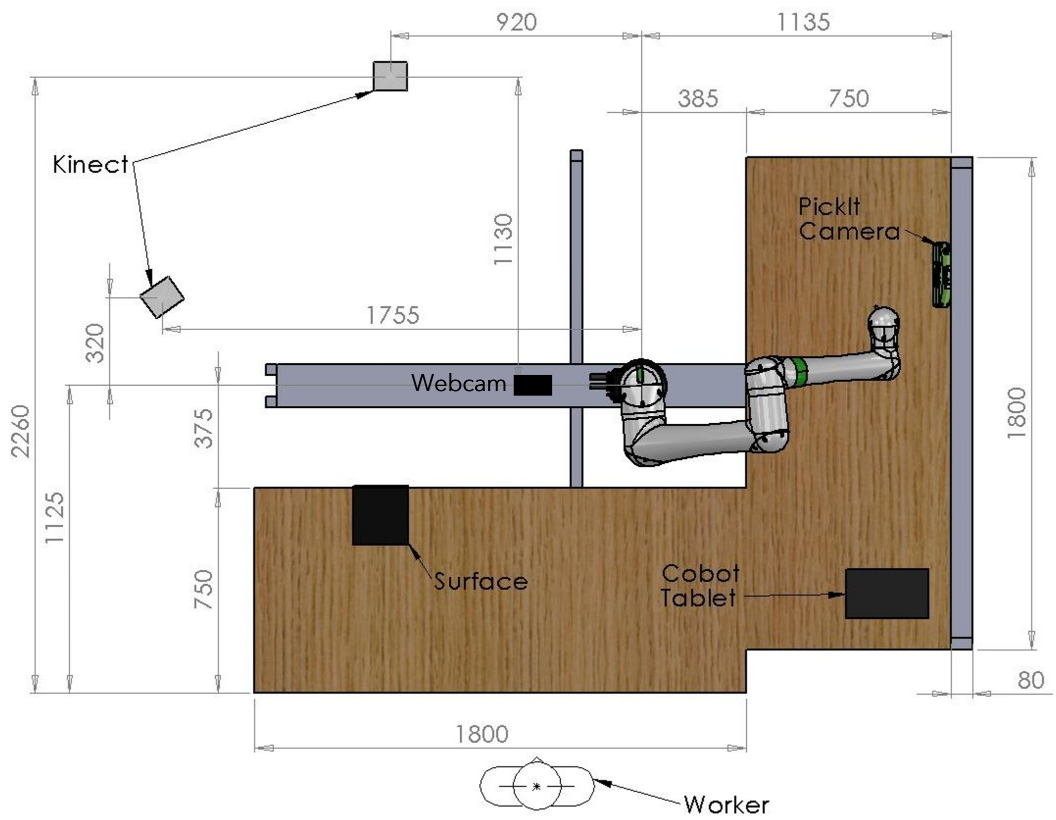


FIGURE 4.1: A top view schematic representation of the MindBot system.

Moving to the array of installed sensors, the collaborative workcell features a Rethink Pickit3D camera (Rethink, 2021), used for the detection of the position and orientation of components. This camera can be mounted on a linear guide, as in Figures 4.1 and 4.2, or mounted on the wrist of the cobot together with the gripper. The first solution decouples the camera movement from the cobot movement making it possible to parallelize the steps of detection and picking with a consequent reduction of waiting times for the execution of cobot tasks. However, such a configuration requires an extremely precise calibration to determine in real-time the transformation matrix connecting the camera reference system with the cobot one. On the other hand, mounting the camera at the wrist of the robot simplifies this aspect and ensures better precision in the robot movements at the cost of a slower execution. Together with the mentioned detection camera, a Logitech C920 HD Pro webcam (Logitech, 2021) is positioned on the aluminum structure right in front of the worker in order to record his/her behavior (e.g., facial expressions, hand gestues, gaze). Additionally, a set of two redundant Kinect Azure depth cameras (Microsoft, 2021b) are positioned at different angles in order to track

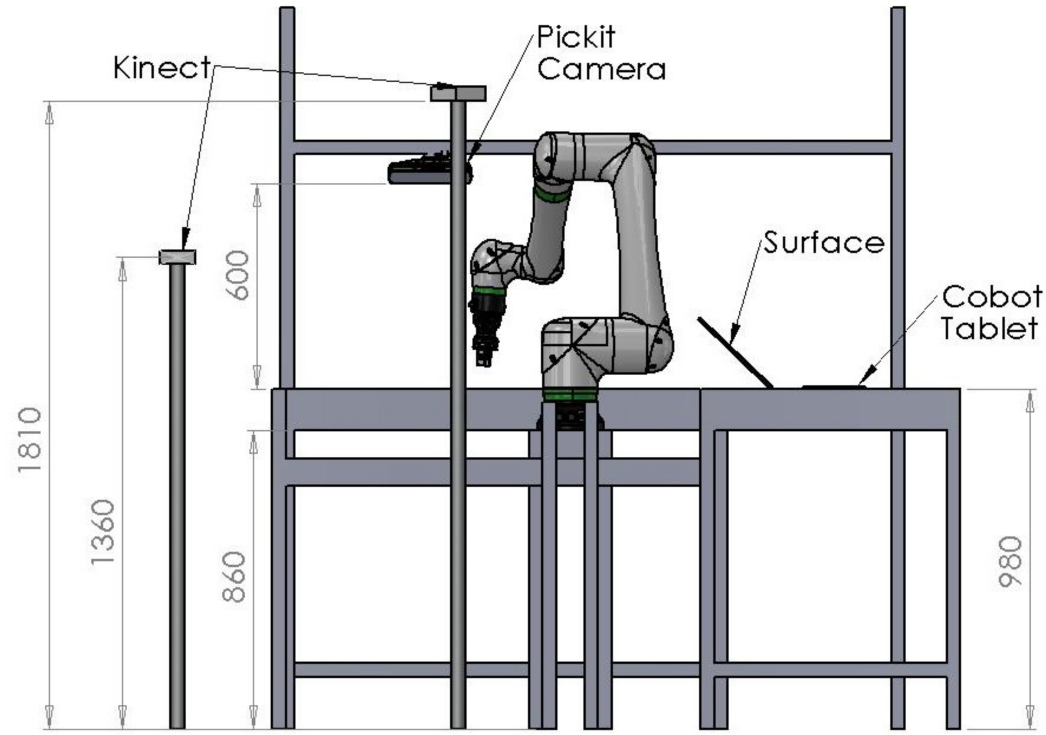


FIGURE 4.2: A side view schematic representation of the Mind-Bot system.

the user's skeleton with high confidence and robustness to occlusions. In terms of wearable sensors, it is very important to be less invasive as possible in order to minimize the discomfort of the subject and speed up setup operations. Depending on the specific experimental needs, a FitBit activity tracker (Google, 2021) and/or a Polar H10 chestband (Polar, 2021) could be leveraged to monitor additional physiological parameters of the user.

Finally, a Microsoft Surface PC (Microsoft, 2021c) is placed on the user's table for different purposes. First, it can be used as a touch screen monitor to administer questionnaires during the experimental sessions and directly collect experience samples from the users. Moreover, a virtual character can be displayed on its monitor and communicate with the user through verbal and non-verbal behaviors.

4.1.2 Design and production of the assembly components

Now that the design and realization of the enhanced collaborative manufacturing workcell has been presented, it is necessary to identify a suitable task to be carried out during the experimental activities. After analyzing different

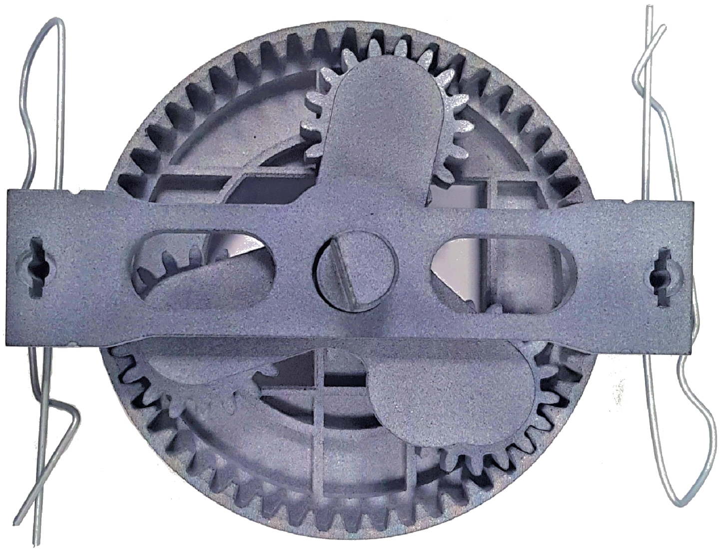


FIGURE 4.3: The custom epicyclic gear train to be used as collaborative assembly task during the experimental activities.

types of applications (e.g., painting, glueing, handling), it was agreed within the project consortium to select a collaborative assembly task. One of the main drivers for this choice was the possibility of implementing close interaction between the cobot and the user and the freedom to, at least partially, modify the order of operations and to implement different levels of collaboration, if required.

For this purpose, the custom epicyclic gear train represented in Figure 4.3 was designed (Redaelli, Storm, and Fioretta, 2021). The resulting product is characterized by large tolerances and the absence of strong or irreversible couplings between the components, in order to satisfy the following project requirements:

- The process required for the assembly of the components should be fairly easy in order to minimize both the time required for the training of the user and the chances of failure for the cobot;
- The finished product should be easily disassembled in order to minimize the time required for the setup of the workcell before each experimental session;
- The product should promote collaboration by design, for instance by requiring more than two hands to avoid undesired disassembly of the components.

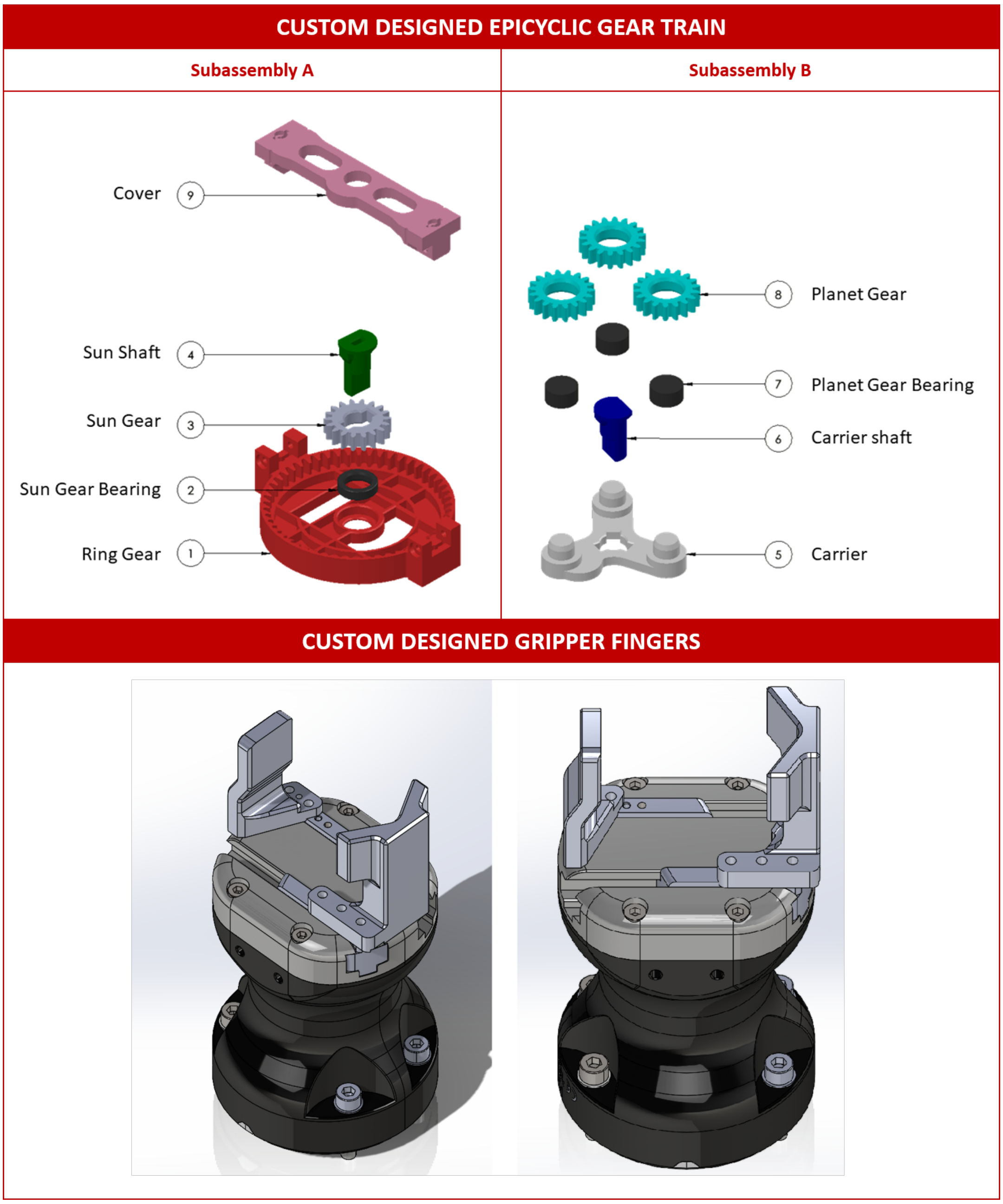


FIGURE 4.4: At the top, an exploded view of all the components making up the custom epicyclic gear train (on the left the parts assigned to Sub_A, on the right the ones assigned to Sub_B). At the bottom, the custom gripper fingers designed to allow the self-adjustment of components during pick and place operations.

At the top of Figure 4.4 all the components making up the complete assembled gear train are shown. Each component has been designed both to facilitate recognition by the Pickit3D detection camera in terms of position and orientation and to be easily handled by the cobot gripper. For this purpose, most of the components include a particular shape feature, made of one flat face and one round face that, together with the custom design of the gripper fingers represented at the bottom of Figure 4.4, allows for a certain degree of self-centering and self-orienting of the parts during pick and place operations.

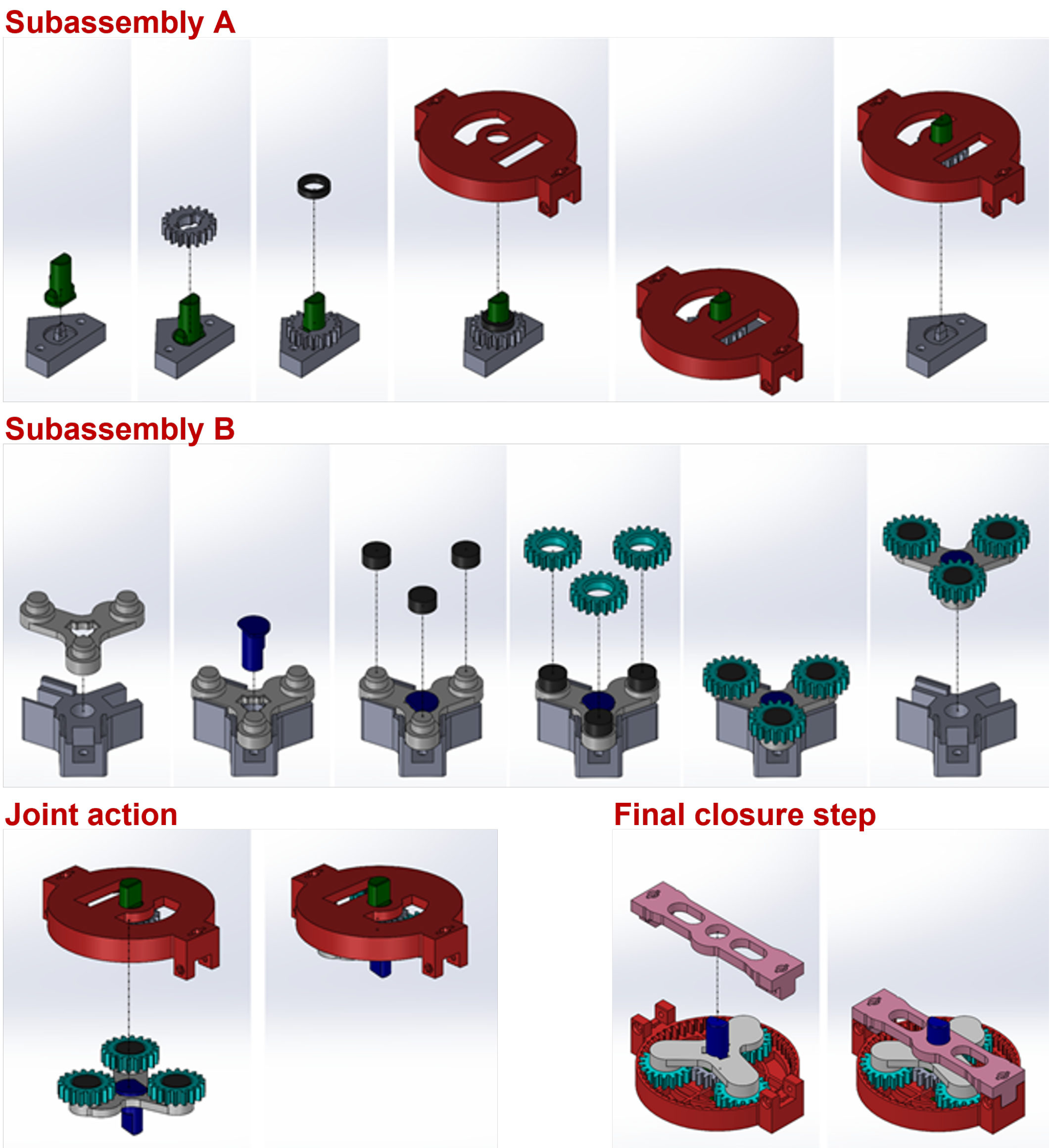


FIGURE 4.5: An exploded view explaining the assembly procedure for the various components. At the top, the steps required for Sub_A. At the center, the steps required for Sub_B. At the bottom, the final joint action (bottom-left) and closure (bottom-right) steps.

As shown, the components can be divided into two groups, called subassembly A (Sub_A) and subassembly B (Sub_B) respectively. The idea is to be able to freely assign the assembly of Sub_A and Sub_B between the user and the cobot depending on the specific experimental needs. In fact, a typical assembly cycle foresees a first step where the two entities independently work on their assigned subassembly, followed by a second step where close collaboration is required to join the two subassemblies and complete the product. Figure 4.5 shows the assembly procedure for Sub_A and Sub_B, the collaborative joining process and the final locking of the gear train. As shown, supporting jigs have been designed to ease the work on the two subassemblies. The collaborative joining phase is of particular interest for the project. One possible implementation of this phase foresees the cobot picking up Sub_A and positioning it near the user to allow the insertion of Sub_B. The pose of the robot should be comfortable for the user to complete the assembly, allowing the operation to take place even if mechanical parts are not tightly coupled. The user has to carefully position Sub_B, inserting mechanical components and paying attention to the correct meshing of the gears, as in Figure 4.6. This requires particular care by the user in order to avoid the planetary gears and their bearings to fall down, for instance using both hands to keep them in position and then aligning the gear teeth. Of particular relevance is the need for a "robotic third hand", designed to promote close collaboration by design.

FIGURE 4.6: A detail of the joint action where the user focuses on the correct meshing of the gears while the cobot helps keeping the parts in place as a third hand.

Once the joint action is completed, the product has to be released from the cobot gripper and handled by the user for the final operations. This requires a trigger, activated by the user through a pedal switch, giving consent for the next cobot assembly cycle. At this point, the user holds the entire assembly, adds the final cover fixed with a couple of cotter pins and places the completed product in a dedicated bin or container.

4.1.3 Implementation of the software interfaces

With the hardware components selected and installed in the laboratory setting, it is necessary to interface them between each other through a comprehensive software framework. For this purpose, the generalized human-driven architecture presented in Chapter 3 is leveraged to deploy a realistic workcell enabling constant monitoring of the worker's psychological strain and adaption of the behavior of the production cell.

First of all, all the individual computers controlling the various hardware components need to be connected to a local network allowing the necessary flow of information, as in Figure 4.7. As anticipated, the architecture heavily relies on ROS to facilitate modular development, even remotely among partners' laboratories by sharing dedicated *rosbag* files (Vigni, Andriella, and Rossi, 2024). In the chosen configuration, the main Orchestrator PC acts as *ros master* while all the other components are implemented as *ros nodes*, exchanging information and commands within the local network. Refer to Appendix B for a complete list of all the implemented topics and services.

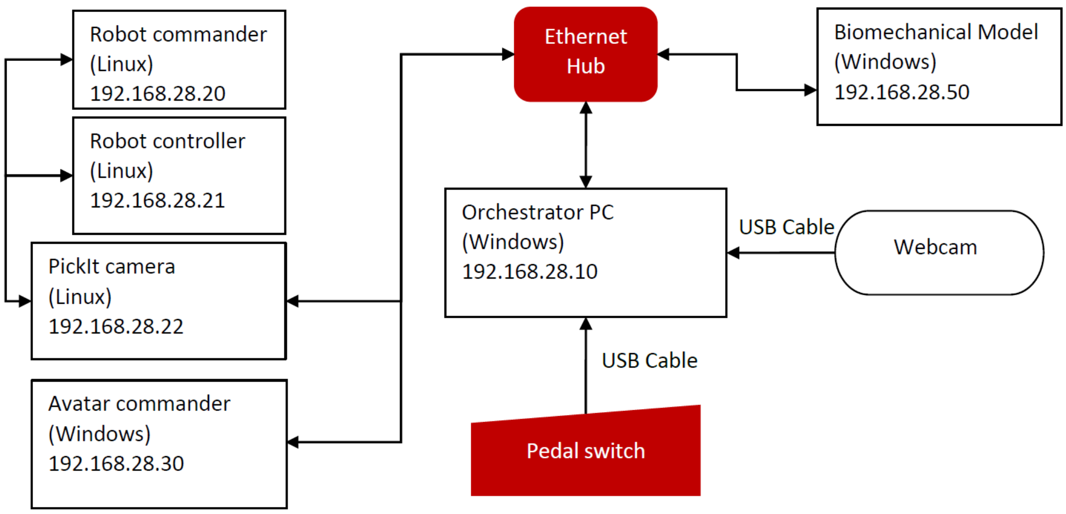


FIGURE 4.7: A schematic representation of MindBot local network.

Robot control

Starting from the robot control system, several software components written in C++ and deeply embedded within the ROS control framework have been implemented. They have been organized in stacks to achieve a clearer and modular structure, easy to maintain and integrate:

- **Cobot stack**: This stack contains all the packages required to interface the cobot with the ROS framework, therefore decoupling the whole MindBot architecture from the specific robot model installed in the production cell. Since the interface packages for the Fanuc CRX10iA/l cobot were not available at the time of integration, they had to be developed from scratch by the author specifically for the MindBot project. Communication between the chosen cobot and ROS has been established using the Fanuc User Socket Messaging (USM) option, allowing the exchange of data packets using TCP/IP communication between the computer and the robot controller, in combination with the Fanuc Remote Motion Interface option, enabling the exchange of semi-formed control commands between the computer and the robot controller. Exploiting these tools, the cobot state is constantly published inside the ROS network, while motion commands can be sent to the robot controller for micro-interpolated execution.
- **Tool stack**: Similarly to the previous one, this stack contains all the packages needed to control the specific tool mounted on the robot wrist remotely, within the ROS framework. Depending on the chosen combination of tool and cobot, interface packages may be available and directly retrievable from the ROS community. Once again, since specific packages interfacing the Fanuc CRX10iA/l cobot and the Robotiq Hand-E parallel gripper were not available at the time of integration, they had to be developed from scratch specifically for the MindBot project. For this purpose, an additional USM TCP/IP pipeline has been created to enable remote control of the tool mounted on the robot wrist. Karel scripting has been leveraged to let the robot controller push the received commands to the end-effector tool using its internal RS485 serial communication.
- **Vision stack**: This stack contains all the packages connected to the vision capabilities needed by the robot control system. Regarding the detection of the assembly components, the cell is equipped with an industrial vision system with ROS functionalities already offered by the producer (Rethink, 2022). Also, the camera system has been instructed through its proprietary web interface and contains all the CAD models of the components to be detected together with information related

to the pick strategy to be adopted for each one of them. User's detection is instead performed by an external module relying on the installed Kinect cameras and pushing simplified volumes for each body part through ROS topics. These are useful to provide the system with the knowledge of the user's position and occupied spaces and leveraged, for instance, inside the planning scene for collision avoidance algorithms.

- **Task stack**: This stack contains all the packages related to the knowledge of the on-going task. In particular, all the static obstacles known a-priori inside the workcell are defined here and published inside the robot virtual scene for collision avoidance purposes.
- **MindBot stack**: This last stack acts as a manager of all the packages introduced above. It is required to dispatch synchronized low-level commands to the hardware as a response to the direct communication established with the Orchestrator module.

Data acquisition and interpretation

As previously mentioned, data acquisition and interpretation is mainly carried out using the SSI tool which needs to be interfaced with the ROS framework in order to allow communication with all the modules making up the system. First of all, SSI only runs on Windows systems while the preferred OS for a ROS installation is Ubuntu. To solve this problem, the *rospy* and *rosbridge* libraries can be used. This combination allows the Windows machine to communicate with the Ubuntu machine through ROS without the need for duplicating the two software installations. Essentially, rosbridge can be used to open a websocket for exchanging JSON messages with any software component outside of the active ROS architecture. Incoming JSON messages are interpreted and republished within ROS in the form of topic, service or action, while any outgoing topic, service or action is translated back into a JSON message. Similarly, rospy allows to create a python script to communicate with ROS without the need for a ROS installation on the involved machine. The available websocket is used to create a connection and messages are sent and received in the JSON format.

Thanks to the tools presented above, the integration between ROS and SSI has been implemented in collaboration with the University of Augsburg by installing a rosbridge node on the Ubuntu machine equipped with the ROS installation and by running the python script leveraging rospy functionalities on the Windows machine equipped with the SSI installation. A *ros_sensor.py* script has been written to subscribe to a ROS topic and to push all the received information into an SSI stream. A second, *ros_consumer.py* script is instead dedicated to take all the data flowing through an SSI stream and publish it in the form of a ROS topic. With this solution, all the data produced by the sensors in the workcell can be received by SSI, interpreted to high-level indicators and published back to the ROS framework, ready to be received by the User Model and Orchestrator modules.

Task management and orchestration

As anticipated in Chapter 3, the Orchestrator module can be realized using the Visual SceneMaker tool. Thanks to its intuitive graphical approach, VSM can be used as an authoring tool allowing the implementation of interactive and complex tasks involving both robots and avatars without the need for any advanced programming experience. Even though this solution has been employed successfully by its creators in past related projects, it has never been deployed within a ROS environment meaning that a dedicated software layer is needed to create the necessary interfaces. For this purpose, VSM provides a plugin system based on the extension of a 'RunTimePlugin' Java class that can be used for the development of dedicated interfaces towards third-party products without involving the recompilation of the core of the system. Leveraging this functionality, a dedicated plugin has been developed in collaboration with DFKI using *rosjava*. This tool allows to create a java script to communicate with ROS without the need for a dedicated installation. The generated messages are sent through the network, translated by the already running rosbridge node and made available to the rest of the modules. Now, the Orchestrator is able to receive information related to the state of the User and send the corresponding rule-based commands to both the Robot and Avatar thanks to the exposed ROS services.

Figure 4.8 shows the top level of the VSM project implemented for commanding the robot, the equipped gripper and the detection system through the execution of the assembly task. As represented, the first node is used to set up a series of robot variables such as maximum speed, maximum acceleration, minimum distance from the operator and so on. Then, the system asks the user to fill in a form collecting some basic data and waits for a "Continue" signal through the pedal switch (see Figure 4.9) before moving to the "Task"

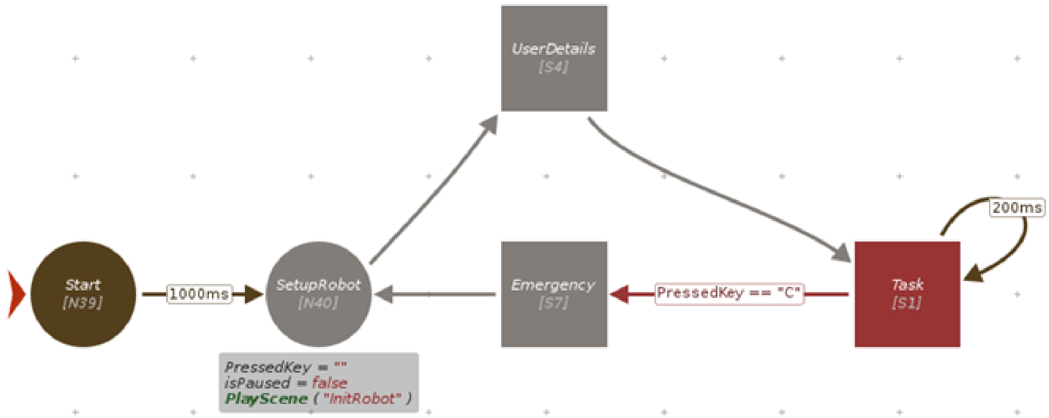


FIGURE 4.8: The top level of the VSM program orchestrating the MindBot assembly task.

supernode. If during the execution of the task the “Stop” signal is received, the execution state immediately jumps to the “Emergency” supernode, responsible for reacting to the request by bringing the robot to a safe state. If, instead, one cycle of the task is completed without problems, the system automatically restarts the production cycle after a brief delay of around 200ms.

FIGURE 4.9: The pedal switch used to send "Continue", "Pause" or "Stop" signals to the system.

At the top of Figure 4.10 the steps that make up the “Task” supernode are depicted. Each of these steps is a supernode by itself, containing all the instructions for the assembly of one of the components of Sub_A. An example is provided at the bottom of Figure 4.10 for the "SunShaft" component. First, the robot is sent over the predefined area where a buffer of the needed component is stored (e.g. “ROI1”). Then, the detection camera is used to look for a reachable component and to define its position with respect to the reference frame at the base of the robot (i.e., “Wait Detection”). Once the component is detected, a kinematic inversion procedure is performed to send the robot to a position just above the detected component (i.e., "Approach Part"). The robot then slowly moves down on the part, closes the gripper, raises the part and places it in the assembly area.

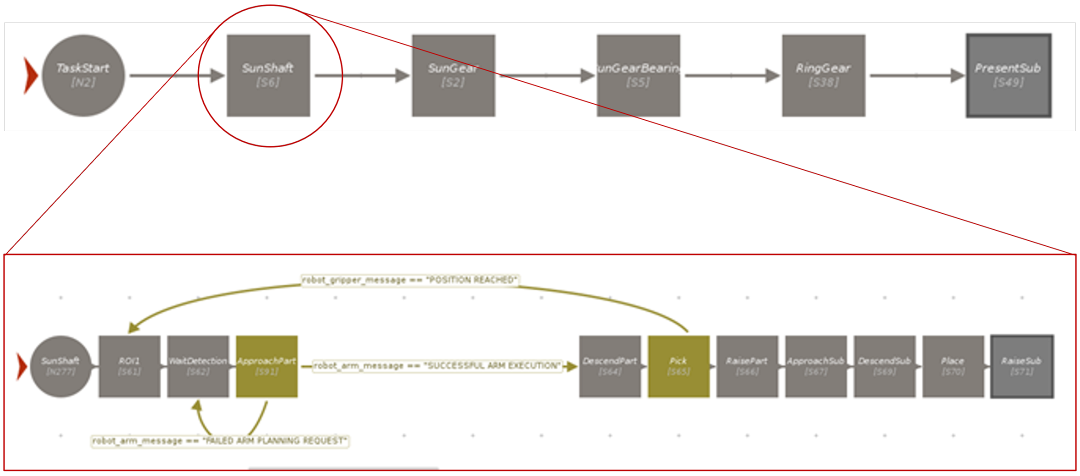


FIGURE 4.10: The lower levels of the VSM program orchestrating the MindBot assembly task.

VSM functionalities are leveraged not only during the actual collaborative assembly task, but also for collateral purposes. In fact, given the several reference systems that are needed to align all the components making up the experimental setup, a calibration procedure should be regularly performed. For this purpose, the robot can be equipped with a calibration table at its wrist in place of the parallel gripper, as represented in Figure 4.11.

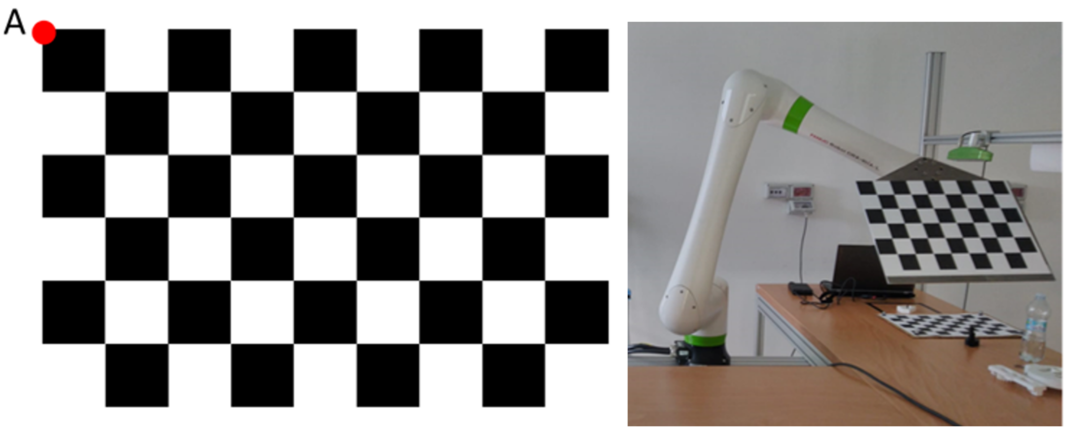


FIGURE 4.11: The cobot equipped with a chess-pattern table used to calibrate the transport matrices connecting all the reference systems in play for the system.

Following a predefined sequence made up of dedicated supernodes inside the VSM project, the robot moves the calibration table in front of the cameras stopping at difference distances and angles in order to promote robustness. Knowing the exact size and shape of the calibration table (square length of 45 mm), a dedicated algorithm is then responsible to analyze the images captured by the cameras and use them to reconstruct all the transformation matrices connecting the reference systems. With this approach, it is possible to express all measures taken by the different devices in use to a unique reference, positioned at the base of the cobot.

4.1.4 The resulting demonstrator setup

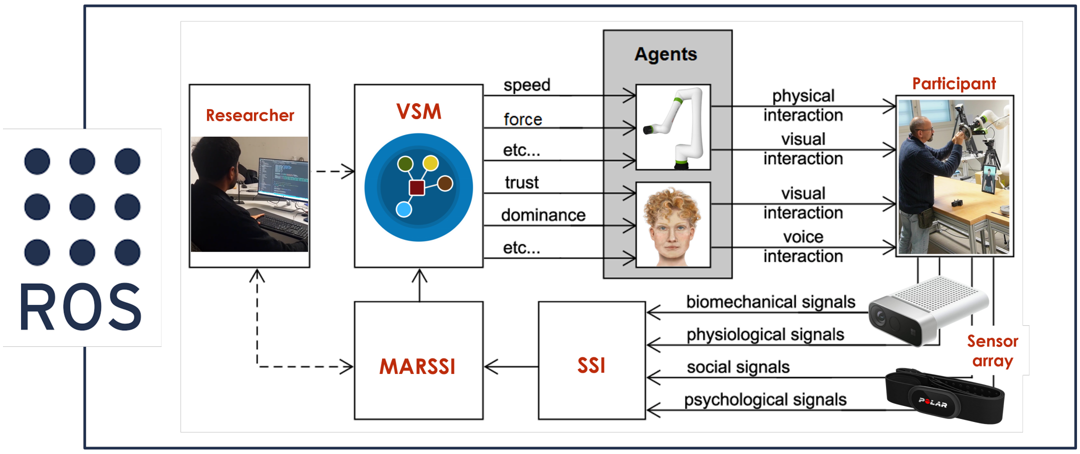


FIGURE 4.12: A schematic representation of the generalized human-driven control architecture deployed for the MindBot system.

The MindBot demonstrator is finally ready to be used for experimental purposes. Figure 4.12 provides a summary of how the human-driven control architecture has been deployed in all its hardware and software components. Moreover, the resulting setup assembled inside one of the laboratories of STIIMA-CNR Lecco is depicted in Figure 4.13. As shown, an operator works on the assembly with the help of a supporting jig, with the spare components of Sub_B on the side. Behind the tables, redundant Kinect cameras and webcams constantly monitor the user's actions and behaviors and inform the system about the inferred interaction experience. To the user's right, a cobot works on Sub_A using all the components placed in their predefined areas on the table. The detection camera is mounted to its wrist together with the chosen gripper, allowing the manipulator to detect the parts and to perform pick and place assembly operations, always with the help of a central supporting jig. In front of the operator, a tablet is held in position by a flexible support and displays the avatar providing a visual representation of the intelligence of the system. Its positioning is designed to be always visible to the user, while hidden speakers allow the user to hear the avatar's utterances even in noisy environments. Finally, to the user's left is the Orchestrator main PC, running the task and available to the researcher, acting as a Supervisor, for any occurring necessity.

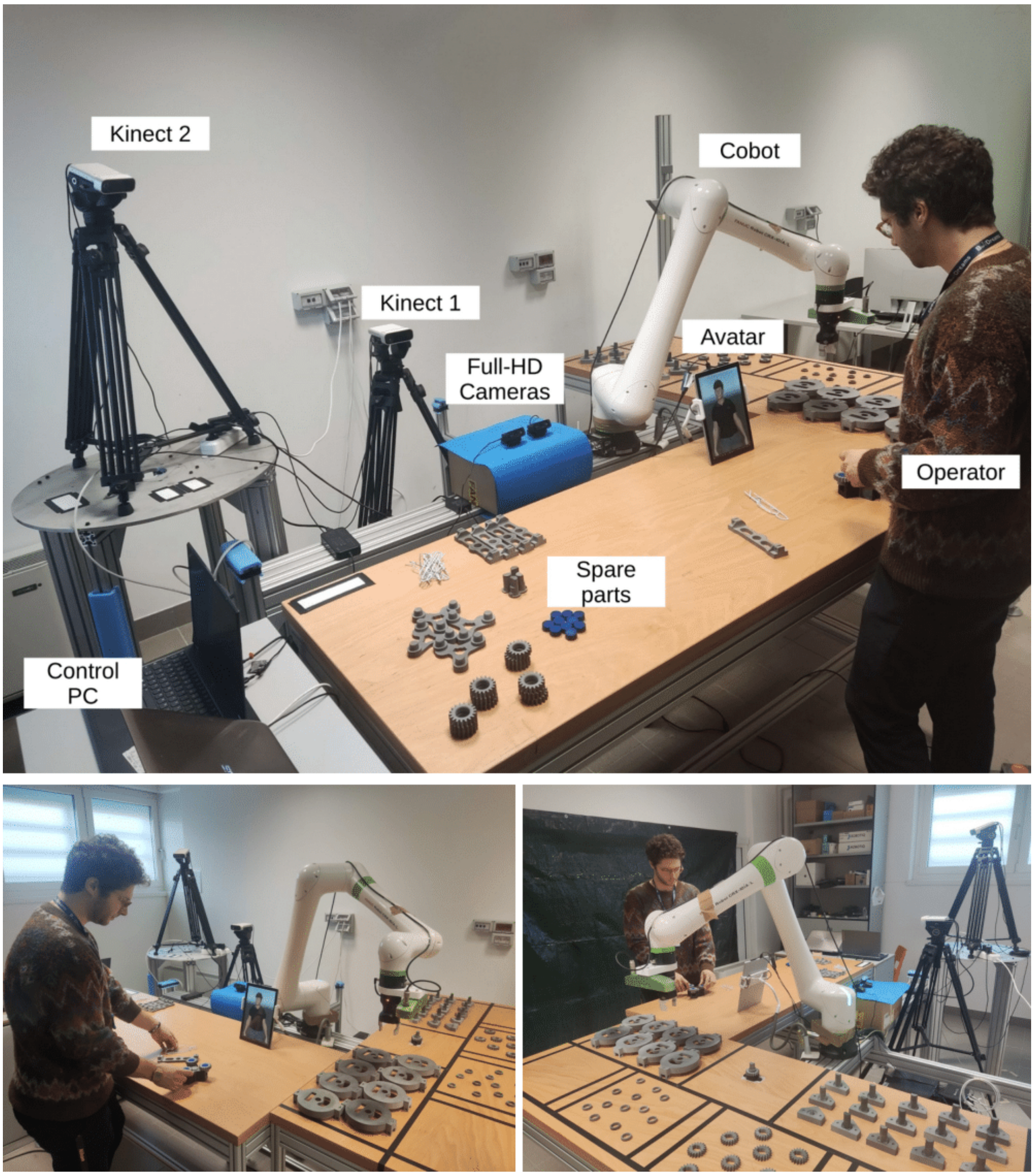


FIGURE 4.13: The resulting experimental workcell seen from multiple points of view.

4.2 Empathetic Neurorehabilitation Trainer

Neurorehabilitation is a widely used medical practice that aims to aid recovery from a nervous system injury. Its purpose is to maximize and maintain the patient's motor control while trying to restore motor functions in people with neurological impairments. Given the constant growth and aging of the world population, the number of patients affected by neuromotor disorders that seek the attention of professionals for their rehabilitation therapy is constantly increasing (Crocker et al., 2013). However, due to a lack of medical personnel, it is impossible to provide the intense training that would be needed for an effective recovery of the patient's capabilities, therefore hindering the actual outcomes of the treatment (Teasell et al., 2005). This situation is both harmful for the patients and constitutes a relevant burden on society and the healthcare system (Wynford-Thomas and Robertson, 2017).

To address this issue, robot-assisted training has been widely investigated as an effective neurorehabilitation approach that helps augment physical therapy and facilitates motor recovery. According to literature, such approaches can help therapists save time and energy while providing patients with a tool capable of assisting the execution of accurate and repetitive moments in high-intensity training sessions (Kwakkel, Kollen, and Krebs, 2008; Zhang, Yue, and Wang, 2017; Qassim and Wan Hasan, 2020). The current situation sees a limited number of this kind of devices, already installed in rehabilitation clinics, hindering their potential as they have to be scheduled over a large number of patients (Maciejasz et al., 2014; Stein, 2012). However, forecasts show that a relevant diffusion of this technology is taking place meaning that, in the near future, we will see an exponentially rising number of the installations of this technology (Morone et al., 2023). Moreover, most of the devices currently available are bulky and expensive but, thanks to the push for telemedicine and telerehabilitation, a new generation of rehabilitation robots is making its way into the market (Washabaugh et al., 2018; Molaei et al., 2022; Mayetin and Kucuk, 2022; Tseng et al., 2024). These affordable and portable solutions would allow for the capillary diffusion of the technology, out of the clinics and directly at home for the patients to use. The application of rehabilitation robots in domestic environments would represent a plausible solution to the lack of treatment intensity that patients are experiencing nowadays. In fact, a system capable of assisting the patient in performing the necessary repetitive motions would relieve a lot of the pressure that is acting on the clinical structures, since the physical presence of medical personnel would be required only for sporadic interventions.

However, a crucial issue for rehabilitation training is user engagement and motivation (Blank et al., 2014), which may be lacking if the rehabilitation system is used without a human medical coach. Since the effectiveness of the treatment has been proven to be related to the patient's level of engagement (Turner-Stokes et al., 2015), it is important for the envisioned system not only to be able to physically assist the patients but also to understand their affective state and react accordingly. Therefore, a neurorehabilitation training system capable of modeling the patient's state and tuning its behavior depending on both the measured performance and the overall inferred state could improve the user's engagement and, consequently, the outcome of the therapy. Once again, a multidisciplinary approach is required to push towards such a complex goal. For this reason, the Department of Affective Computing of the University of Augsburg (UA) and the German Research Center for Artificial Intelligence (DFKI), already involved in the MindBot project introduced above, collaborated on the realization of a so-called Empathetic Neurorehabilitation Trainer.

4.2.1 The PLANarm2 prototype

As a first step for the realization of the Empathetic Neurorehabilitation Trainer system demonstrator, a proper robotic device needs to be selected. For this purpose, instead of purchasing a commercial device, it was decided together with the project partners to use the PLANarm2 rehabilitation prototype (Yamine et al., 2020), designed from scratch by the STIIMA–CNR institute and reported in Figure 4.14. This choice was driven by the ease with which the device's control system can be accessed and interfaced to the other modules envisioned for the present project. Moreover, the prototype satisfies all the requirements of portability and affordability mentioned above.

In terms of mechanical design, the well-known 5R planar kinematic chain (Giberti, Cinquemani, and Ambrosetti, 2013) was considered a promising solution as it makes it possible to place both motors on a fixed base. As a result, the robot is characterized by a relatively high stiffness and lower moving masses if compared to serial manipulators, therefore providing higher dynamic performances, a lighter structure and, potentially, better positioning accuracy. Even though this architecture has already been adopted to realize similar devices, such as the one developed by Klein, Roach, and Burdet, 2014, some key improvements have been made. Starting from the parametric

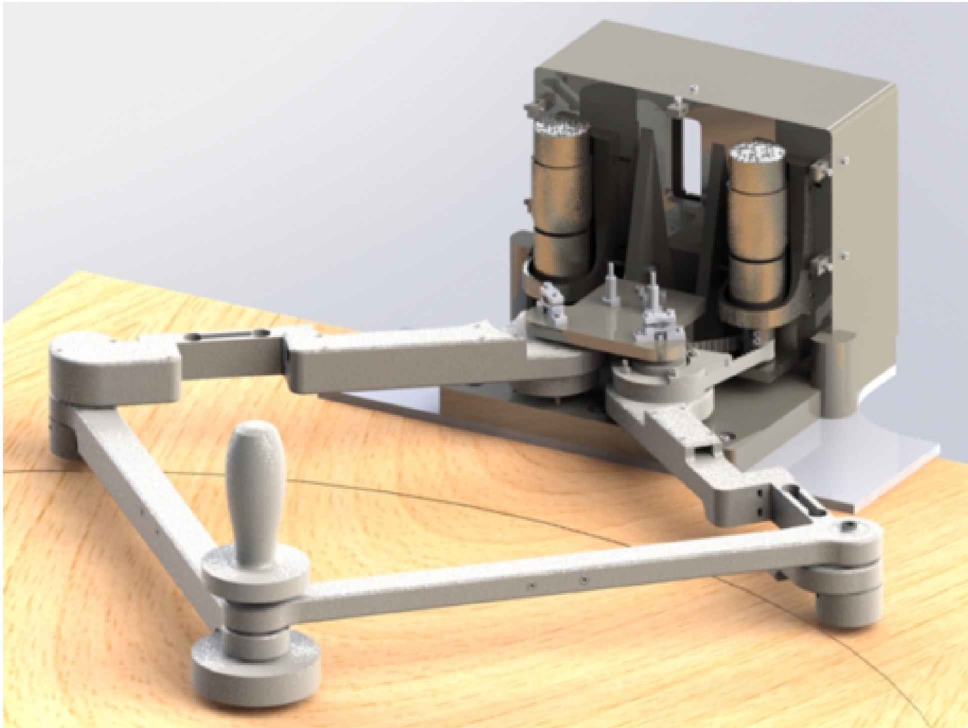
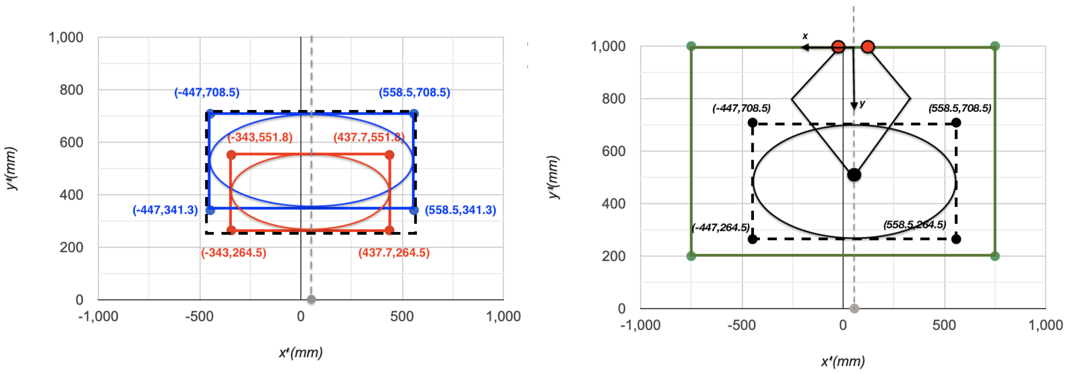


FIGURE 4.14: A rendering of the PLANarm2 rehabilitation prototype.

model of the 5R kinematics, the length of the links of PLANarm2 have been optimized to have good kinematic performances in the large majority of its workspace. Moreover, the workspace itself has been properly dimensioned to overlap the range of motion of the upper limb. The theoretical reachable workspace for upper limb neurorehabilitation in Cartesian coordinates was defined by Corona-Acosta and Castillo-Castaneda, 2015 through a transformation from articular to Cartesian coordinates, performed using the direct kinematics of the human arm. With reference to the left side of Figure 4.15, the desired workspace is defined as the union between the workspace defined for minimum limb lengths and the one defined for the maximum limb lengths. This identifies an ellipse with center $c = [0, 513.5] mm$, $minor\_axis = 222mm$ and $major\_axis = 502.75mm$. Since the population studied by Corona-Acosta and Castillo-Castaneda, 2015 was right-handed, the authors of that research centered the reachable workspace at $x = 55.75mm$. On the contrary, the y-axis of PLANarm2 has been translated of the same distance in order to have it aligned with the center of the reachable workspace, as shown in Figure 4.15. As a result, the obtained structure is inherently characterized by a symmetrically distributed kinetostatic behaviour with respect to the user's sagittal plane and therefore usable both by right-handed and left-handed patients. Since the manipulator is designed for domiciliar use, it should be possible to install it on a regular home table or desk. An average sized table is assumed to have a length of, at least, 1500mm and a width of about 800mm. Furthermore, the patient should be sitting in front of the device at a distance of around 200mm away from the table. As shown on the right side of Figure 4.15, the design of the device is perfectly fitted to the assumed dimensions and can therefore be easily clamped to a common table, facilitating both portability and fast installation inside already furnished environments. As opposed to the device described in Klein, Roach, and Burdet, 2014, which is characterized by a self-supported manipulandum, the PLANarm2 manipulandum slides on the surface of the table, automatically supporting vertical loads. Consequently, the links of the parallel structure only transmit horizontal forces, limiting bending loads and enabling production through additive manufacturing techniques, in line with the affordability requirement.



(A) (B)

FIGURE 4.15: (a) The total reachable workspace (dashed) including patients with minimum limb lengths (red) and those with maximum limb lengths (blue). (b) The placement of PLANarm2 with respect to the reachable workspace.

Moving to the electronic components of the prototype, they have been selected in accordance with the expected functionalities of the device. Impedance and admittance control strategies are today part of the state of the art in physical HRI and essential for rehabilitation devices. Impedance control requires a direct force/torque control Hogan, 1985 with a consequent preference for backdrivable motors. However, the high-torque and low-velocity features needed for this application, clash with the characteristics of electrical motors that generally express high velocity and low torque. Torque motors are available on the market, but they are expensive and not suitable for the low-cost device described. Instead, PLANarm2 is moved by two 24V DC motors equipped with a non-backdrivable 49:1 gearbox (resulting in a no-load speed of 143*rpm* and a stall torque of 19.6*Nm*) further reduced by a 3:1 pulley belt transmission connected to the link. Due to this design choice, an admittance control strategy has to be preferred over the impedance approach and, therefore, force sensing and good position/velocity control are required. For this purpose, the links have been designed to embed a Cantilever Beam load cell measuring the transmitted shear force. By multiplying this force by the arm length, it is possible to evaluate the motor torque, as necessary for the selected low-level control strategy. Moreover, each actuator is equipped with an incremental encoder sensor with a final resolution on the link rotation of 0.00213*rad*, providing enough precision for position and velocity measurement. Proximity sensors are used to detect the end stroke of each arm, as a reference for the incremental encoders. Finally, the low-level firmware, in line with the affordability nature of the device, is installed on an Arduino DUE board that, thanks to a VNH5019 dual motor driver, enables full control over the robot movements. The Arduino board, the motor drivers and other electronic elements required to operate, are installed on a PCB and mounted on the device. The MCU controller is, then, connected with an external PC, responsible for the management of the rehabilitation task, through serial communication. Further details about the design and construction of the PLANarm2 prototype are provided by Yamine et al., 2020.

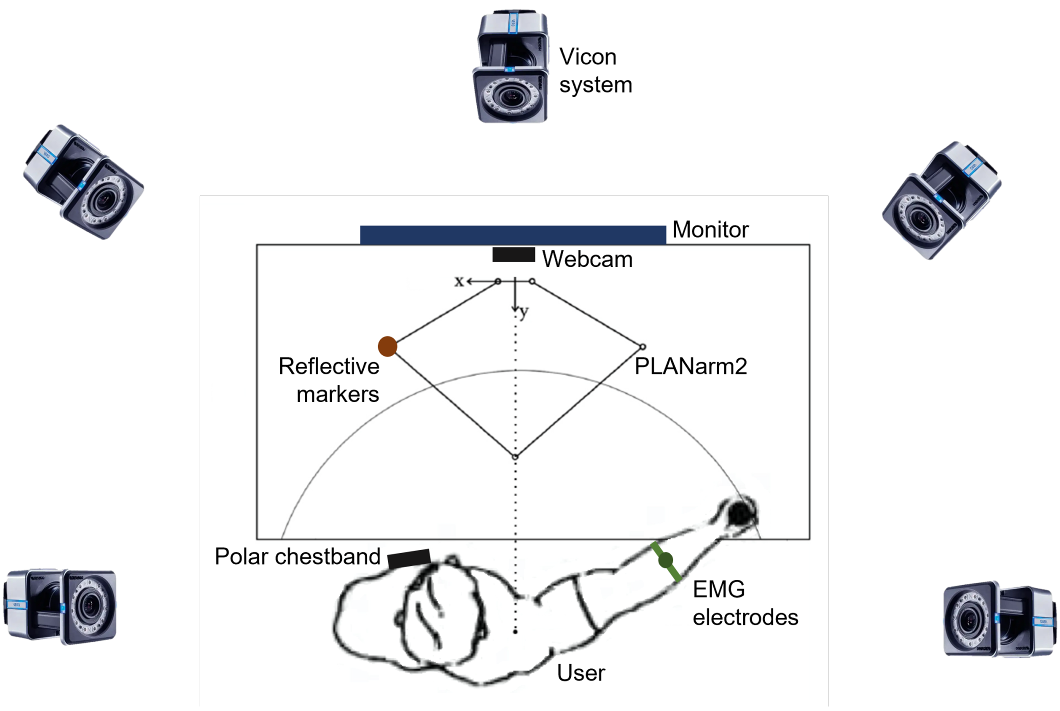


FIGURE 4.16: A schematic overview of the experimental setup realized for the Empathetic Neurorehabilitation Trainer.

To complete the setup, the prototype just presented is clamped on a generic table and placed in the middle of the area visible to a Vicon 10 TVC system, able to track with high precision both the user and the device thanks to a series of adesive reflective markers. In the same lab-space, a 16-channels Cometa EMG system is also available and can be synchronized with the Vicon system to collect the user's muscle activations. Of course, the choice of these two sensors was purely driven by experimental validation needs and is not foreseen in a future real-life application. Inheriting from the setup realized from the MindBot project, a second set of affordable, less invasive and portable sensors is introduced. With reference to Figure 4.16, a monitor is placed in front of the user to display the graphical user interface running on a dedicated laptop. On top of the monitor, a Logitech C920 HD Pro webcam (Logitech, 2021) is mounted to record the user in terms of behavior, gaze direction and facial expressions during the execution of the rehabilitation task. Additionally, a Polar H10 chestband (Polar, 2021) can be used to track physiological data.

4.2.2 Unity3D-based rehabilitation tasks

A Graphical User Interface (GUI) was specifically developed to enrich the PLANarm2 prototype with intuitive setup functionalities together with a list of motor tasks in the form of serious games. In particular, three different rehabilitation exercises, inspired from the related literature, have been implemented.

1. **Clock game**. A total of 9 targets are generated within the reachable workspace in a sunburst formation. One target represents the Center (C) of the sunburst while the other 8 cover all the remaining cardinal points: North (N), North-East (NE), East (E), South-East (SE), South (S), South-West (SW), West (W) and North-West (NW). Given these points, two different types of exercise can be performed. As a first option, the patient is requested to perform linear movements, as in Figure 4.17, from the center to a cardinal point and then back to the center. This task repeats until all the cardinal points have been reached. Alternatively, the patient is asked to move from the center to a certain cardinal point and from there reach the other appearing cardinal points through circular motions.

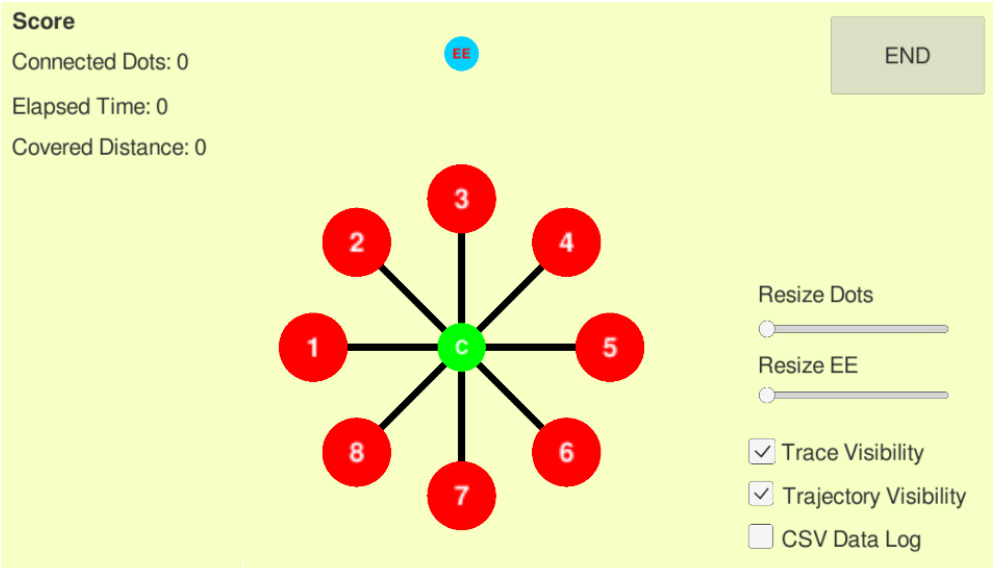


FIGURE 4.17: A screen capture of the *Clock* game.

2. **Connect the dots**. The GUI generates a series of numbered targets within the reachable workspace either randomly or on the basis of a past saved setting. A visual feedback of the current position of the end-effector is also displayed on screen, as shown in Figure 4.18. The patient is asked to move the handle of the PLANarm2 device in order to reach all the generated targets, connecting them one by one in order.

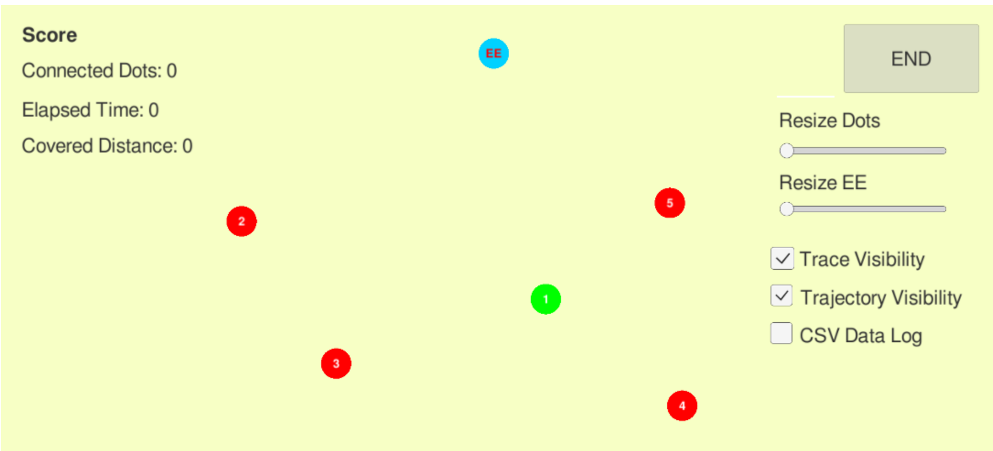


FIGURE 4.18: A screen capture of the *Connect the dots* game.

3. **Draw trajectory**. A starting and an ending point are generated within the reachable workspace and connected through a certain trajectory, as in Figure 4.19. This trajectory can either be generated on the spot by the therapist by moving a number of waypoints around or loaded from a previously saved setting. The patient is asked to bring the end-effector

on the starting target and, from there, follow the displayed ideal trajectory as close as possible until the end target is reached.

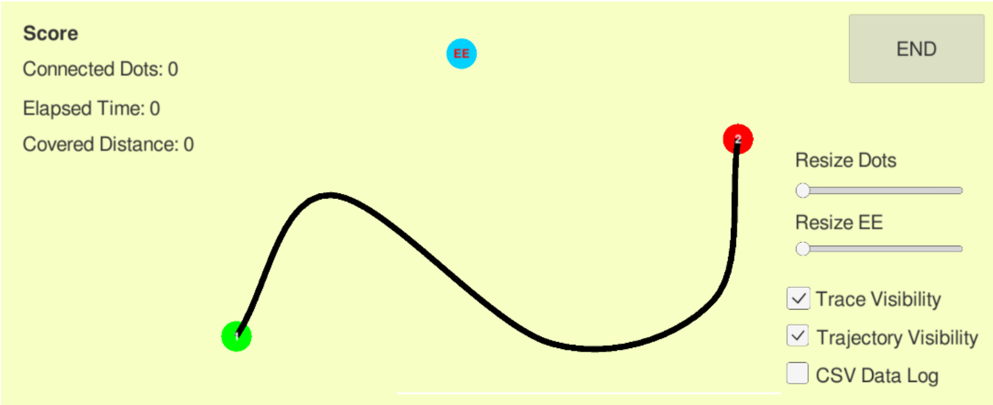


FIGURE 4.19: A screen capture of the *Draw trajectory* game.

During the execution of the task, the system logs all the relevant data related to the exercise: duration, traveled distance, exchanged forces, number of reached targets, average error, etc. Each exercise can be performed under a certain behavior of the robotic device: either completely passive, completely active or assistive-as-needed, as explained in detail in Section 4.2.3.

4.2.3 Software modules and interfaces

Once again, ROS is used to allow communication among all the developed software modules. Differently from the MindBot project, where some of the commercial devices required the use of dedicated computers communicating through a local network, all the modules developed for the Empathetic Neurorehabilitation Trainer can run on a single PC. The same software layers previously presented are used here to interface both SSI and VSM with the ROS environment. Additionally, the *Unity Robotics Hub* package (Unity Technologies, 2022) is used to connect the rehabilitation GUI with the rest of the system. Now, the details of how the main software modules have been implemented are reported.

Robot control

In order to develop an effective and modular control architecture, the author decided to leverage again the functionalities of the *ros_control* package (Chitta et al., 2017). Figure 3.3 specializes the functionalities of this package, already presented in Section 3.3.4, for the specific needs of the PLANarm2 prototype and application. As shown, the control structure can be split in low-level control and high-level control. The low-level portion of the control architecture is represented by the PID loops for position and velocity control running on the Arduino DUE board. These capabilities are often built-in for commercial robotic devices but, in this case, given the use of a general purpose Arduino DUE for cost-effectiveness and flexibility reasons, they had to be redesigned from scratch as needed for the custom made high-level controllers running on the external computer.

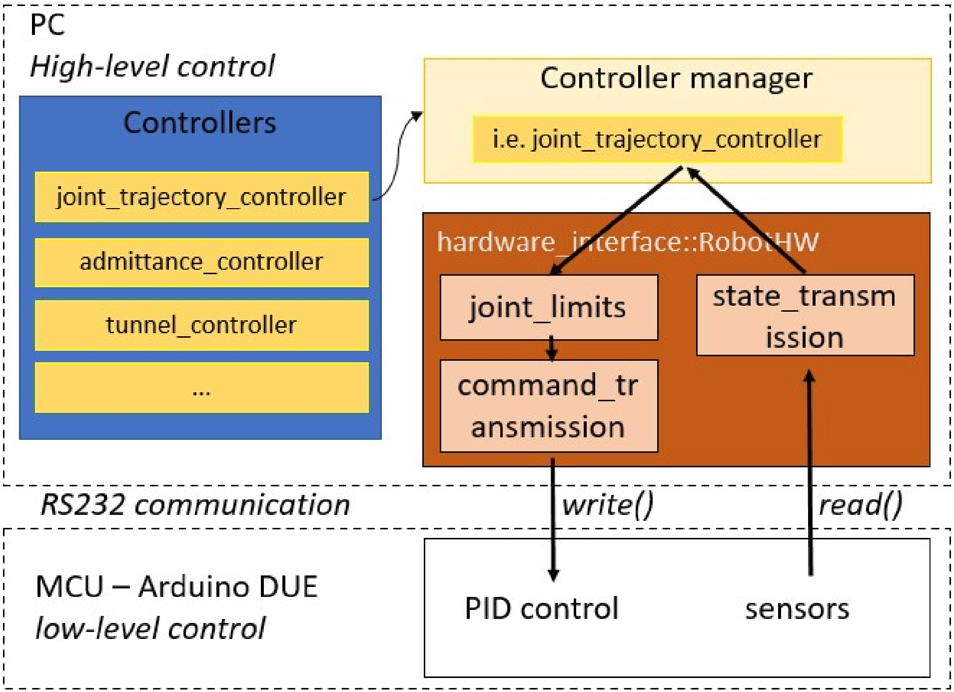


FIGURE 4.20: The *ros_control* functionalities deployed for the PLANarm2 prototype.

The reader can refer to Yamine et al., 2020 for more details about the reconstruction of the low-level controllers, while a brief overview of the three rehabilitation-specific high-level controllers developed for the project is reported here.

1. **Trajectory controller**. This controller can be used to perform passive rehabilitation exercises. Since it is a common tool, the author decided to exploit the so-called *joint_trajectory_controller*, available as part of the *ros_control* package. This controller takes as input trajectories specified as a set of waypoints to be reached at specific time instants and attempts to execute them as well as the mechanism allows. The author chose to interpolate between waypoints using quintic 1D splines, in order to guarantee continuity at the acceleration level. Thanks to the trajectory controller, PLANarm2 is capable of following any path that lays within

the workspace, while dragging along the arm of the patient that is holding onto the device end-effector. It is important to note that a software limit of force can be specified in order to avoid any harm to the patient due, for instance, to forced overextension of spastic muscles.

2. **Admittance controller**. Starting from Hogan's work (Hogan, 1985), indirect force control strategies such as admittance control can be considered the most proper and efficient way to control a robot interacting with its environment. For this reason, the author choose a strategy similar to the one described in Seraji, 1994. Given a reference force $F_r(t)$, coming from the digital environment connected to the device, it is possible to control the motors with a velocity reference (v_r) obtained through a PI control loop over the force error F_e, where $F_e(t) = F_r(t) - F_m(t)$ with F_m being the measured force. For the sake of simplicity, Equation 4.1 has been written only for one of the controlled axes:

$$v_r = \frac{1}{D_e} \cdot F_e(t) + K_i \cdot \int_0^t F_e(t')dt' \tag{4.1}$$

The proportional parameter in Equation 4.1 is called $\frac{1}{D_e}$ to highlight that the transparency felt by the user will increase while D_e, that can be associated to a virtual damping, decreases.

3. **Tunnel controller**. Corrective rehabilitation is proven effective when aiming to improve motor coordination. To provide this functionality, the author decided to develop a so-called tunnel controller, taking inspiration from Ding et al., 2014. It takes as input a predefined trajectory and builds a virtual tunnel of user-defined width around it. The patient is allowed to move freely along the path and, whenever the tunnel's boundaries are exceeded, a restoring force is produced in order to correct the undesired movement. A schematic representation of this concept is reported on the left of Figure 4.21. Differently from the trajectory controller, for which input trajectories are time-parametrized, the tunnel controller requires paths expressed in terms of curvilinear abscissa s. In order to guarantee coherence with the other controllers, a method that automatically transforms a time-parametrized trajectory into its corresponding s-version has been implemented so that the same computed trajectory can be applied to all the available controllers. Also, a new coordinate system $(\vec{t},\vec{n})$ has been defined on the trajectory $f(s)$ at any instant, denoting by $\vec{t}$ and $\vec{n}$ the tangential and the normal vectors

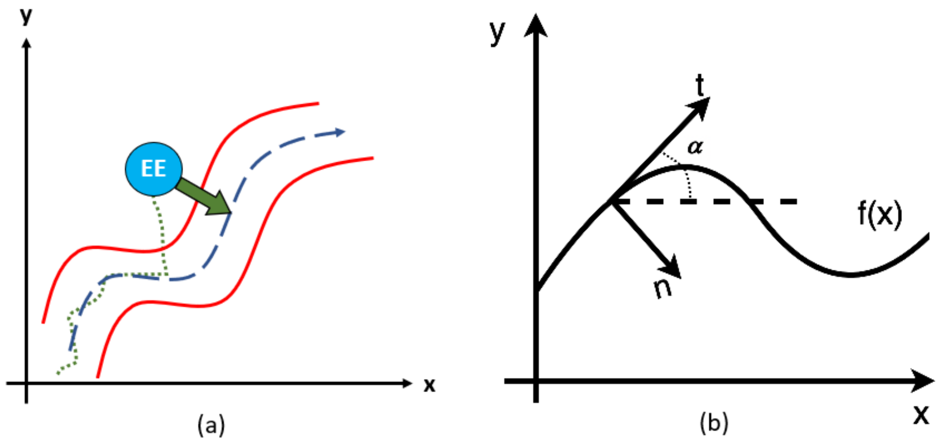


FIGURE 4.21: (a) Schematic representation of the tunnel built around the ideal trajectory and the corrective force produced on the device end-effector. (b) The new coordinate system built onto the ideal trajectory following its curvilinear abscissa.

respectively, as shown on the right of Figure 4.21. The patient's force on the end-effector is projected from the Cartesian reference frame to the new reference frame according to the instantaneous slope α of the requested trajectory. Then, the controller's basic working principle is similar to the one of the admittance controller. For every control cycle, the normal distance n_{ee} between desired and actual position of the end-effector, with respect to the given trajectory, is calculated. If that distance is smaller than the user-defined tunnel half-width W, tangential and normal measured forces are given as input to a high-level PI loop set with a reference of $0N$. On the contrary, if the end-effector is detected outside said tunnel, the force F_{refN} used as reference for the PI loop related to the normal direction is computed as in Equation 4.2, where K_v represents the stiffness of the virtual spring responsible for the generation of the corrective force.

$$F_{refN} = n_{ee} \cdot K_v \tag{4.2}$$

The effect of this approach is that the patient is allowed to move freely inside the virtual tunnel but, whenever the boundaries are exceeded, a virtual spring generates a corrective force that compensates the error and guides the end-effector back inside the tunnel. On top of this, an acceleration limit has been implemented within the controller's logic for safety reasons: if any spasm or sudden movement of the patient occurs, it can be absorbed.

Data acquisition and interpretation

Affective signals collected from the patients can be used to infer useful information about their experience. Home-based healthcare systems frequently leverage a diverse range of affective signals (Majumder et al., 2017; Philip et al., 2021; Wang et al., 2021). Of these, the project partners decided to focus on the more relevant ones when considering a neurorehabilitation scenario:

1. **Attention**. Motivation and attention serve as crucial modulators of neuroplasticity, influencing the outcomes of rehabilitation therapy (Cramer et al., 2011). Distractions, stemming from factors like boredom or lack of motivation, can disrupt the user's engagement during training sessions. Hence, the user's attention level becomes a pivotal input for the system's motivational strategy in neurorehabilitation. While previous studies in various domains have demonstrated the prediction of attention through physiological signals such as EEG (Acı, Kaya, and Mishchenko, 2019; Souza and Naves, 2021), these methods require proper sensor placement and additional user training on sensor usage. A more practical alternative lies in camera-based solutions, which capitalize on a common behavioral cue associated with distraction: looking away from the task. Research in other domains (Zaletelj and Košir, 2017; Smith, Shah, and Vitoria Lobo, 2003; Prajod et al., 2023) has indicated that facial and body pose features, including gaze direction, head orientation, and body posture, can effectively detect loss of attention. Inferring attention from such features is contingent on the setup (e.g., screen position), and detection models need to be appropriately calibrated.

2. **Pain**. Research on the occurrence of pain within the neurorehabilitation population and the consequent necessity for medical interventions has been extensively explored in works dedicated to neurorehabilitation (Benrud-Larson and Wegener, 2000; Castelnuovo et al., 2016). In the realm of healthcare applications, numerous systems employ image or video-based automatic pain detection (Kunz et al., 2017; Sellner, Thiam, and Schwenker, 2019). These approaches typically entail the identification of pain based on facial expressions captured by a frontal camera.

3. **Stress**. Detecting stress becomes crucial, especially with the introduction of gamification elements in the training session, where the patient may experience stress, particularly if the exercise surpasses their

current skill level. Extensive research has explored diverse modalities for stress detection, encompassing physiological signals, speech, gestures, and contextual behavioral patterns (Koceska, Koceski, and Simonovska, 2021; Larradet et al., 2020; Giannakakis et al., 2019; Heimerl et al., 2023). Physiological signals, including ECG, BVP, EDA, and respiration, have demonstrated high efficacy in stress detection (Gedam and Paul, 2021; Prajod, Mahesh, and André, 2024; Smets, De Raedt, and Van Hoof, 2018). Audio or speech analysis is another prevalent modality for automatic stress recognition (Dillon, Teoh, Dillon, et al., 2022; Lefter, Burghouts, and Rothkrantz, 2015). However, this approach typically involves substantial verbal interaction with the agent, a scenario not anticipated during neurorehabilitation exercises.

Given the above considerations, SSI functionalities are leveraged in collaboration with the University of Augsburg to deploy dedicated data acquisition and interpretation pipelines. With reference to Figure 4.22, images coming into the system through the Logitech webcam undergo a cropping phase to extract the detail of the user's face by leveraging the face detection model (Bazarevsky et al., 2019) provided by MediaPipe. This represents the fundamental information fed to the attention and pain detection pipelines.

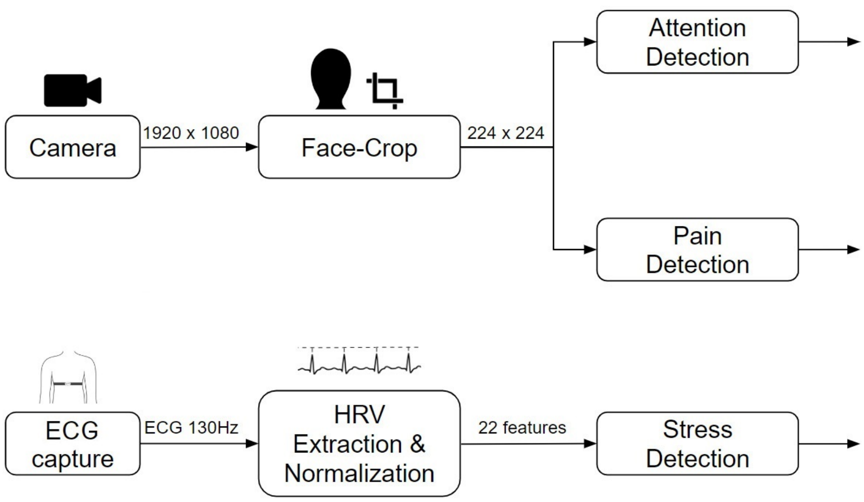


FIGURE 4.22: The SSI pipelines deployed for the Empathetic Neurorehabilitation Trainer.

Starting from the detection of the level of user's attention, a VGG16 network for gaze estimation is trained using the ETH-XGaze dataset (Zhang et al., 2020). After that, a transfer learning approach is leveraged to detect task specific states of distraction. For that, the prediction layer of the gaze estimation network is fine-tuned using a specifically collected dataset. Thanks to this approach the system is capable of discerning between states of attention (user looking at the monitor) and distraction (user looking away) with an average accuracy of 84.6%.

Moving to the topic of pain detection, a major challenge is posed by the fact that pain datasets are typically small for training deep learning models (Wang et al., 2018; Hassan et al., 2019; Xiang et al., 2022). To circumvent this, a solution can be found in the transfer learning approach described by Prajod et al., 2021, which involves leveraging features learned for emotion recognition in pain detection. To this end, an emotion recognition model is trained using a large dataset called AffectNet (Mollahosseini, Hassani, and Mahoor, 2019). The model is then fine-tuned using images from two pain datasets: UNBC-McMaster shoulder pain expression database (Lucey et al., 2011) and BioVid heat pain dataset (Walter et al., 2013). Both these datasets are derived from video sequences and thus, have virtually repetitive images. To mitigate this redundancy, images can be selected following the strategy proposed by Prajod, Huber, and André, 2022. Thanks to this step, the prediction layer of the emotion recognition model is modified for a 2-class prediction of pain and no-pain classes with an average accuracy of 78% on the test set.

Additionally, ECG data collected through the Polar H10 chestband are sent via bluetooth to the PC and injected inside a dedicated pipeline responsible for the extraction of hand-crafted HRV features. This choice is based on the observations presented by Prajod and André, 2022, where HRV features showed more generalizability than models based on raw ECG signals. The ECG signals from the WESAD dataset (Schmidt et al., 2018) are used to derive the HRV features for training the stress detection model. A total of 22 HRV features are computed from the time domain, frequency domain, and poincaré plots. A Support Vector Machine with the radial basis kernel function is then trained to predict if the user is stressed or not. To mitigate the individual differences in the signal and derived features (e.g. resting heart rate), the signals undergo MinMax normalization with a final average accuracy of 87%.

Task management and orchestration

Once again, the Visual SceneMaker tool is used as a powerful state machine capable to orchestrate the rehabilitation task by aligning the behavior of both the PLANarm2 prototype and the interactive virtual agent. The goal is to achieve a good integration between the two entities so that the avatar can be perceived as the virtual embodiment of a technologically reconstructed medical coach that can physically interact with the user through the adaptive behavior of the robotic device. A simple logic that can be used to control these two agents is reported in Figure 4.23.

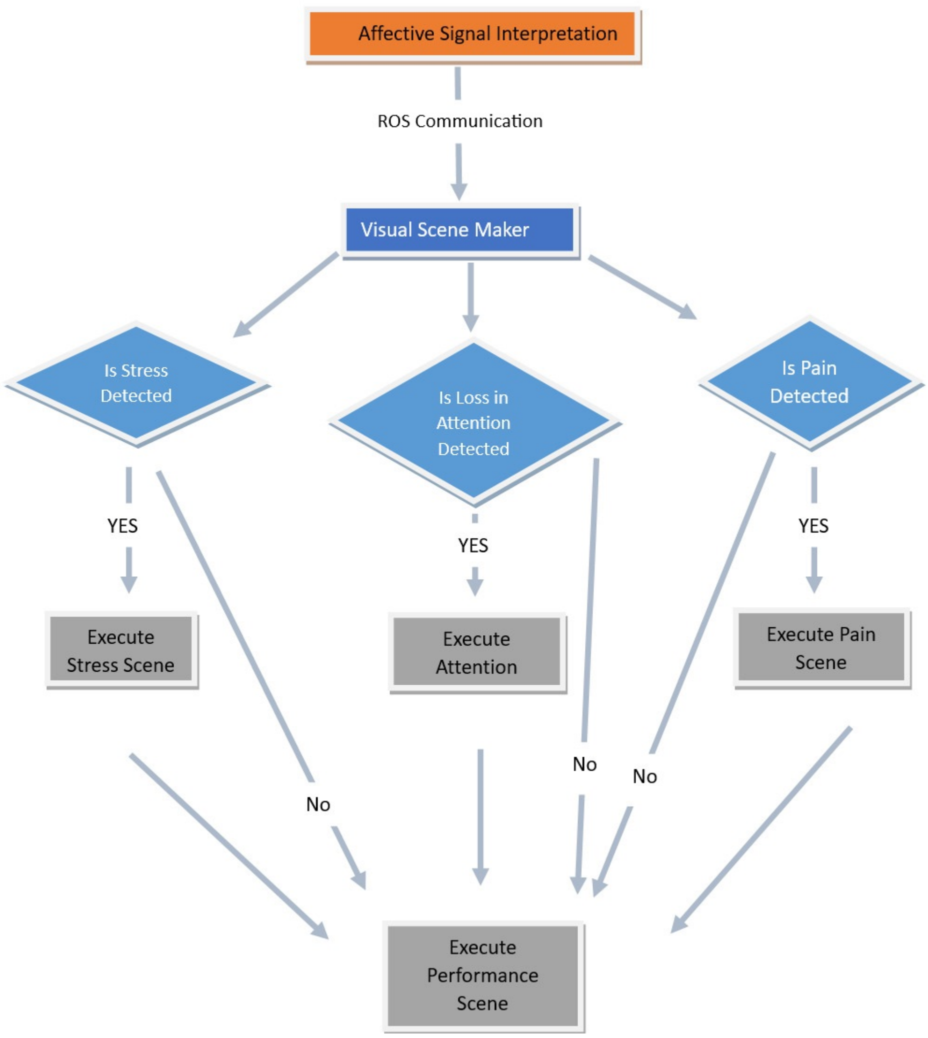


FIGURE 4.23: The logical structure of the VSM program developed for the Empathetic Neurorehabilitation Trainer.

The interpreted affective signals produced by the SSI pipelines are pushed within the ROS framework in the form of topics so that they are directly accessible from the VSM program. The depicted logic works as a generic decision tree where, depending on the specific detected state, a different reaction of both the robot and the avatar can be triggered. Thanks to the collaboration with DFKI, the virtual character is equipped with advanced speech synthesis and animation capabilities, making it possible to command lifelike, contextually appropriate interactions based on social cues. The speech output is generated using the Nuance Text-To-Speech system, which supports precise lip-syncing and manipulation of speech patterns. Animations are, instead, controlled through direct manipulation of the model's skeletal joints, enabling nuanced and dynamic physical responses (Gebhard et al., 2014). Overall, the avatar can perform 54 conversational gestures, captured via motion capture technology and adjustable in real-time, and express a range of 14 facial expressions, including the six basic emotions defined by Ekman (Ekman, 1992). On the robotic side, instead, the parameters available for control are the ones related to the level of assistance provided to the user and to the level of difficulty of the proposed task. Given all these controllable variables, the reaction logics can be personalized depending on the specific user. For instance, an exercise that is too challenging for the user could lead to the detection of a stress state. As a reaction, the avatar could provide mental support while the robot could increase the level of assistance and make the proposed exercise easier in order to balance the task complexity with the user's abilities. On the contrary, an easy task could be boring for the user with a consequent decrease in the detected level of attention. In this case, the avatar could intervene vocally in order to draw the user's attention back towards the exercise, while the robot could reduce the provided assistance and raise the game complexity so that the user is forced to focus on the task in order to perform well.

4.2.4 The resulting demonstrator setup

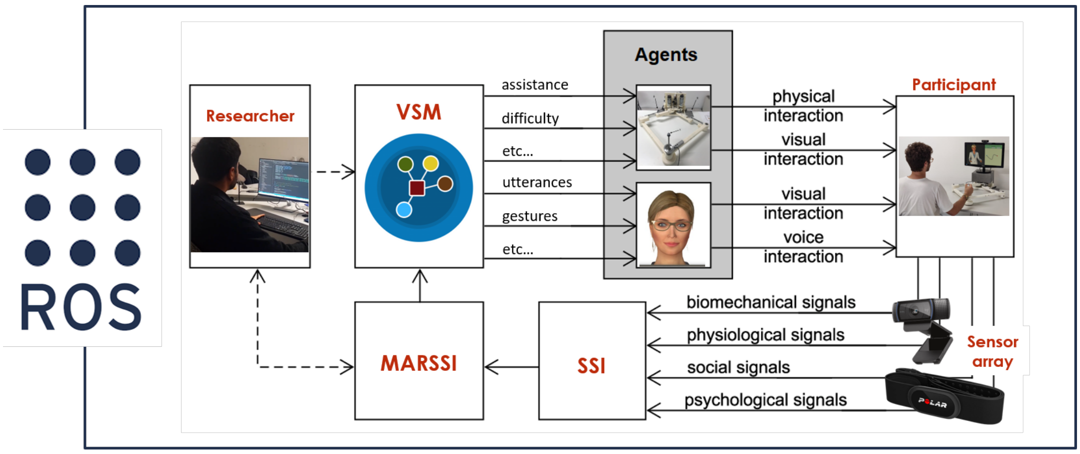


FIGURE 4.24: A schematic representation of the generalized human-driven control architecture deployed for the Empathetic Neurorehabilitation Trainer.

Thanks to the successful implementation of all the hardware and software features envisioned for the Empathetic Neurorehabilitation Trainer, the final demonstrator is ready to be used for the experimental activities. Figure 4.24, shows how the conceptual control architecture presented in Chapter 3 has been specialized and deployed for this specific application. It is important to notice that most of the implementation effort is shared between the two use-cases under analysis within the present project. The main differences between the two deployments consist in the choice of the hardware (robot and sensors), effectively absorbed by the inherent flexibility and modularity provided by ROS and SSI, and in the definition of the task, anyway simplified by the intuitive authoring framework offered by VSM. Now, the resulting experimental setup is depicted in Figure 4.25. As shown, a user is sitting in front of a table wearing the Polar H10 chestband under the t-shirt. The PLANarm2 prototype is clamped to the same table and a monitor is placed on top of it. The user is in direct physical contact with the robotic device through the ergonomic end-effector handle. In front of him, the monitor displays the rehabilitation task within the Unity3D-based GUI, enriched by the presence of an avatar acting as a virtual coach. The ECG data collected by the chestband is completed by the face-cropped images coming from the webcam placed in front of the user. During the exercise, the system monitors both the execution performance and the inferred user state in order to proactively adapt the behavior of the two interactive agents coherently and consistently and provide a better interaction experience.

FIGURE 4.25: The Empathetic Neurorehabilitation Trainer.

4.3 Approval from the ethical committee

Given the nature of the project, the involvement of human participants is required for the collection of real-life measures. For this reason, before being able to run the foreseen experimental campaigns, it is necessary to obtain ethical clearance from a designated committee. Since the studies are performed in the facilities of STIIMA–CNR in Lecco, the necessary documentation has been submitted to the institute's Commission for Research Ethics and Integrity. Moreover, since the project foresees the recruitment of adults characterized by ASD, the additional opinion of a clinical institution was requested. The ethical committee of IRCCS Eugenio Medea La Nostra Famiglia, partner of the project, was therefore involved to express their feedback from a clinical point of view. In these terms, a couple points should be highlighted:

- The proposed studies do not have any diagnostic or clinical goal.
- All the recruited participants, including those with ASD, must be healthy adults capable of giving their voluntary, autonomous and informed consent for the participation in the study.

On these bases, the required documentation was prepared. First of all, the study protocol was written to explain the goals and planned methodologies of the studies. A list of the chosen robotic devices was provided including both commercial and prototypical solutions. For the commercial devices, the available certifications guaranteeing their safety were attached to the proposal. Regarding the prototypes, instead, a risk analysis compliant with the Italian regulation for medical devices was produced. Inclusion and exclusion criteria were clearly stated together with the foreseen actions to be taken in case of dropouts or deviations from the protocol. The list of data to be collected was defined and completed by a description of all the planned data treatment activities. In compliance with the ethical principles of the Declaration of Helsinki, all data are made anonymous using alphanumeric codes. A protected file storing the link between the mentioned codes and the corresponding participants is kept in order to guarantee the rights to erasure, rectification and restriction. This file will be permanently deleted as soon as the goal of the data collection is achieved or the present project reaches its end. On this regard, a Data Protection Impact Assessment (DPIA) was prepared and positively evaluated by the institution's Data Protection Officer (DPO). Finally, the following documents were written to collect the signatures of participants before the start of the experimental sessions:

- *Informative sheet*: A document explaining the goals of the study to be signed for acknowledgement.
- *Data treatment sheet*: A document explaining the criteria and procedures that will be used for the treatment of personal data to be signed for acknowledgement.
- *Informed consent sheet*: A document to be signed to agree with the participation in the study, including the consent for data collection and treatment.

All the documents presented above, including the signed DPIA, have been submitted to the ethical committee of STIIMA-CNR who approved the full study under the condition of a positive feedback from a clinical institution regarding the involvement of ASD participants. For this, the ethical committee of IRCCS Eugenio Medea La Nostra Famiglia was consulted and granted its clearance, concluding the process of study protocol approval. Thanks to this achievement (see Appendix C), fundamental for the purposes of the present project, it is now possible to run a series of experimental activities, presented in detail in Chapter 5.

Chapter 5

Experimental campaigns and results

This Chapter is dedicated to the research activities carried out for the present project and to the obtained results. Each of the following studies have been published in international peer reviewed journals and conferences. Here, the main content is reported, but references to the full publications are available for further details. At the end of each section, a recap of the main take-aways is provided, before the final conclusions in Chapter 6.

First, Sections 5.1, 5.2 and 5.3 are dedicated to the inference of the user's biomechanical, social and psychological state, respectively, and to their integration within the generalized human-driven control architecture. Then, these measures are also used to drive the behavior and interventions of a virtual character in Section 5.4. Finally, Section 5.5 provides a comparison between neurotypical and ASD participants, fundamental to achieve true personalization of the HRI experience.

5.1 Biomechanical assessment and ergonomics

One of the first macro topics of interest is the biomechanical assessment of a user in interaction with a robotic device. In fact, in all fields of application it is of utmost importance to be able to monitor the user's physical fatigue. With this approach, the interaction experience can be improved by optimizing the task to limit instances of pain or discomfort. Also, a proper biomechanical analysis of the user's posture can provide great benefits in terms of task ergonomics. This is true for both of the use-cases of interest for the project: robot-based industrial manufacturing and rehabilitation scenarios.

5.1.1 Biomechanical assessment of the upper limb for determining fatigue, strain and effort: a review

Industry 5.0 aims at creating a synergy between humans and autonomous machines (Nahavandi, 2019), driving the transition to a human-centered and sustainable industry (Xu et al., 2021). These recent developments are pushing companies and stakeholders to introduce measures designed to ensure the overall wellbeing of their workers, a fundamental step to improve working conditions and reduce work related musculoskeletal disorders (WRMSD) (Sorensen et al., 2019). Of course, this translates into significant investments to transfer all those techniques, sensors (Scano et al., 2020) and findings coming from research in the bioengineering field to real-life industrial applications. For instance, upper limb fatigue, strain and effort have been repeatedly measured and assessed for various purposes, including the customization of working cells, load reduction and improvement of ergonomics (Kadir, Broberg, and Conceição, 2019). Moreover, recent research projects are starting to focus on these topics also within collaborative scenarios, where humans interact with cobots. However, since the adoption of these practices is still on-going, often there is no correspondence between the tests made in laboratories and those carried out in real working environments, leading to a gap between potential and actual applications. As a starting point to try and bridge this gap, a systematic review of the studies aimed at evaluating and assessing motor performance of the upper limb in the industrial field was carried out. Here the main highlights are reported, but the reader can refer to Brambilla et al., 2023b for additional details. Considering the available literature, the goal is to provide an answer to the following questions:

- What are the main obtained findings?
- Which kind of setting and equipment were used for the studies?
- What were the demographics of the involved participants?
- Which type of motor tasks were studied?
- Which analysis techniques were employed?

In order to do that, papers applying biomechanical analysis to industrial applications were considered using the international guidelines established by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al., 2009). A collection of articles was obtained by screening Scopus, and Web of Science (WOS) using the following logical query:

(shoulder ***OR*** *elbow* ***OR*** *wrist* ***OR*** *"upper limb"* ***OR*** *upper-limb* ***OR*** *"upper extremity"* ***OR*** *arm)* ***AND*** *(fatigue* ***OR*** *strain* ***OR*** *effort)* ***AND*** *(worker* ***OR*** *workplace* ***OR*** *industry* ***OR*** *industrial)* ***AND*** *(assessment* ***OR*** *index* ***OR*** *evaluation* ***OR*** *biomechanics* ***OR*** *measure* ***OR*** *measurement)*

A total of 1375 articles were found but, by considering only full journal articles, written in English and published after 2000, only 288 of them were considered eligible and included in the review. Below, a summary of the outcomes of this study is reported while the reader can refer to Brambilla et al., 2023b for further details.

Main findings

In terms of physiological conditions, the articles highlight that fatigue has effects on joint kinematics, torques and coordination and that it causes an increase of the power spectrum of velocity and acceleration. Duration, complexity and precision of the task increase muscle fatigue and it is possible to detect this state thanks to the measurements coming from EMG, IMU or Kinect sensors. Considering the characteristics of the task, also work pace, handled load, height and direction of movement have an impact on fatigue, pain and endurance. Assessments are therefore required to identify and mitigate the risk of associated muscolo-skeletal disorders. REBA, RULA, Strain index and the OCRA checklist are the most cited techniques to identify bad working postures for which ergonomic interventions are needed. Some prevention strategies are also mentioned, among all the use of exoskeletons reducing muscular effort, heart-rate and oxygen consumption and therefore delaying global fatigue.

Setting and equipment

With reference to Figure 5.1, 50% of the screened studies were performed in a laboratory environment while 47% of them were performed directly at the workplace or using data collected at the workplace. The remaining few studies either suggested protocols not yet implemented or worked on simulations.

As clearly shown, the selected studies split almost equally into two groups: those made in laboratory environment, and those performed in working places. Considering the number of recent publications, it is possible to identify a surprisingly increasing trend for the first group while the second group seems

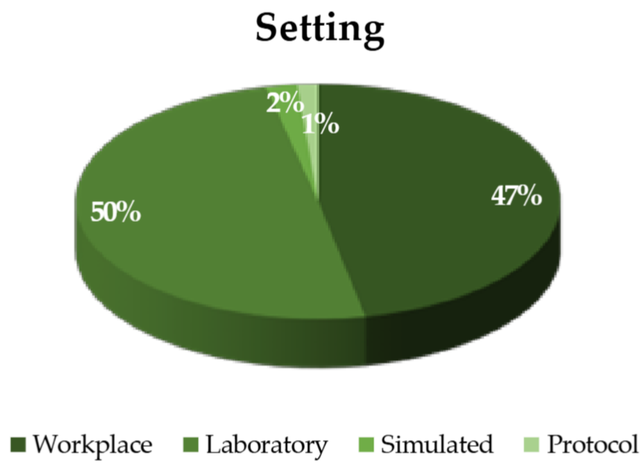


FIGURE 5.1: Settings distribution for screened papers.

to remain stable. Even though this trend seems to suggest that the research interest is focusing more on laboratory activities than their translation to the workplace, it should be commented in the light of the fact that the restrictions due to the COVID-19 pandemic may have had a strong impact on the field.

Demographics of participants

Keeping the distinction between lab and real-life settings identified above, a first interesting point of discussion can be found in the analysis of the cohorts of involved participants. As shown in Figure 5.2, 71% of laboratory studies have enrolled volunteers with no working experience related to the topic of the study and only 23% enrolled workers (the remaining 6% is simulated data). Conversely, 99% of the studies performed directly at the workplace enrolled workers.

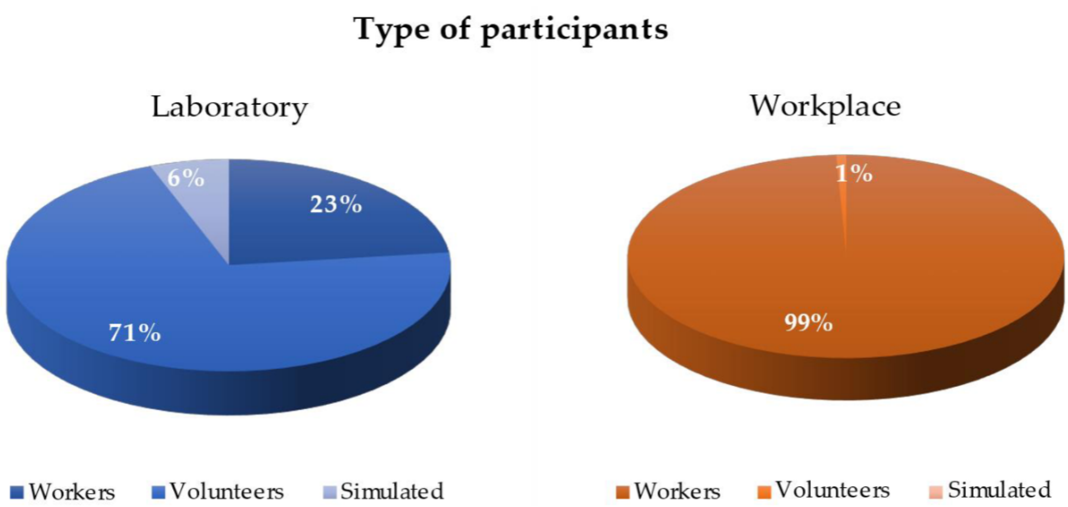


FIGURE 5.2: Participants type distribution for screened papers.

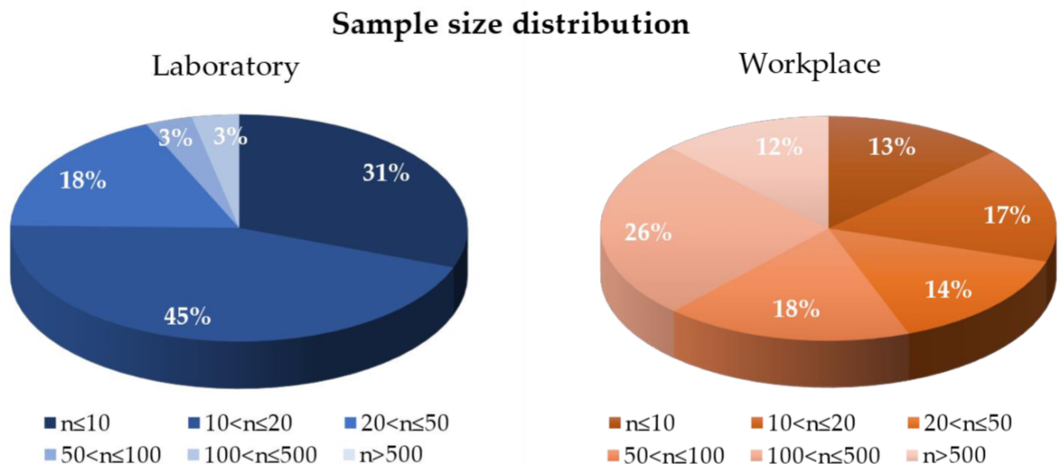


FIGURE 5.3: Sample size distribution for screened papers.

A great difference can also be found in the number of involved participants for the two groups of studies. Most of the laboratory studies involved less than 20 participants and only 10 papers involved more than 50 subjects. On the contrary, most of the studies in the workplace setting involved a high number of participants (>50) with 16 papers peaking over 500 subjects each. Figure 5.3 shows how the sample size distributes among the two groups.

Analyzed motor tasks

Starting from the task design distribution shown in Figure 5.4, laboratory studies are equally divided in repetitive (49%) and controlled (49%) movements, subject by nature to experimental limitations, and only 2% of them considered unconstrained movements, representing more realistic working conditions. On the contrary, most of the workplace studies (55%) considered unconstrained movements, 38% of them was based on controlled design, and only 7% was conducted in repetitive conditions.

Another relevant difference can be found in the type of supports (e.g., robots, tools, handles) used during these tasks. Since most laboratory studies regarded interaction with the environment and simulation of controlled tasks, 42% of them required tools and handles, including screwdrivers, hand supports, and others; 38% were based on free movements, while other supports (3%), end-effector robots (1%) and exoskeletons (16%) were leveraged in the other cases. On the other hand, in workplace studies participants usually performed their work during the entire workday and therefore the majority of them (75%) reported free movements, 20% used tools, while only 3% employed exoskeletons.

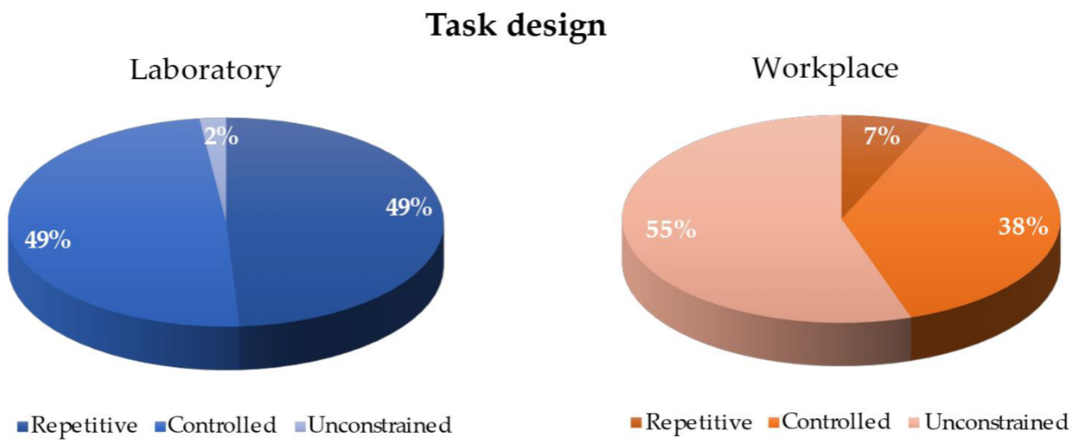


FIGURE 5.4: Task design distribution for screened papers.

Analysis techniques

Several approaches were employed in the screened studies as shown in Figure 5.5. Some instrumental approaches were based on EMG and kinematics, but also model-based approaches often included biomechanics and kinetics, with human models or recorded forces. Other approaches were, instead based on scales and questionnaires or a mix of the others. Interestingly, once again the type of assessments differs consistently between laboratory and workplace settings. In laboratories, EMG and kinematics are the most used methods to assess biomechanics, effort, fatigue and strain, employed in more than 50% of the studies. For workplace settings, instead, questionnaires and scales are by far the most employed ones (more than 80% of the studies).

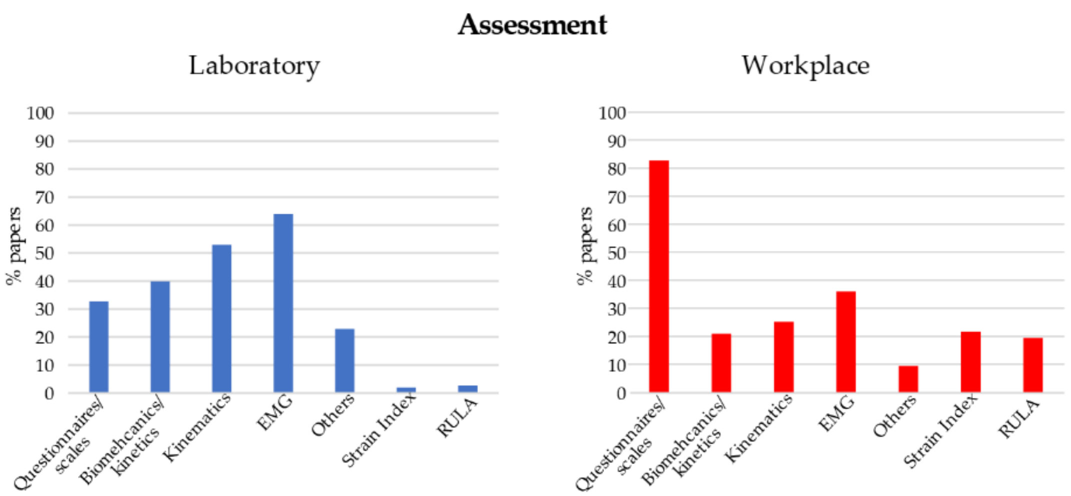


FIGURE 5.5: Assessment type distribution for screened papers.

All the discrepancies identified above put a question mark on the repeatability of the results obtained in controlled laboratory environments within actual working scenarios. It is therefore of utmost importance that a push towards the translation of lab findings to the real-world is promoted in order to actually leverage the obtained knowledge for the benefit of all the involved stakeholders.

5.1.2 Azure Kinect performance evaluation for human motion and upper-limb biomechanical analysis

In Section 5.1.1, a systematic review of the available literature on fatigue monitoring during industrial tasks highlighted key differences between studies performed in laboratory environment and those carried out in real working scenarios. One of the points of discussion was related to the type of sensors employed for the analyses. In general, marker-based optoelectronic systems are considered the gold standard in human motion tracking, but their use is not always feasible in industrial environments. On the other hand, marker-less sensors are relatively inexpensive, noninvasive and easy to use, but their accuracy can depend on sensor positioning, light conditions and body occlusions.

With the aim of establishing a common approach suitable for all kind of settings, it was decided to investigate the performance of the Microsoft Azure Kinect sensor, one of the most popular marker-less tracking systems on the market, in computing kinematic and dynamic measurements of static postures and dynamic movements. Here, a summary of the experimental procedures and results is reported while the reader can refer to Brambilla et al., 2023a for further details.

Materials and methods

Previous studies revealed that the tracking accuracy of Kinect cameras can be influenced by multiple factors, such as distance and orientation from the subject (Scano et al., 2020), light conditions (Romeo et al., 2021a) and body occlusions (Cai, Liu, and Ma, 2021). All these factors are likely to be variable and non-controllable in real-life scenarios. It is therefore important to analyze how much they affect the reliability of the collected tracking data in order to understand if they can be safely applied within various settings. For this purpose, 25 healthy adults (age: 28.5±4.9; height: 175.6±9.6 cm;

weight: 69.4±10.0 kg; 18M and 7F) were recruited and asked to perform specific movements while being tracked both by a Vicon Vero system (marker-based golden standard) and four Azure Kinect cameras positioned at different angles. The participants were asked to wear very tight vests or to be bare-chested at the time of trial, to facilitate placement of the markers on the anatomical landmarks. Also, an L-shaped table was used to simulate body occlusions, while a lux meter helped to standardize variable light conditions. Using these tools, both static and dynamic movements were assessed. With reference to Figure 5.6, in static trials the subject held three postures for 10 seconds: (i) with the right arm raised frontally, (ii) with both arms raised laterally and (iii) with both arms raised laterally and the elbows flexed. For dynamic movements, instead, both frontal and lateral reaching were assessed as motion primitives for several upper-limb tasks. The subjects started from the resting position, defined as the standing position with the arms relaxed by the side, extended elbows and non-elevated shoulders. Then, they were asked to raise the right arm, either frontally or laterally, at 90° with the palms facing downwards before going back to the resting position. Each movement was repeated ten times.

As mentioned, four Kinect cameras were used to investigate four different positions and orientations with respect to the subject in simultaneous acquisitions. In order to cover a wide field of view, as shown in Figure 5.6 the four chosen positions of the camera were: (a) in front of the subject (frontal view) at 1-meter height from the ground (frontal Kinect – KFront); (b) in front of the subject, from above at 1.80-meter height from the ground with an inclination of -22.5° (pitch) pointing downwards (frontal up Kinect – KFrontUp); (c) on the lateral side at 45° (yaw), from above at 1.80-meter height from the ground with an inclination -22.5° (pitch) pointing downwards (lateral 45° Kinect – KLat45); (d) on the lateral side (lateral view) at 90° (yaw) at 1-meter height from the ground (lateral Kinect – KLat). The tasks were performed with and without the joint occlusions provided by the presence or absence of the L-shaped table to quantify the resulting inaccuracies. Moreover, in order to standardize light conditions, tests were performed using artificial light, with no light coming from the outside. Two light conditions were assessed: illuminance was 268.76±23.3 Lux in the 'light' condition and 18.2±6.1 Lux in the 'no light' condition. Once again, the four combinations of occlusion and light intensity are shown in Figure 5.6.

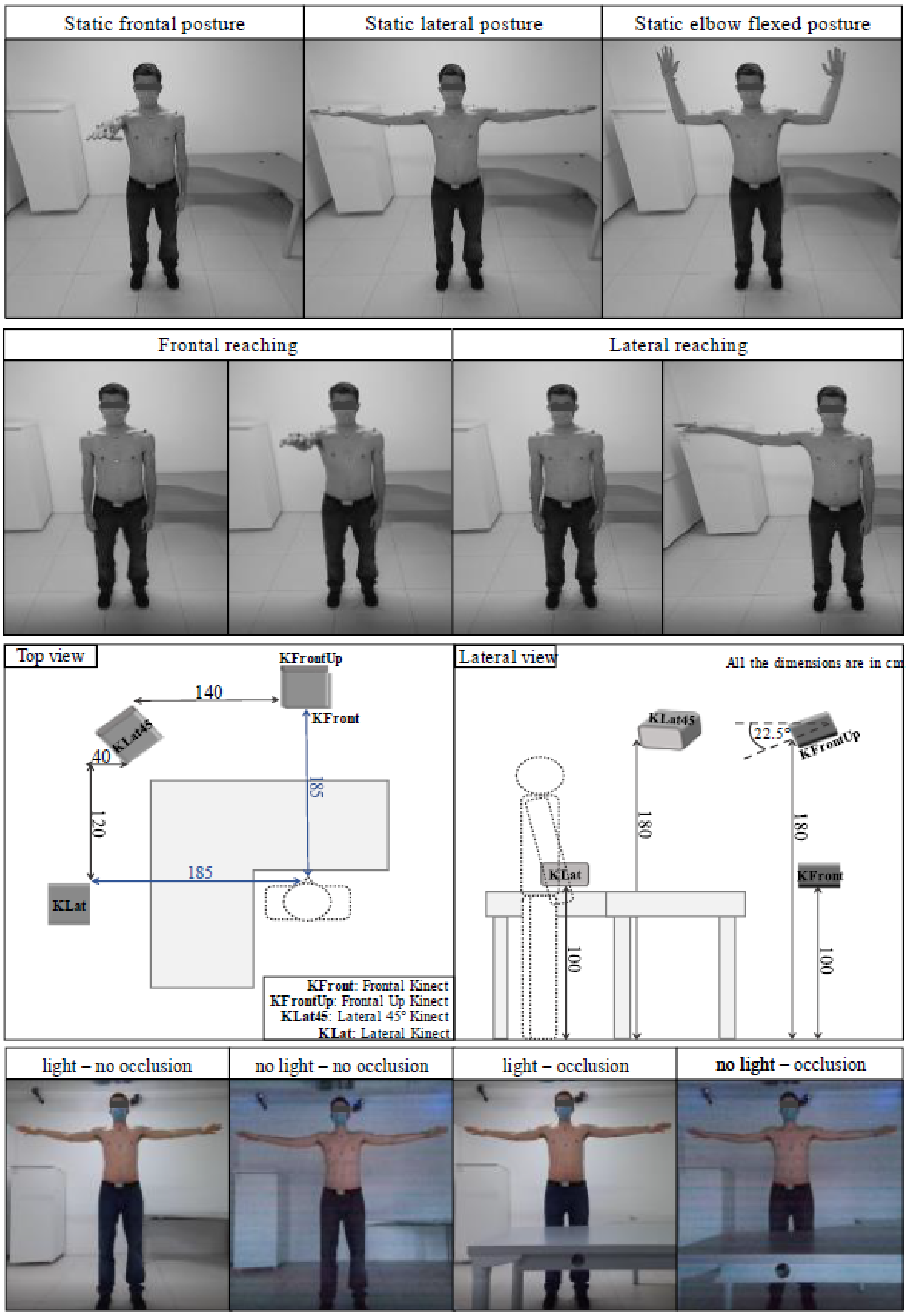


FIGURE 5.6: A summary of all the motor tasks and assessment conditions analyzed for the evaluation of the Azure Kinect sensor performance.

Markers were attached to the subject on anatomical landmarks corresponding to the Vicon Upper limb model requirements (Nexus, 2007). The data acquired by the Vicon were elaborated in the Vicon Nexus software to track and label the markers: the output obtained was the 3D coordinates of markers and joint center positions sampled at 100Hz. The acquisition from the Azure Kinect cameras was elaborated with the Microsoft Azure Body Tracking SDK (v1.1.1): the output was the 3D positions of 32 joints and quaternions indicating segment orientation sampled at 30Hz. Since the Vicon system and the Azure Kinect cameras were not temporally synchronized, movement onset and offset were identified in the data from each system, and the corresponding phases were aligned with a post-processing procedure (Ceseracciu, Sawacha, and Cobelli, 2014) using MatLab.

Data analysis

For data analysis, the raw data obtained from both systems were filtered with a 4th-order Butterworth low pass filter at a cut-off frequency = 5 Hz in order to remove noise artifacts. Since the acquisition systems have different sampling rates (100 Hz for the Vicon system and 30 Hz for the Kinect), the Azure Kinect data were up-sampled to 100 Hz with a shape-preserving piecewise cubic interpolation (Fritsch and Carlson, 1980) to allow data comparison between the two systems. Then, data were elaborated with a biomechanical model, which allowed the computation of kinematic and dynamic variables and motor control parameters. The biomechanical model takes as input the 3D position of markers and joint center for the Vicon system and the 3D coordinates of joints and the quaternions for the Kinect. Then, it reconstructs the 3D coordinate system of each segment and joint and computes joint angles. Joint moment and forces are computed with Newton-Euler equations. Power exerted at joint level, expended energy and normalized jerk are computed as explained in Brambilla et al., 2023a. For both Vicon and Kinect, the mass properties for each subject (mass, inertia matrix and center of mass of each segment) are estimated by accessing anthropometric tables (Christensen et al., n.d.) by height and weight of each participant.

The parameters computed for the analysis were divided into three categories based on the level of detail they provide. *Basic parameters* were not associated with a biomechanical model and included execution time, as the time needed to execute each movement and normalized body segment lengths (arm and forearm). Normalization is computed as the limb length divided by the subject's height and is needed for reliable inter-subject comparisons.

Kinematic parameters are, instead, extracted thanks to the mentioned biomechanical model and include joint angles (absolute, minimum, maximum and range of motion) and angular velocities, computed as the derivative of joint angles. Finally, *Dynamic parameters* are obtained using both the biomechanical model and the Newton-Euler equations and include: joint torques, peak power, expended energy and normalized jerk. A summary of all these parameters is graphically reported in Figure 5.7.

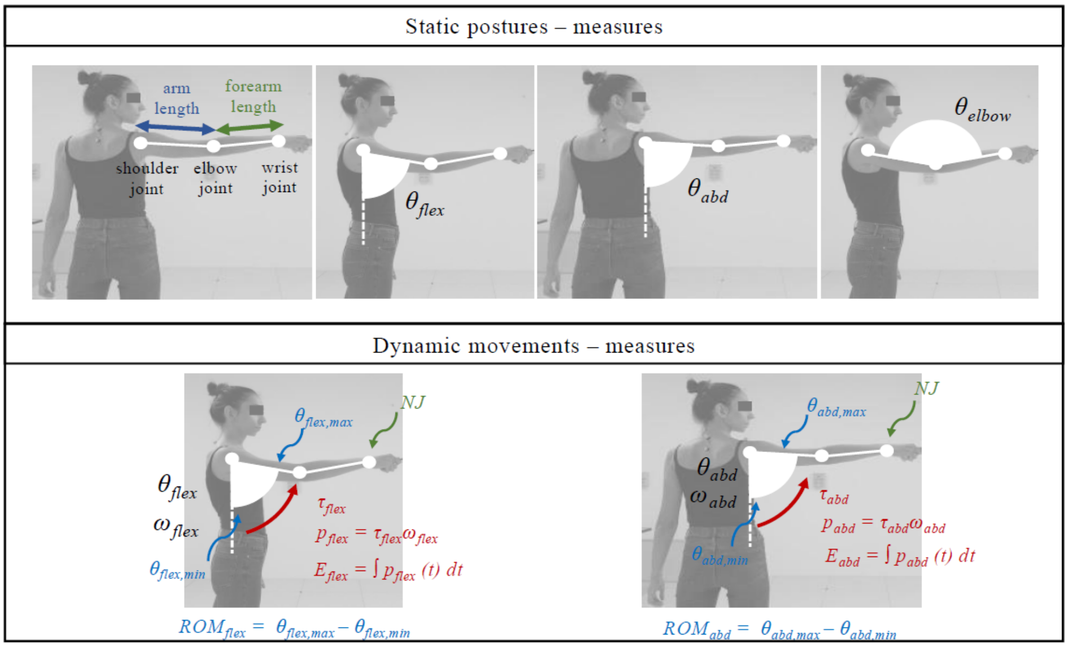


FIGURE 5.7: A graphical summary of all the assessment parameters used for the evaluation of the Azure Kinect sensor performance.

Results

First of all, the repeatability of all the parameters of the Vicon data is tested among the conditions to state whether it could be a standard, without statistical differences among the trials (i.e., in all conditions, the performed movements were the same). Since this condition is met, the parameters computed from Kinect data can be compared to the Vicon ones.

A first general consideration can be done considering all the tested conditions. After the normalization of the limb lengths obtained dividing the measures by the height of each subject, the Azure Kinect underestimates the length estimates. When considering articular angles, the Azure Kinect systematically underestimates the shoulder angles, and overestimates the elbow flection angles. The main reason of these different estimations can be due to the way the Vicon system provides the 3D coordinates of the articular centers. On the contrary, the Kinects provide an estimation of the joints based on the RGB and depth streams, processed by the SDK body tracking making it difficult to establish the exact correspondence of the detected joint positions provided by the two sensors. Now, comparing the results obtained with the different camera positions and orientations, the lateral Kinect had the lowest correlation with the Vicon system, with a worse estimation of biomechanical parameters and a higher variability between subjects. Instead, both the frontal and frontal raised Kinects showed high correlation coefficients with Vicon data, providing the most precise tracking when evaluating limb lengths and articular angles, thanks to the better point of view with respect to the analyzed motion tasks. Considering the presence of occlusions, the tracking accuracy was negatively influenced for all cases. As expected, this means that conditions without any object obstructing the field of view of the cameras lead to better results and the use of marker-less systems may be tolerated or not depending on the application. Quantification of these effects, useful to decide weather the Kinect sensors are suitable or not for the specific use-case, are reported in Brambilla et al., 2023a. On the contrary, light conditions do not seem to affect the performance of the Kinect tracking system making its use suitable for settings in which the light intensity cannot be easily controlled. Presented results are summarized below in Table 5.1.

TABLE 5.1: Summary of the results obtained for the different Kinect placements in terms of Vicon correlation, precision and robustness to light conditions and occlusions.

	Limb length	**Shoulder angle**	**Elbow flex angle**	**Vicon correlation**	**Occlusions**	**Light conditions**
KFront	under estimation	under estimation	over estimation	best correlation	disrupted performance	unaffected performance
KFrontUp	under estimation	under estimation	over estimation	best correlation	disrupted performance	unaffected performance
KLat45	under estimation	under estimation	over estimation	average correlation	disrupted performance	unaffected performance
KLat	under estimation	under estimation	over estimation	worst correlation	disrupted performance	unaffected performance

The performed analyses work on a complete characterization of the subject, and include parameters relevant for identifying fatigue and monitoring mental health conditions during repetitive activities and working cycles. Moreover, the computation of dynamic parameters allows the evaluation of risks related to musculoskeletal injuries. Overall, results demonstrate that the Azure Kinect sensors can be considered acceptable for the biomechanical assessment of the workers in industrial applications since the obtained biomechanical parameters are well correlated with the gold standard measures. Particular attention should, however, be paid to the presence of occlusions by mitigating the inevitable performance loss with properly chosen camera angles and, eventually, the fusion of data coming from different points of view. Azure Kinect cameras can also be employed for obstacle avoidance in human-robot collaboration, but the differences in the estimation of limb lengths and articular angles with respect to the Vicon reference should be considered while implementing the algorithms.

As a final note, it is important to understand that, even though the performance of the Azure Kinect sensor was tested with industrial applications in mind, the results can be directly applied also to the other use-case considered in the present project. In fact, the cost, invasiveness and complexity of use of marker-based system would not be feasible for robotic neurorehabilitation sessions, especially if considering a future where we hope to apply this kind of technologies in domestic environments.

5.1.3 A framework for human-robot collaboration enhanced by preference learning and ergonomics

Section 5.1.1 highlighted the importance of introducing biomechanical assessments in the workplace in order to minimize WRMSDs. Moreover, Section 5.1.2 evaluated the Kinect Azure sensors as suitable for the same purpose. Building on these findings, the next goal is to develop a preference-based optimization algorithm to improve working conditions by introducing an ergonomic assessment for HRC scenarios.

Assembly workers in manufacturing industry, are at risk for physical and mental health problems which are common and costly for both workers and their employers (Govaerts et al., 2021). Considering in particular collaborative assembly manufacturing tasks, engaging in repetitive motions, adopting awkward body positions, and consistently exerting excessive force can lead to overloading the musculoskeletal system. Thus, assessing an individual's ergonomics to properly implement HRC frameworks is crucial. However, it is also important to consider the preferences (Yan and Jia, 2022) and motivations of the individual to create a working environment that is both physically and mentally conducive to productivity and well-being.

For this purpose, it has been decided to work on the development and testing of an algorithm designed to optimize the operator's posture ensuring comfort and minimizing the risk of musculoskeletal disorders while, at the same time, matching the operator's personal preference. This is a novelty with respect to the approaches used in related studies, which either considered the topic of preference or ergonomics without trying to conciliate the both of them. Here, a summary of the study and related findings is reported, but the reader can refer to Falerni et al., 2024 for additional details.

Materials and methods

The flow of the proposed algorithm is schematized in Figure 5.8. As shown, an iterative minimization process is foreseen after an initial training. In this process, user preferences are used to address the cognitive workload of the collaborative task (e.g., user engagement, work-related stress), while the a quantitative index is used to improve task ergonomics.

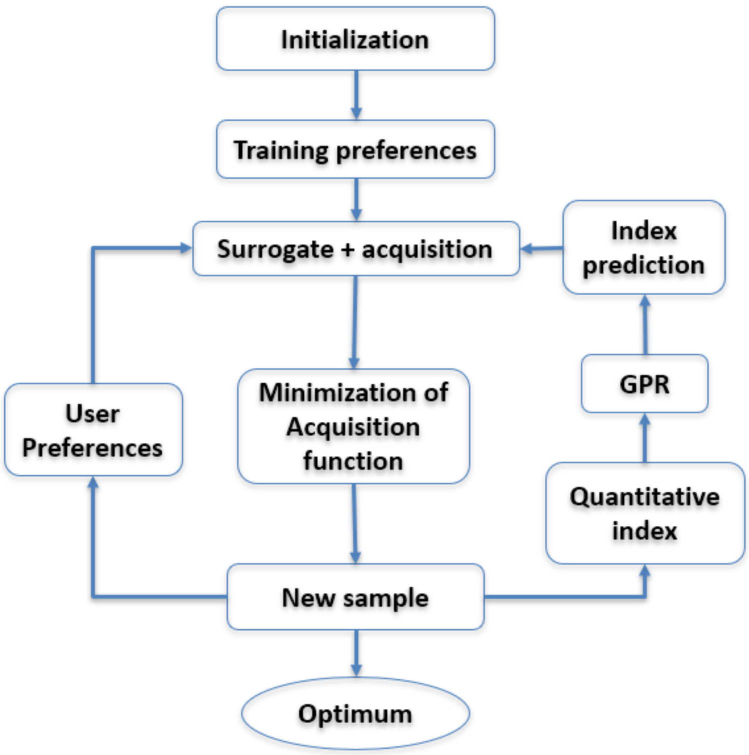


FIGURE 5.8: Block diagram representing the flow of the ergonomics and preference based optimization algorithm.

Preference-based optimization is a semi-automated technique for solving black-box optimization problems, where the explicit mathematical expression of the objective function is either expensive or impossible to obtain (Bemporad and Piga, 2021). In these cases, it is possible to exploit human preferences as a guide for the optimization process toward the optimal solution. For instance, Active Preference Learning (APL) (Bemporad and Piga, 2019; Bemporad, 2020) techniques can be leveraged to find the global optimum of an unknown function using only the preferences of a human decision-maker. For the purposes of this experimental activity, the AmPL algorithm proposed by Dao et al., 2023 is used as a starting point, as it expands the original approach with the introduction of multiple levels of preference, making it easier for the subject to compare between options. In AmPL, the objective function $f(x)$ is assumed to be non-accessible. Let $\mathbb{R}^n$ be the space of decision variables and x_1 and x_2 are two n-element vectors so that x_1, $x_2 \in \mathbb{R}^n$. Because the values of $f(x_1)$ and $f(x_2)$ cannot be quantifiable, only their comparison in the form of preference p: $\mathbb{R}^n \times \mathbb{R}^n \to \{-2,-1,0,1,2\}$ and corresponding certainty level c: $\mathbb{R}^n \times \mathbb{R}^n \to \{1,2,3,4\}$ are accessible. The overall *preference function* is a composition of these two elements and is defined as:

$$\begin{aligned} \pi : \mathbb{R}^n \times \mathbb{R}^n \to \{-2,-1,0,1,2\} \times \{1,2,3,4\} \\ \pi(x_1,x_2) = (p(x_1,x_2),c(x_1,x_2)), \end{aligned} \tag{5.1}$$

where:

$$p(x_1,x_2) = \begin{cases} -2 & \text{if } x_1 \text{ is "much better" than } x_2, \\ -1 & \text{if } x_1 \text{ is "better" than } x_2, \\ 0 & \text{if } x_1 \text{ is "as good as" } x_2, \\ 1 & \text{if } x_1 \text{ is "worse" than } x_2, \\ 2 & \text{if } x_1 \text{ is "much worse" than } x_2, \end{cases} \tag{5.2}$$

and:

$$c(x_1,x_2) = \begin{cases} 1 & : \text{not so sure}, \\ 2 & : \text{quite sure}, \\ 3 & : \text{sure}, \\ 4 & : \text{absolutely sure}. \end{cases} \tag{5.3}$$

Using this definition, at every iteration the user is asked to express his/her preference between two options and a level of certainty on the decision itself. Then, the algorithm works on the basis of three functions, as from Equation 5.4:

$$a(x) = \frac{\hat{f}(x) - \min\{\hat{f}(x_i)\}}{\Delta\hat{F}} - \delta z(x), \tag{5.4}$$

- $\hat{f}(x)$ is a *surrogate function* which takes into account the preferences of the user and tries to reproduce the answers to make predictions for decision vectors not explored yet;
- $z(x)$ is an *exploration function* used to avoid falling in local minima during the global optimization, with δ being the gain used to balance between exploitation and exploration behaviors.
- $a(x)$ is the *acquisition function* to be minimized, through global optimization, to get the next sample to be compared. It is the sum between the surrogate function and the exploration function.

Starting from the algorithm above (further details can be found in Dao et al., 2023), the acquisition function defined in Equation 5.4 is modified to the form of Equation 5.5, introducing two new elements to the optimization problem:

$$a(x) = \frac{\hat{f}(x) - \min\{\hat{f}(x_i)\}}{\Delta\hat{F}} - \delta\frac{z(x)}{\Delta Z} + P(x) + \eta\,\kappa(x), \tag{5.5}$$

- $P(x)$ is a *penalization function* used to avoid some zones of the feasible space when the user realizes that they are not optimal and it would be a waste of time to let the algorithm explore them.
- $\kappa(x)$ is a *quantitative function*, used in this case to account for the ergonomics of the task, with η being the gain used to modulate its impact on the overall acquisition function.

Among all the existing approaches for the evaluation of ergonomics, the Rapid Upper Limb Assessment (RULA) has been selected (Yazdanirad et al., 2018) as it provides a better estimation of posture-related risks with respect to other methods (Kee and Karwowski, 2007). Following this choice, user's body joint frames and positions can be acquired using Kinect Azure cameras and processed to extract the corresponding RULA score, as listed in Table 5.2. The higher the obtained score, the higher the WRMSD risk, with a consequent heavy influence of the quantification function on the overall acquisition function pushing the optimization process away from that unfavorable zone of the feasible solution space. Then, a Gaussian Process Regression modeling (GPR) is implemented to map the RULA index to the optimization variables and to the subject's anthropometric parameters.

Score	Level of WRMSD risk
1-2	Negligible risk, no action required
3-4	Low risk, change may be needed
5-6	Medium risk, further investigation
7	Very high risk, implement change now

TABLE 5.2: RULA score values based on the WRMSD risk.

Thanks to the use of ROS, this approach can be easily built into the generalized human-driven control architecture presented in Section 3.2. The involved components are highlighted in Figure 5.9. User posture and expressed preferences flow as biomechanical signals into the Interpreter module, which incorporates the presented algorithm. The User Model contains the anthropometric parameters used to map the obtain high-level indexes to the specific user and, consequently, obtain the information needed by the Orchestrator to optimize the task. The Supervisor is represented by the researcher that oversees the experimental session and inputs the data required by the penalization function, if needed. As a result, the Orchestrator commands a new collaborative assembly pose to the cobot, computed to optimize ergonomics while accounting for the user's personal preference.

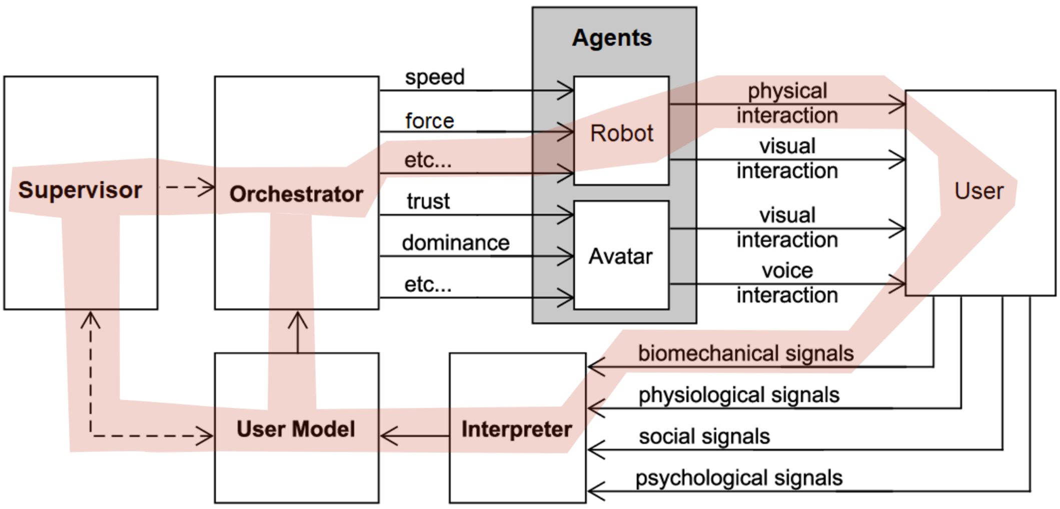


FIGURE 5.9: Highlighted control scheme showing how the presented algorithm is built into the generalized architecture.

Experimental assessment

Taking advantage of the above optimization algorithm deployed within the control architecture introduced in Section 3.2, of the hardware and software setup described in Sections 4.1.1 and 4.1.3 and of the assembly task presented in Section 4.1.2, an experimental campaign is prepared. The goal is to optimize the end-effector pose of the MindBot cobot when performing the direct collaboration step of meshing the gears of the epicyclic mechanism. With reference to Figure 5.10, the parameters to be optimized are:

- x, y and z representing the relative 3D spatial coordinates of the end-effector with respect to the base frame of the robot;
- θ_x and θ_y representing the vertical and horizontal orientation of the end-effector with respect to the base frame.

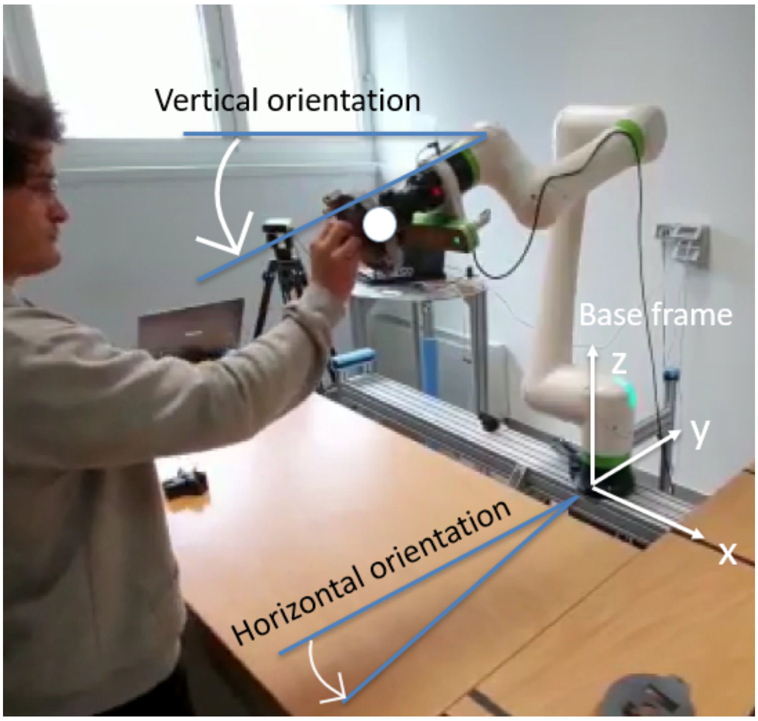


FIGURE 5.10: A representation of the optimization problem: The user and the cobot are in direct collaboration for the gear meshing step and the algorithm proposes new sets for the position and orientation of the end-effector.

Table 5.3 provides the domain and the variation step of the optimization variables used in the experiments. The *Step* parameter is tuned to make sure that the change in the robot configuration is big enough to be perceived by the user.

	X [m]	Y [m]	Z [m]	θ_x [°]	θ_y [°]
Upper bound	0.5	0.3	1.75	90	45
Lower bound	-0.3	0.0	1.15	0	-45
Step	0.1	0.1	0.1	10	10
Range size	0.8	0.3	0.6	90	90

TABLE 5.3: The domain and variation step of the optimization parameters chosen for the experimental sessions.

Each experimental sessions is composed of three phases. In the first *training phase*, an expert operator instructs the participant on how to assemble the components and express his/her preferences. During a second *optimization procedure*, pairs of robot configurations are tested. For every iteration, the user expresses his/her preference between the cobot pose that was considered the best so far and a new one suggested by the optimization algorithm. As an exit criterion, the maximum number of optimization iterations is set equal to 30 (experimentally found to provide a good balance between exploration and exploitation in the considered experimental scenario; other exit criteria based on, for instance, the satisfaction of the user can be used). In the final *validation phase*, the participant is asked to use the cobot manual guidance functionality to realize the configuration that s/he considers the best. The parameters of this configuration are saved so that a comparison with the optimum achieved by the algorithm can be produced.

A total of 20 participants (24-40 years old) have been involved in the experimental campaign. The mean participants' height is 171.5cm, with a standard deviation of 8.5cm. In the selection of participants, an effort was made to balance the number of right (9 participants) and left (11 participants) handed subjects in order to have a similar representation of the two groups within the tested population.

Results

Table 5.4 presents the obtained results in terms of pose "error" between the optimum obtained by means of the developed framework and the one set manually by the user through the manual guidance of the robot. Even though the number of iterations within a single experiment was relatively low considering the number of parameters to be optimized (Dao et al., 2023; Bemporad and Piga, 2019; Bemporad, 2020; Bemporad and Piga, 2021), they are sufficient to reach an admissible configuration of the robot with respect to the user expectations. As a matter of fact, according to the performance results presented in Table 5.4, the algorithm's error for each variable was found to be lower than the corresponding problem resolution specified in Table 5.3. This means that the distance between the pose computed by the optimization algorithm and the one chosen directly by the participant is smaller than what can be effectively perceived, so much that most of the volunteers could not distinguish between the two, when asked.

	x [m]	y [m]	z [m]	θ_x [°]	θ_y [°]
$\bar{e}$	0.06	0.05	0.09	7.66	11.65
$\bar{\sigma}$	0.07	0.03	0.06	11.18	7.73

TABLE 5.4: Mean error ($\bar{e}$) and standard deviation ($\bar{\sigma}$) for each variable.

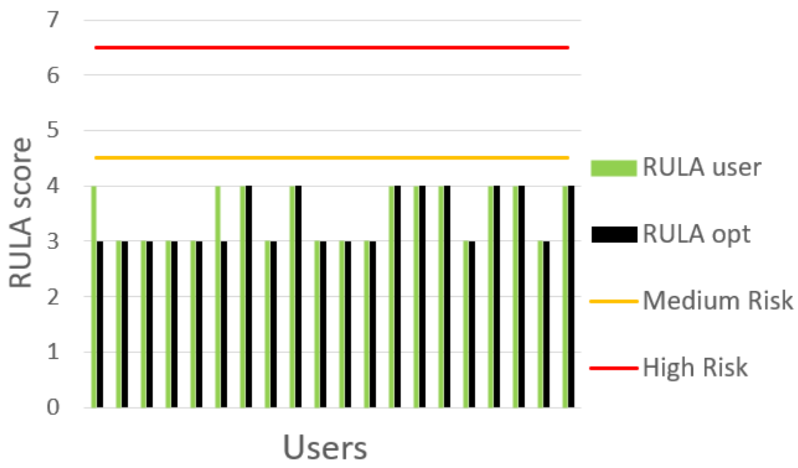


FIGURE 5.11: The RULA scores obtained by the optimization algorithm and by the user choice compared to high and medium risk levels.

Now, considering the results related to ergonomics, Figure 5.11 can be used to compare for each participant the RULA score corresponding to the configuration chosen by the algorithm and the one chosen by the subject. The most important achievement that can be inferred from the depicted data is that all RULA scores remain under the threshold of medium risk, meaning that the configuration of the robot allows the user to maintain a good posture that contributes to minimizing WRMSDs. Moreover, it is clear that, for each participant, the two RULA scores are very close to each other. This result is aligned with the expectations since the small difference between the two compared configurations, reported in Table 5.4, should also result in similar user postures. It is interesting to notice that the configuration manually chosen by some of the participants actually leads to a higher RULA score than the one obtained by the algorithm. This can be explained by the fact that postural comfort and ergonomics are two different concepts. While ergonomics focuses on postural parameters to guarantee the safety and well-being of users and prevent health problems, comfort refers a wider range of factors such as cognitive, physiologic, and environmental factors (Naddeo and Cappetti, 2014). Therefore, the most comfortable posture may be different from the ergonomically optimal one, as reflected by the results.

Even though the obtained results are strictly connected to this specific experimental setting and cannot be generalized, they serve as a positive indication that the algorithm is capable of optimizing user preference and ergonomics at the same time.

5.1.4 The effects of robotic assistance on upper limb spatial muscle synergies in healthy people during planar upper-limb training

The importance of introducing biomechanical assessments and monitoring in working environments, as presented up to now, can be directly applied also to the medical field and, in particular, to rehabilitation where the muscolo-skeletal sphere is at the center of focus. Since the pioneering approaches in the 90s (Aisen et al., 1997), in the last decades several studies (Krebs et al., 2004; Riener, Nef, and Colombo, 2005) showed that robotic rehabilitation is an effective technique inducing comparable or better motor improvements in respect to standard treatment (Volpe et al., 2000). However, the mechanisms underlying motor recovery and neuroplastic effects induced by robotic therapy are not well known and understood yet. It is in fact a matter of debate which guidelines should be followed when assisting rehabilitation with robots. While several good practice guidelines have been proposed, including: assist-as-needed paradigms (Reinkensmeyer et al., 2012), adoption of transparent robots (Just et al., 2018), biomimetic controllers (Abboudi et al., 1999), human-in-the-loop approaches (Nam et al., 2019), adaptive controllers (Yang et al., 2016), some relevant scientific questions are still open. In particular, detailed evidence on the effect of robot assistance on motor control have not been exhaustively identified (Broekens, Heerink, Rosendal, et al., 2009; Rodgers et al., 2019).

For the purposes of the present project, knowledge about how the introduction and regulation of robotic assistance influences motor control is key to design a system capable to proactively adapt its behavior in order to promote both motor recovery and positive interaction experiences. Therefore, the control architecture introduced in Section 3.2 and deployed for the experimental setup presented in Section 4.2 is leveraged to run a dedicated experimental campaign. Figure 5.12 highlights the leveraged hardware and software modules. Here, a summary of the study protocol and results is reported, but the reader can refer to Cancrini et al., 2022 for further details.

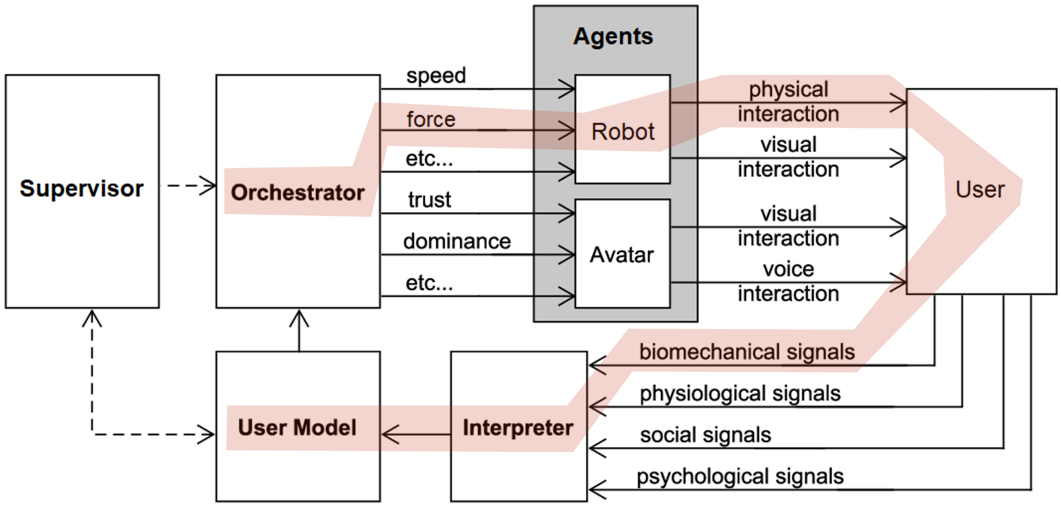


FIGURE 5.12: A highlight of the modules of the proposed control architecture used for the evaluation of muscular synergies.

Materials and methods

According to the theory of Muscular Synergies (MS), the central nervous system exploits a reduced set of pre-shaped neural pathways, called synergies, to achieve a large variety of motor commands. The potential of this method has already been exploited to gain deeper insights concerning motor impairment. For instance, MS have been used as metrics for the evaluation of robot-assisted interventions (Lencioni et al., 2021), finding that post-stroke subjects who followed robotic rehabilitation showed larger improvements in axial-to-proximal muscle synergies with respect to those who underwent usual care. It was also shown that robot-therapy induced subject-dependent modification of synergies (Tropea et al., 2013) and slight modifications of the original synergies (Scano et al., 2018). However, only a few studies have evaluated the effects of assistance and challenging conditions during human-robot interaction on healthy people. This assessment is a missing piece in the understanding of how robot assistance influences motor coordination and in the identification of which modes and approaches can maximize the patients' motor recovery.

Using the setup presented in Section 4.2, 10 healthy adult volunteers (33±9 years, 2F, 8M) have been recruited to take part in an experimental activity following a protocol inspired from Dumas, Cheze, and Verriest, 2007. In order to have an appropriate model-based tracking with the Vicon system, each participant was asked to wear a total of 11 markers positioned on T8 and C7 vertebrae, jugular notch, xiphoid process, greater tubercle, medial and lateral humeral epicondyles, styloid process of the ulna and the radial, second and fifth metacarpal heads. Six of the same markers were placed also on the robot: three on the base of the device as reference axes, one on the top of the right driven pulley, one on the right joint and one on the end-effector. Moreover, subjects were equipped with 16 s-EMG electrodes positioned according to the SENIAM guidelines (Hermens et al., 1999) on the following muscles: Erector Spinae (ES), Middle Trapezius (MT), Upper Trapezius (UT), Infraspinatus (IF), Deltoid Anterior (DA), Deltoid Middle (DM), Deltoid Posterior (DP), Pectoralis (PC), Triceps Long Head (TLo), Triceps Lateral Head (TLa), Biceps Long Head (BCl), Biceps Short Head (BCs), Brachioradialis (BR), Pronator Teres (PT), Wrist Flexors (WF) and Wrist Extensors (WE). The right side of Figure 5.13 shows the instrumented robot and subject.

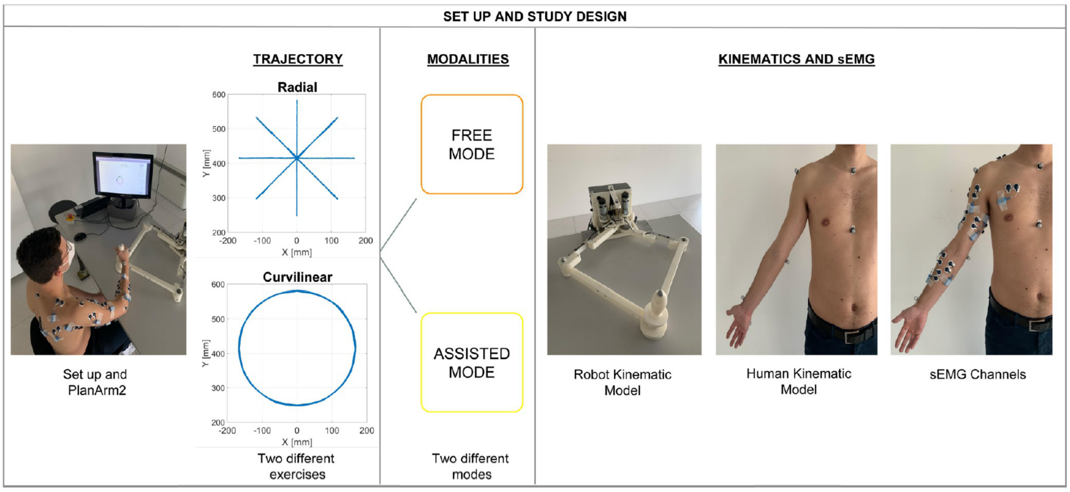


FIGURE 5.13: From left to right: the study setup, the trajectories proposed to the user (radial and curvilinear), the available robot modes (free and assistive) and the kinematic models used for motion tracking with the Vicon system.

First, a short adaptation phase was proposed to the participants in order to familiarize with the task to be executed. During this preliminary adaptation, the robot is moved by the *trajectory controller* (see Section 4.2) from target to target following rectilinear and curvilinear paths with trapezoidal-shaped biomimetic velocity profile. In this phase, each subject is asked to follow the robot with their dominant limb without pushing nor being pushed, thus keeping interaction forces as low as possible. Then, two experimental conditions of interaction with the robot are considered. A first condition, called "*Free*", exploits the functionalities of the *admittance controller* (see Section 4.2) set with a 0N force tracking to realize a feeling of transparency to the user's push. With this settings, each subject is asked to reproduce the trajectories performed during the preliminary adaptation without any assistance, while visual feedback is provided on screen. As a kinematic metric for task performance, the root mean errors of each of the samples of the performed trajectory with respect to the ideal one are measured. The second condition, called "*Assisted*", is build on the same trajectories but using the functionalities of the *tunnel controller* (see Section 4.2) with an assistance width set to 10mm.

The acquisition protocol includes a comprehensive variety of movement trajectories, based on standard radial paradigms (Aisen et al., 1997) and curvilinear trajectories. While radial paths are a standard for this set-up, curvilinear tasks are only marginally considered in the literature of robot-assisted planar movements. However, implementing such trajectories enables to elicit a wide variety of upper-limb tasks to promote challenge and the recruitment of the synergies available to people (Frère and Hug, 2012). Targets are oriented towards the main cardinal directions (NE, E, SE, S, SW, W, NW, N) in a circumference with a radius of 170 mm (see *Clock game* in Section 4.2), comparable to previous studies (Tropea et al., 2013). The left side of Figure 5.13 can be used as reference for the robot modes, the proposed trajectories and the overall experimental setup.

Data analysis

Thanks to the collected experimental data, coming either from the robot, from the Vicon system or from the EMG sensors, three outcome measures can be computed (see Figure 5.14):

- Kinematics.
  The main metric for the assessment of the performance in the two conditions is the average root mean error found comparing the actual end-effector path with the ideal trajectory.

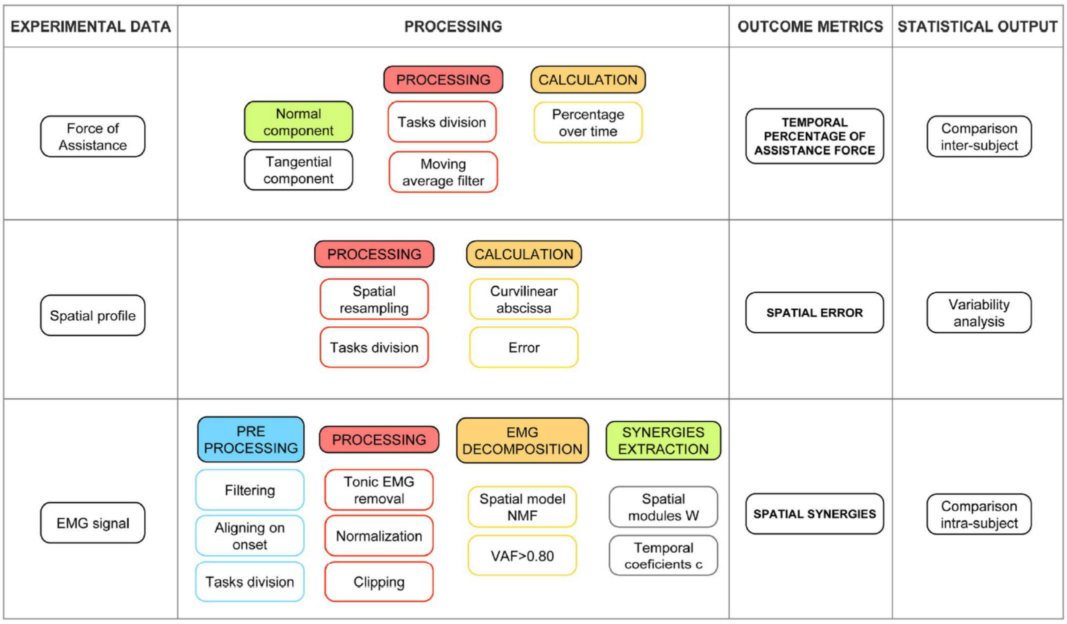


FIGURE 5.14: Experimental data processing and type of analysis for all measures of interest.

- Assistance.
  The amount of assistance provided by the robot is estimated as the percentage of time during which the corrective tunnel force is different from zero.

- Synergies.
  First of all, kinematic recordings are used to separate motion phases leveraging the velocity profile associated with the marker on the second metacarpal head of the dominant limb as a signal for detecting movement onsets and offsets. Then, all the movements are aligned by considering data 0.25 s before the task onset and 0.25 s after the task offset, to ensure the complete capture of EMG waveforms which begins before movement kinematic onset and ends after reaching the target (d'Avella et al., 2006). Since the nominal duration of the tasks is not fixed, a procedure for time-scaling and resampling is needed before the extraction of spatial synergies. Then, EMG data is band-pass filtered between 20–450 Hz to remove aliasing effects inside the sampling, rectified and low-pass filtered with a cut-off frequency of 6 Hz to extract the envelope. The obtained envelopes are further analyzed to extract the phasic component of the EMG, removing the postural (tonic) EMG activity from the original signal (Flanders, Tillery, and Soechting, 1992), following the approach used in previous works (d'Avella et al., 2006).

A normalization procedure was also performed in order to allow intra-subject comparisons. Then, the processed EMG envelopes are arranged to generate the pooled matrix data to be given as input to the NMF synergy extraction algorithm (Lee and Seung, 1999). For each subject, the extracted spatial synergies in Free and Assisted conditions are matched by similarity so that each synergy can be coupled with the most similar one found in the other dataset (García-Cossio et al., 2014). Then, each couple of matched synergies is assigned a similarity score computed as the cosine product between the vectors containing the synergy loads.

In all conditions, normality of the data is tested with the Shapiro-Wilk test. Since many distributions result to be not normal, the non-parametric Kruskal-Wallis Test is used.

Results

Starting from the kinematic outcomes, results show that error between the actual followed path and the ideal trajectory is significantly higher in the Free condition with respect to the Assisted condition for all three tested exercises. As expected, these results underline the effect of the robotic assistance during exercises, reducing the kinematic error, which is always maintained within the size of the so-called corrective tunnel. Moreover, as represented in Figure 5.15, participants generally make more mistakes in the Curvilinear trajectories with respect to the Radial ones.

Figure 5.16 provides a comparison on the levels of assistance elicited for the Radial and Curvilinear exercises. Interestingly, the assistance in the Curvilinear exercises is in general higher that the one required for the Radial ones. This result can be explained with a series of considerations. Radial trajectories are usually considered more intuituive by participants since they closely reseamble reaching movements that are often required for daily activities. On the contrary, Curvilinear trajectories represent a more challenging task which causes the subjects to follow the ideal path less precisely. As mentioned above, this results in higher kinematic error meaning that the corrective force provided by the tunnel controller needs to be activated more frequently.

Moving to MS, the mean number of extracted synergies was 6±1 regardless of the type of exercise (Radial or Curvilinear) and the exercise mode (Free or

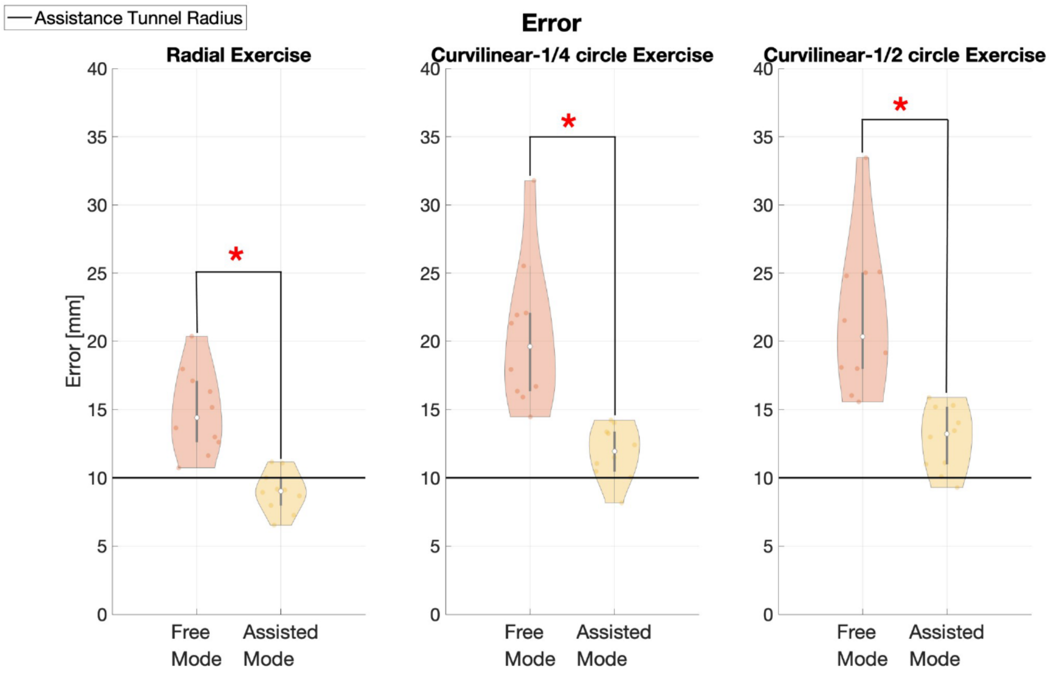


FIGURE 5.15: A graphical summary of the kinematic error for the three exercises both in Free and Assisted modes.

Assisted). On these basis, it was decided to consider 6 synergies for all participants so that comparison can be carried out more directly. An example of matching between paired synergies of the Free and Assisted datasets is reported in Figure 5.17. After that, the mean distributions of synergy matching scores for each subject can be quantified. Even though statistical differences can be found in this sense among subjects, the same cannot be said when comparing between the two types of exercises.

Overall, results are in accordance with previous findings, showing that assistance allows improving task performance (i.e., reducing errors in respect to ideal trajectories), but only alters the spatial structure of muscle synergies in a limited way. These findings support the employment of robotic devices for rehabilitation. In fact, the robotic intervention could, in principle, alter the motor control strategy adopted by the subject, acting as an external perturbation, despite achieving a reduced kinematic error. Verifying that robotic assistance is not disrupting the muscular synergies seems instead to suggest that assist-as-needed paradigms, where robot intervention comes into play only when error overpasses a certain threshold, succeed in letting the user experiment motor learning in a natural way.

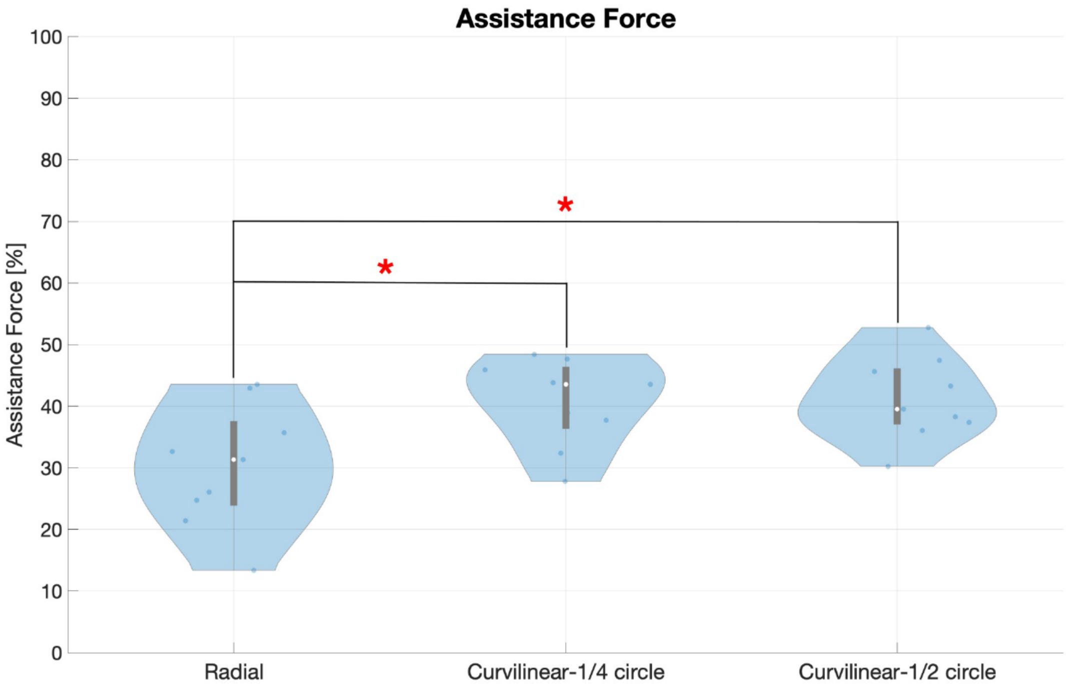


FIGURE 5.16: A comparison of the assistance levels elicited for Radial and Curvilinear exercises.

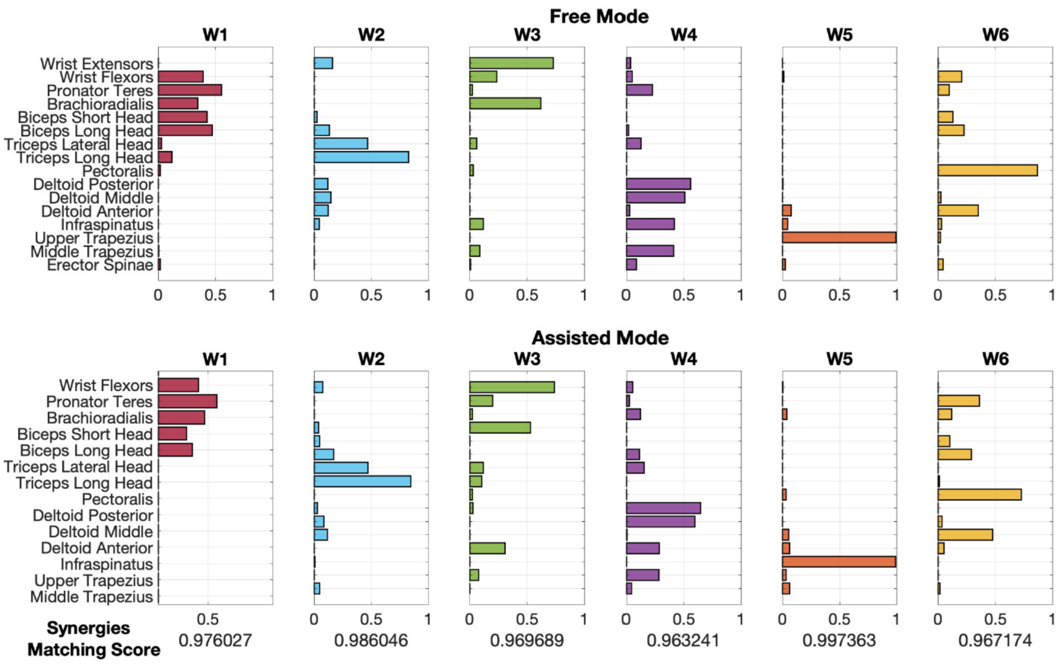


FIGURE 5.17: A typical example of paired synergies coupled by similarity following the extraction from Free and Assisted datasets for one of the subjects.

Biomechanical assessment and ergonomics.
Main take-aways:

The need for a thorough biomechanical assessment of the user in production environments is highlighted, both to mitigate physical fatigue and pain instances and to provide better task ergonomics. Overall, **substantial differences exist between studies carried out in labs and those performed in real working environments**. Therefore, a push towards the translation of results in industry is needed to provide actual benefits to all the involved stakeholders.

The use of non-invasive techniques is to be preferred over other methods, which are suitable only for lab-based controlled studies, both to minimize setup times and to provide a more natural interaction experience. **Marker-less systems, such as the Microsoft Azure Kinect cameras, are validated as a reliable option** since they can provide measures that correlate well with the gold standards if particular attention is posed to possible occlusions.

As a notable example, the Azure Kinect sensors are employed within a robotic collaborative assembly task and used as input for the generalized human-driven control architecture. This way, **user preferences and ergonomics can be leveraged simultaneously to personalize the robot collaborative pose in real-time**.

The importance of understanding the effect of robotics on the biomechanics of an interacting user is even more evident when considering a scenario of neuromotor rehabilitation. As a starting point, it is demonstrated that **the introduction of robotic assistance allows improving exercise performance, with only a limited alteration of the spatial structure of muscular synergies**.

5.2 Gaze behavior analysis and exploitation

One of the crucial aspects of designing a human-robot collaborative (HRC) production system is the tuning of the assigned workload since it can significantly impact the operator's well-being. For example, a high workload is associated with distress, high blood pressure, and other indicators of low well-being (Ilies, Dimotakis, and De Pater, 2010). On the other hand, boredom at work leads to distress and counterproductive work behavior (Hooff and Hooft, 2014). Both this scenarios are possible when working together with an automatic system which is intrinsically blind to how the operator subjectively perceives the workload throughout his/her shift. Due to these considerations, it is important to adapt the production rhythm to the level of productivity of the operators. Another aspect greatly impacting the well-being of operators is the experience of social isolation when working inside a robotic productive work cell where the usual human colleague is substituted by an automatic system. In non-industrial settings, for instance in hospitals or elderly care, studies show that specifically designed robotic solutions can be effective in reducing social isolation (Sarabia et al., 2018). Extending this concept to the industrial context, a cobot capable of interacting with the operator in a natural and social manner may be effective in reducing social isolation. To achieve both goals, human-robot collaboration strategies should be inspired by everyday human-human interactions, which rely on a variety of perceptual cues (Bull and Connelly, 1985; Argyle, Cook, and Cramer, 1994; Hadar et al., 1983). For instance, individuals instinctively direct their gaze towards their intended collaborators before initiating collaborative activities (Cary, 1978). If such behavior can be elicited during interactions with cobots, gaze direction can serve as a natural cue to communicate the intention to collaborate and therefore help in the automatic adaptation of the production rhythm.

The next sections address exactly these topics. First, the development of a vision model capable of distinguishing human gaze direction towards specific areas of interest is presented in Section 5.2.1. Then, in Section 5.2.2 the same model is applied to an HRI assembly scenario and evaluated in terms of objective performance and subjective participant satisfaction. Finally, Section 5.2.3 is dedicated to a second vision model for action recognition that can complement the information inferred though gaze behavior analysis to mitigate situations of uncertainty.

5.2.1 Gaze-based attention recognition for human-robot collaboration

Attention recognition is a key factor in improving human-robot collaboration. Previous studies (Saran et al., 2018; Tayibnapis, Choi, and Kwon, 2018; Huang and Mutlu, 2016) have proposed camera-based solutions to identify the area/object that has the user's attention. A gap in the validation of these systems stems from the guided gaze behaviors in the setup. The setup typically involves a stationary participant (sitting or standing) who is asked to gaze at a labeled area. This heavily reduces variations in viewing angle, head poses, etc. Even in their driver attention use case, where the user is expected to be seated, Ahlstrom, Kircher, and Kircher, 2013 point out that achieving "true distraction" in an artificial setting is difficult. Instead, leveraging the experimental setup presented in Section 4.1, an attention recognition model could be evaluated using videos from a human-robot collaboration task that resembles an industrial assembly. In fact, the literature highlights a preference for testing such models in a full-fledged setup than a well-controlled setting. Moreover, the chosen setup provides multiple opportunities for the cobot to adapt its behavior depending on the operator's attention, which can improve the collaboration experience and reduce psychological strain. Here, an overview of how the gaze-based attention detection model is realized and the related results on robustness and performance are reported, but the reader can refer to Prajod et al., 2023 for additional details.

Material and methods

In order to train and test the proposed gaze-based attention recognition model, the collaborative assembly scenario described in Section 4.1 is exploited. A simple pilot of the setup and of the assembly task is used to get a rough estimation of the duration of a single production cycle and of the time synchronization with the cobot. Generally, the complete cycle takes around 60-70 seconds with the operator finishing first and waiting for the cobot for around 10-15 seconds before tackling the collaborative assembly of the two sub-assemblies. The assembly is shared equally between the cobot and human operator, making this setup a level 3 (highest level) collaboration according to Christiernin's categorization (Christiernin, 2017). With the advent of Industry 4.0, this level of collaboration is becoming more common in the manufacturing process. Thus, this setup will allow for effective exploration on how the collaborative experience of the operator can be improved.

While piloting the setup, it was observed that there exist two key areas that the operator pays attention to for an extended period of time: the cobot and the work table. It is also possible that the operator gets distracted when working for long hours. So, three classes of attention are defined based on the gaze of the operator: attention on the cobot, attention on the table or distracted (looking in some other direction), as represented in Figure 5.18. It is a reasonable heuristic that if the operator's attention is on the table, they are working on the sub-assembly. Similarly, if they are looking at the robot, they are plausibly waiting for the robot to bring its sub-assembly. Now, two scenarios can take place. If the cobot assembles faster and visibly waits for the operator, the operator might feel pressured to speed up in order to synchronize with the cobot. So, if the operator is still assembling (attention to the table), the robot should wait inconspicuously or proceed to assemble the next part till the operator is ready. On the other hand, if the operator is faster and is waiting for the cobot (look at the cobot), then the cobot should increase its pace to avoid boredom. The ultimate goal is to enable the cobot to adapt its behavior in response to the operator's social and affective cues (e.g. gaze).

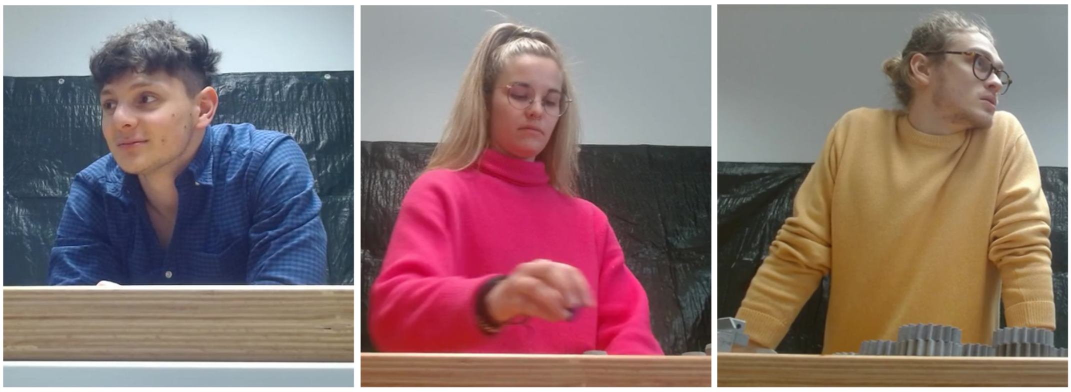


FIGURE 5.18: From left to right the three classes: attention to the cobot (while waiting), attention to the table (while assembling) or distracted (while waiting).

As a first step towards this goal, a gaze-based attention recognition model should be trained to detect the area the user is currently focusing on. To do that, a transfer learning approach is leveraged. Transfer learning involves using the parameters learned for task A to train a related task B. Since the attention classification proposed here is based on the gaze of the operator, gaze estimation is an ideal source task. For this purpose, the ETH-XGaze dataset (Zhang et al., 2020) is perfect as it contains over a million high-resolution images collected from 110 participants varying in gender, age, ethnicity, gaze angles, head poses and illumination. Using this dataset, a gaze estimation model is trained using the VGG16 (Simonyan and Zisserman, 2014) neural network architecture pre-trained on ImageNet (Russakovsky et al., 2015). Then, as mentioned, a transfer learning technique is exploited by re-using the weights learned by the gaze estimation model and freezing the convolutional layers, which are not modified anymore in the following training steps. Instead, the prediction layer is modified to classify the input image into three classes (cobot, table, and distracted). According to this classes, a second custom dataset (*Attention Areas dataset*) is produced. Always within the experimental setup presented in Section 4.1, a total of 8 participants are asked to stand in front of the Logitech camera and to look either towards the cobot, the work table, or anywhere else, with different configurations of their head orientation and gaze direction. For each of the three conditions, 30 pictures (1920×1080) per person are collected and labeled accounting for a total of 720 images. The Attention Areas dataset is then used for the final training of the model, making it possible to automatically map the gaze direction upon the predefined areas of interest.

Evaluation

In order to explore the robustness of the obtained model during human-robot collaborative tasks, a video dataset (*HRI Gaze dataset*) of participants working on the collaborative task described in Section 4.1.2 is collected and annotated. A total of 8 adult healthy participants (5F and 3M, age: 18-30 years) are asked to work as operators on the task for 3.5 hours a day, for 5 consecutive days, thus simulating the experience of a week of work. In order to obtain a naturalistic human-robot collaboration dataset, no guidance with respect to their gaze behavior is given to the participants. Three sessions of approximately 10 minutes each are recorded (1280×720, $25 fps$) during the first workday (beginning, middle, and end of the workday). Likewise, three additional videos are acquired during the last workday of the experiment. With this approach, one hour of videos for each participant, for a total of 8 hours of recording, are available to test and validate the gaze-based attention recognition model.

With reference to Figure 5.19, three main phases of the assembly task can be identified: gathering parts, independent assembly, and collaborative joining. Moreover, by design the operator has to wait between the end of the independent assembly and beginning of the the joint action with the cobot. During this waiting phase, it is observed that the operator either looks at the cobot

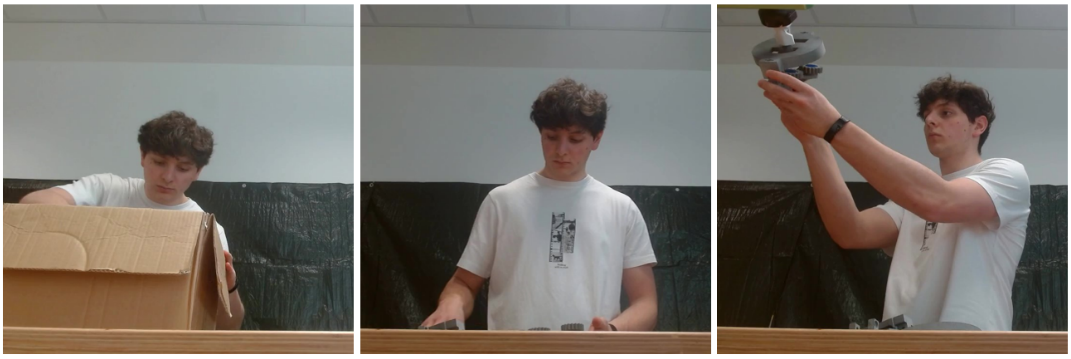


FIGURE 5.19: From left to right the three main task phases: the operator gathers the parts, independently works on his/her sub-assembly, and finally collaborates with the cobot to finish the product.

or in a random direction. With this distinction, the collected videos are annotated by indicating the activity the operator is doing. From the videos, sets of labeled images are extracted by removing the ones where face detection fails (blurry because of movement or covered by objects in the field of view) or where the eyes are not clearly visible, resulting in a test set with 833 images of attention to cobot, 940 images of attention on the table, and 962 images of distraction.

Results

Leveraging the labeled HRI Gaze dataset, the performance evaluation of the gaze-based attention recognition model yields a satisfactory result, with an F1 score of approximately 82%. With reference to the confusion matrix reported in Figure 5.20, it is clear that the main contributor to the performance loss is the "distraction" class, which is often misclassified as "attention to the work table". Further inspecting this aspect by manually checking the misclassified images, it is clear that during many instances of the waiting phase operators are looking towards the table because distracted by something in that direction and not because they are assembling parts. Examples of this behavior are reported in Figure 5.21.

To conclude, the performance of the developed model is satisfactory and robust for the tested experimental setup. However, it could be further improved by incorporating more informative data: the output of an additional action recognition model based on the dataset presented below in Section 5.2.3 could, for instance, mitigate instances of uncertainty.

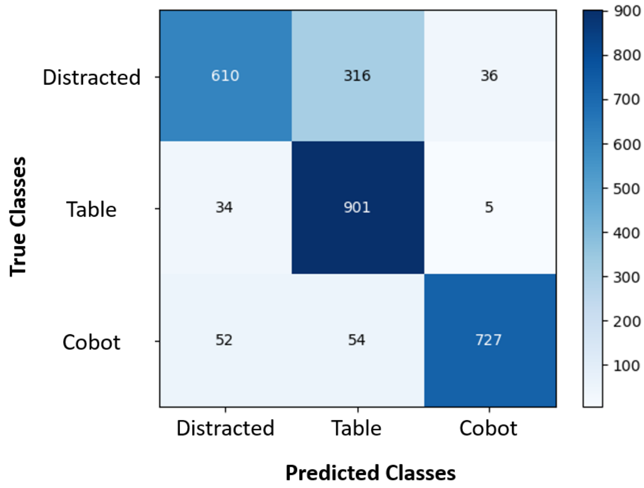


FIGURE 5.20: Confusion matrix for the evaluation of the gaze-based attention recognition model.

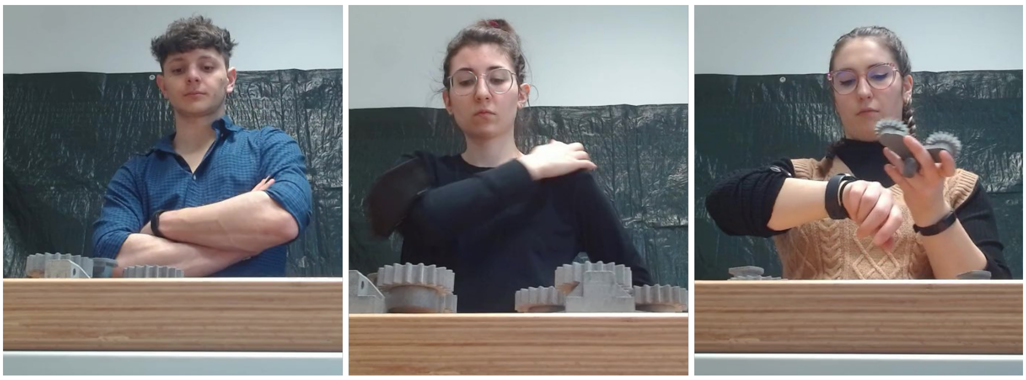


FIGURE 5.21: Examples of misclassified distraction images.

5.2.2 Gaze detection as a social cue to initiate natural human-robot collaboration in an assembly task

Leveraging HHI patterns to adapt the behavior of a collaborative robot holds promise to the reconstruction of social experiences akin to working with a human colleague within the working environment. As mentioned, gaze is a perceptual cue often present in this kind of scenarios: individuals instinctively direct their gaze towards their intended collaborators before initiating collaborative activities. Therefore, the gaze-based action recognition model developed and evaluated in Section 5.2.1, can serve as a starting point for an adaptive collaborative workcell capable of tuning the workload using natural interaction strategies designed to improve the operator's experience. Using the setup introduced in Section 4.1, two dedicated experimental activities are carried out. Here, a summary of the two experiments and their results is reported, but the reader can refer to Lavit Nicora et al., 2024 for further details.

Experiment 1

The goal of this first experiment is to analyze the natural behavior of users directly collaborating with a cobot on an assembly task and in particular to understand if gaze towards the cobot can serve as a natural cue to initiate joint action.

For this purpose, a total of 37 adult volunteers (29M and 8F) ranging from 18 to 48 years old (mean=29.03, SD=7.08) have been recruited. Each participant took part in a 15 minutes video-taped experimental session, carefully timed to ensure an adequate number of assembly cycles (approximately 15 to 20 complete products) enabling a comprehensive analysis of their recurring gaze behavior. With reference to the task presented in Section 4.1.2, each participant had to assemble Sub_B while the robot hovered with the detection camera over the pre-assembled Sub_A as if it was scanning for ready-to-pick sub-assemblies. As the volunteer's task got close to completion, a researcher acting as Wizard-of-Oz pressed a button on the laptop to trigger the robot. As a response, the robot smoothly interrupted the ongoing scanning motion, moved towards one of the sub-assemblies, picked it up, and brought it in front of the user at a convenient angle for the final joining. This iterative process continued throughout the 15-minutes experimental session, regardless of the number of completed gearboxes. To ensure a smooth workflow, ten pre-assembled sub-assemblies were initially placed on the cobot's table. The researcher restocked the sub-assemblies as necessary. Importantly, participants were unaware of the trigger given by the researcher to prevent any potential biases in their behavior during the interaction with the cobot. Also, the participants were informed of being filmed for ethical reasons but the aim of studying their gaze behavior was revealed only at the end of the experiment, again to avoid any possible bias.

The collected videos are given as input to the gaze-based attention recognition model presented in Section 5.2.1 in order to label the recordings on the basis of three classes: attention towards the cobot, attention towards the worktable or distraction. Additionally, manual annotation is required to select the frame where the cobot enters the camera field of view for the collaborative joining phase, considered the start of the joint activity for each assembly cycle. An example of the output of this procedure is reported in Figure 5.22.

Now, the gaze pattern for each participant is analyzed in two steps. First,

FIGURE 5.22: A snapshot showing the predictions from the attention recognition model (top track), and the annotated joint activity start points (bottom track, red lines).

the gazes towards the cobot within 15 seconds prior to the joint activity are computed (*pGazeJoint*). These 15 seconds are chosen because, after the WoZ trigger, the cobot takes around that time to move over the part, grab it, pick it up, and bring it to the collaborative joining position. This step helps to determine how often the joint activity is preceded by gazing towards the cobot, which therefore represents a cue to initiate the activity. Second, gazes towards the cobot outside the above-mentioned 15 seconds and outside the joint activity itself are also computed (*pUnexpectedGaze*). This step allows to make sure that the gaze pattern is prominent around the time of the joint activity, and not a frequent behavior irrespective of the activity.

Figure 5.23 visualizes the *pGazeJoint* and *pUnexpectedGaze* values from the 37 participants as box-plots. The mean *pGazeJoint* value is 83.74, i.e., on average, 83.74% of all collaborative joining instances were preceded by a gaze towards the cobot. Similarly, the mean *pUnexpectedGaze* is 9.67%, which implies that only few gazes at the cobot were outside the expected time frame. In other words, looking at the cobot occurs predominantly around the time of the collaborative joining activity.

These results indicate that people actually use gaze as a social cue preceding joint activity even when interacting with a cobot. Therefore, the second experimental activity can build on this outcome to implement and evaluate an

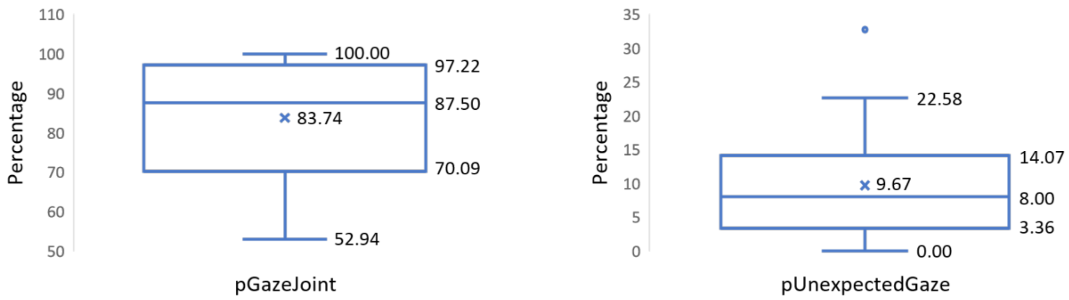


FIGURE 5.23: Box-plots computed from 37 participants representing *pGazeJoint* values on the left and *pUnexpectedGaze* values on the right.

automatic gaze-based trigger system substituting the role of the Wizard-of-Oz. Moreover, the data related to the unexpected gazes further supports the need for additional information to distinguish between uncertain scenarios. This need is preliminarily addressed in Section 5.2.3.

Experiment 2

The goal of this second experiment is to pilot the full integration of an augmented collaborative cell where joint action is automatically triggered on the basis of the detected gaze behavior of the user.

For this purpose, the same protocol described for Experiment 1 is used but, instead of having a Wizard triggering the joint action, the triggering process is automated thanks to the attention recognition model presented in Section 5.2.1. In practice, the robot automatically moves towards the participant to perform the joint action only if the latter looks towards the robot for longer than a threshold tuned to avoid slowing down the collaboration flow but also to avoid unwanted activations due to quick glances. The actual implementation of the automated system is based on the control architecture introduced in Section 3.2. Figure 5.24 highlights the modules that are involved for this specific experimental campaign. A total of 10 volunteers, with a balanced gender distribution (5M and 5F) and an age range between 18 and 30 (mean=23.8, SD=5.14), have been recruited for this second experiment. Again, none of the participants has prior experience with the robot and they are not told about the gaze-based automatic triggering system. In order to keep the experiment as short as possible but still make sure to collect enough experience samples, no fixed duration is set. Instead, each experimental condition lasts for the time required to assemble 10 complete gearboxes. At the end of the sessions, the participants are asked to report their impressions on the system and their responses are transcribed for post-analysis.

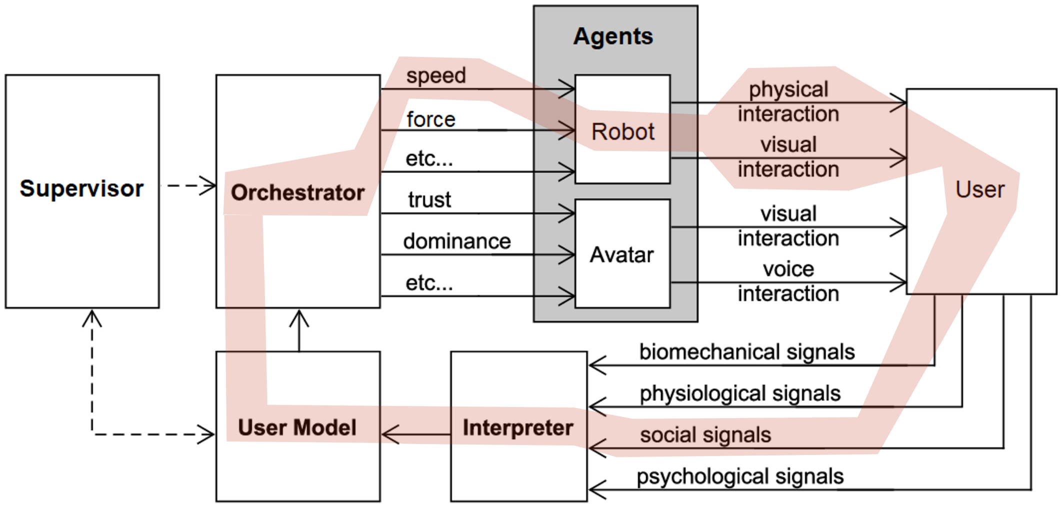


FIGURE 5.24: A highlight of the modules of the proposed control architecture involved in the automated gaze-based trigger system.

To evaluate the system, each iteration is considered "successful" if the participant is able to trigger the joint action at the expected moment (right before/after finishing his/her part) and within a reasonable time (maximum of 5s after finishing his/her sub-assembly, inspired by the threshold used by Eldardeer, Sandini, and Rea, 2020). Once again, manual annotation is required to select the frames corresponding to the moment when the participant is done with his/her part of the assembly and the moment when the cobot receives the trigger and starts moving towards its subassembly. Thanks to this annotation step, it is possible to compute the amount of time passed between these two instances for each participant and for each assembly cycle. As a result, the system achieves a success rate of 88.64%. Interestingly, for all the iterations that are not considered successful, the participants actually looked at the robot and triggered the joint action but did that after the 5s threshold set for the analysis. Also, as shown in Figure 5.25, it is important to notice that the system scores higher than what is observed during Experiment 1 (83.74% of joining instances preceded by a gaze towards the robot) meaning that the full integration of the system can be considered successful.

Moreover, some before-activations (i.e., the robot receiving the trigger before the end of the operator's assembly task) are observed. Overall, this situation occurs 19.21% of the times with an average anticipation time of 2.19s (left plot in Figure 5.26). A possible explanation for this result is that, over

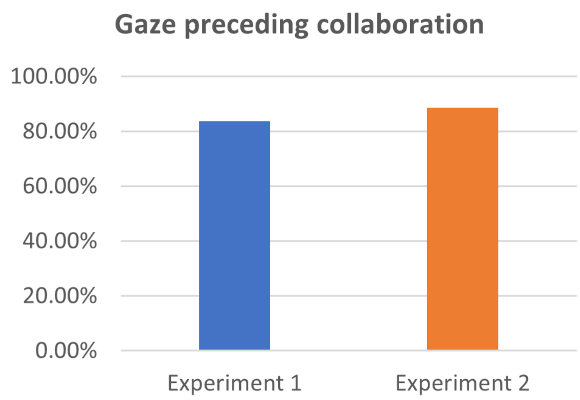


FIGURE 5.25: A comparison between the percentage of gazes preceding collaboration detected during the two experiments.

time, some of the volunteers may have guessed the role of their gaze in the process and started looking towards the robot before finishing their part in order to reduce the waiting times. A comparison between the average percentage of before activations of the group of participants who, at the end of the experiment, stated that they understood the gaze-based mechanism (before activations: 43.06%) and the others (before activations: 2.86%) seems to confirm the hypothesis (right plot in Figure 5.26).

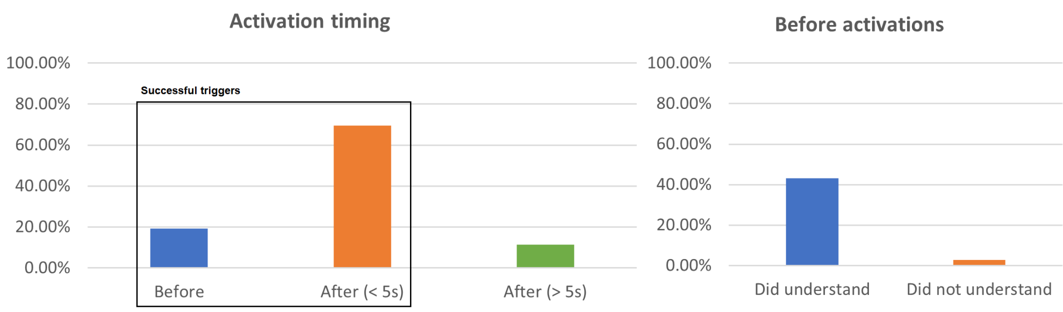


FIGURE 5.26: On the left, the distribution of timings for the detected gaze behaviors. On the right, the percentage of before activations clustered between those who understood the trigger system and those who didn't.

Following up on the results obtained with the first experiment, this second research activity confirms that the operator's gaze information can in fact be used as a natural cue to trigger joint action with a cobot. In general, most of the participants reported a pleasant and natural interaction experience, strengthening the hypothesis that the introduction of human-human interaction mechanisms within HRI scenarios can improve user experience.

5.2.3 A dataset on human-cobot collaboration for action recognition in manufacturing assembly

Sections 5.2.1 and 5.2.2 highlighted the need for additional information to clarify situations of uncertainty where the use of gaze is not sufficient to infer the actual state of the user and of the on-going task. In fact, in manufacturing processes human action recognition and segmentation are crucial for an effective human-robot collaboration. The accurate recognition and segmentation of the actions, including the timing of when the actions commence and conclude, is essential for the cobot to understand and interpret the intended actions of the human collaborator, to synchronize its actions, respond in real-time, and ensure smooth cooperation with the human collaborator (Cicirelli et al., 2015; Maselli et al., 2023). The information derived from skeleton joints enables researchers to capture temporal variations in body movements and offers flexibility in focusing on either the entire body or specific body parts allowing for a more comprehensive representation of the action being recognized and bypassing eventual privacy concern (Romeo et al., 2022). Literature is rich in RGB-D datasets for human action recognition (Lopes, Souza, and Pedrini, 2022) prevalently acquired in indoor/outdoor unconstrained settings. They are mostly related to daily actions, two-person interactions, or gaming actions. Few papers present assembly action datasets mostly acquired from the worker perspective (Ragusa et al., 2021; Sener et al., 2022). Moreover, few vision-based datasets exist on human-cobot cooperation for object assembly in industrial manufacturing and they all have some limitation (e.g., low level of collaboration, no depth information). For this reasons, the MindBot project consortium decided to put together a new and more complete dataset, called HARMA, as foundation for developing and testing advanced action recognition/segmentation systems in the context of HRI. Here, the most relevant information on how the dataset is built and tested is reported, but the reader can refer to Romeo et al., 2024 for additional details.

Dataset acquisition

The HARMA dataset is built leveraging on the setup presented in Section 4.1.1 and the collaborative task explained in Section 4.1.2. A total of 27 participants have been recruited and recorded by the two Kinect cameras present in the setup while working on the assembly (see Figure 5.27). Each subject performs the task multiple times, resulting in a total of 240 task executions in the dataset. By design, the product to be assembled and its components offer some flexibility to the user in the order of assembly and in the manipulation strategy (e.g., using both hands or only the dominant hand). Therefore, in order to build a dataset representative of the variability in the way an operator may approach the task, the participants are instructed to proceed with the assembly according to their preference by freely choosing how to perform each step.

FIGURE 5.27: A participant working on the task as seen by the two Kinect Azure cameras.

After the acquisition, manual data annotation is required to segment and label each action in the collected videos. The start frame of each action is determined when the subject begins to move the arm toward the component to be grasped. The end frame, instead, is set when the subject releases the component. As a result, total of 2885 actions have been annotated, including the “don’t care” action (annotated 245 times overall) to classify eventual pauses between action transitions or to unexpected events such as the loss of a component during the assembly.

Dataset analysis and validation

The first analysis that can be done on the resulting HARMA dataset is an exploration of its temporal characteristics. Videos are recorded at 30 frames per second (fps) and each action has a different duration in terms of number of frames. Moreover, subjects perform the task at their own comfortable and self-selected speed, so there is a high temporal variance between different subjects, as clearly visible in Figure 5.28.

An additional analysis of the spatial movement of skeletal joints during the execution of the actions can be helpful in getting information about the main direction and spatial displacement of the actions. As a reference, both wrists

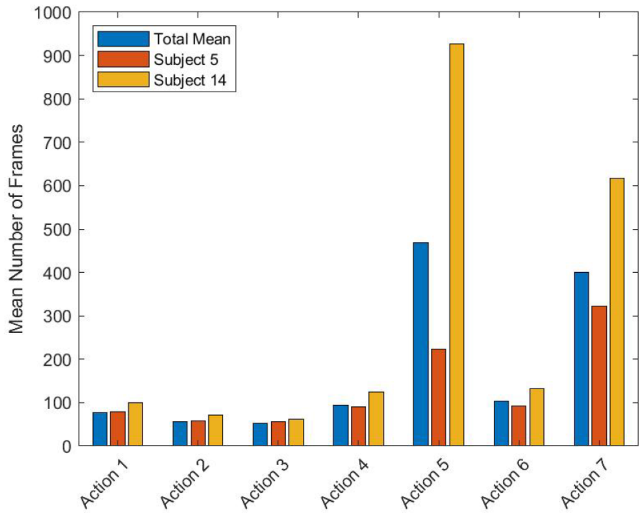


FIGURE 5.28: A comparison between the duration of each action as performed by two different subjects.

of all operators are considered as they are mainly involved in the action execution. Figure 5.29 shows the standard deviation of the coordinates (X, Y, Z) of the right and left wrist joints for each action in all videos. As represented, each action has its own spatial characteristics. However, it must be considered that this analysis may be influenced by how the operators performed the task since no precise rules were imposed in order to achieve maximum variability of the dataset.

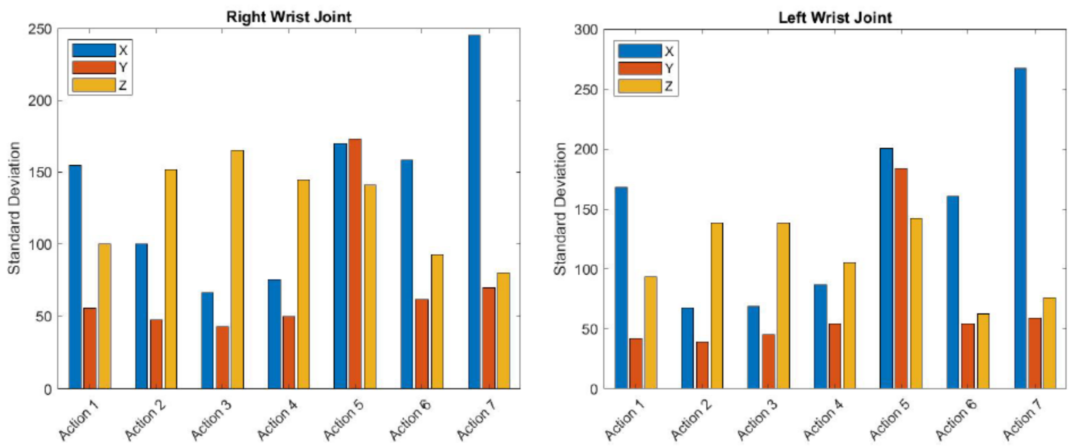


FIGURE 5.29: A comparison between the duration of each action as performed by two different subjects.

A final validation is provided through a practical application of the dataset, by testing action recognition and segmentation methods on it. In particular, the well-known ASFormer model (Yi, Wen, and Jiang, 2021) is used as reference since it is one of the first transformer-based architectures for temporal action segmentation. RGB and skeletal data are analyzed for three cases: data acquired from the frontal and lateral cameras, respectively, and combining the data from both cameras. The HARMA dataset is split into non-overlapping training and testing sets by considering the 70% of videos for training and the remaining 30% for testing ensuring that videos of the same operator do not appear in both training and testing sets. Figure 5.30 shows the best obtained result, with 97.44% (RGB) and 95.95% (skeleton) accuracy rates.

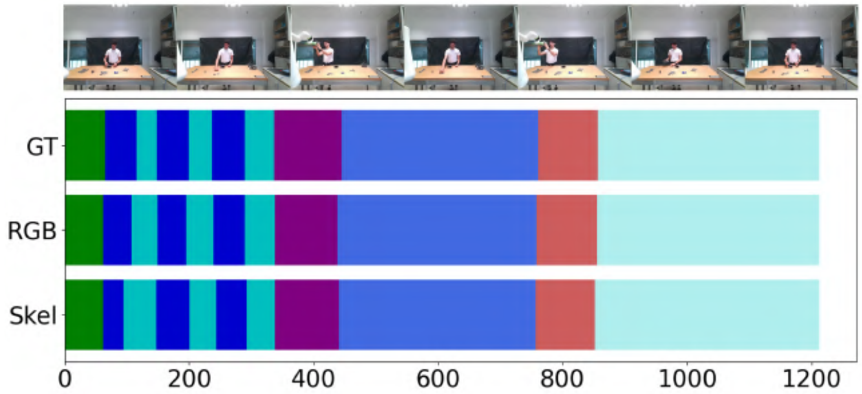


FIGURE 5.30: The best segmentation result obtained for the tested videos of the dataset.

High-performance rates are obtained when using RGB features while Performance worsen in the case of skeletal data. In general, RGB data provides rich visual information about the scene but typically requires higher storage space and computational complexity compared to skeleton-based data representation. On the other hand, by using skeleton data is possible to abstract away detailed appearance information and focus solely on the spatial configuration of body joints and movements. Therefore, it's essential to carefully find a good trade-off and select the data modality that best aligns with the goals and constraints of the working context.

Gaze behavior analysis and exploitation.
Main take-aways:

The reconstruction of behavioral patterns typical of human-human interactions is proposed as a promising approach to provide positive HRI experiences. For instance, **gaze can serve as a social cue to communicate the intention to collaborate**, useful to promote a natural interaction and a personalized task speed.

First, **a gaze-based attention recognition model is trained using transfer learning techniques**. The model achieves a satisfactory 82% F1 score, but some sources of uncertainty are also identified. As a result, this new tool is robust for the considered experimental setup, but it could also benefit from additional information regarding the on-going actions.

Then, the developed model is introduced in the generalized human-driven control architecture and leveraged for exploratory research activities. A first analysis reveals that around 84% of the collaborative instances are preceded by a gaze towards the robot, meaning that **people actually use gaze as a social cue even when interacting with machines**. Once again, some situations of uncertainty are identified, strengthening the need for the integration of action-related information. Secondly, **gaze-based attention recognition is proven to be successful for the automatic triggering of robot collaboration**, achieving a success rate of around 89%.

In order to complete the information that can be inferred through gaze behavior analysis, **a new dataset is produced as foundation for developing and testing action recognition/segmentation systems**. The HARMA dataset is therefore collected, validated and made available to the community for future advancements in the field.

5.3 Psychology-in-the-loop: flow and locus of control

Up to now, the exploitation of biomechanical and social signals have been explored and successfully introduced in the generalized human-driven architecture. However, in order to gain a complete and heterogeneous description of the user's experience of interaction, psychological measures cannot be neglected. With reference to Section 3.1, these kind of signals can be heavily influenced by the mentioned social filter and it is therefore important to include strategies of psychological inference based on quantitative and objective data as well as subjective measures. Regarding the latter, data collection approaches such as questionnaires and interviews cannot, by nature, be collected in real-time and leveraged for the adaptation of the behavior of the robot as the experience of interaction unfolds. However, they still retain great importance as they can be used for the validation of objectively inferred measures and for a posteriori evaluation of the acceptability of the system. With this goal in mind, the following sections attempt to address these topics by focusing specifically on the well-established concepts of Flow and Locus of Control.

5.3.1 Flow in human-robot collaboration: multimodal analysis and perceived challenge detection in industrial scenarios

The concept of Flow is often described as a state of optimal experience. It is characterized by high levels of engagement, motivation, a sense of control, and complete immersion in an activity (Csikszentmihalyi, 2000). This state emerges when the challenges presented by the task match the individual's skills and abilities. While extensive research (Nah et al., 2014; Stamatelopoulou et al., 2018; Santos et al., 2018; Pearce, 2005) has been conducted on the concept of Flow across various domains, such as sports, education, and gaming, its application in industrial settings remains relatively unexplored. Considering the significance of Flow in optimizing performance and well-being at work (Csikszentmihalhi, 2020; Csikszentmihalyi and LeFevre, 1989), it is imperative to bridge this research gap and explore the Flow experience in industrial environments (Fullagar, Delle Fave, and Van Krevelen, 2018; Beyrodt et al., 2023).

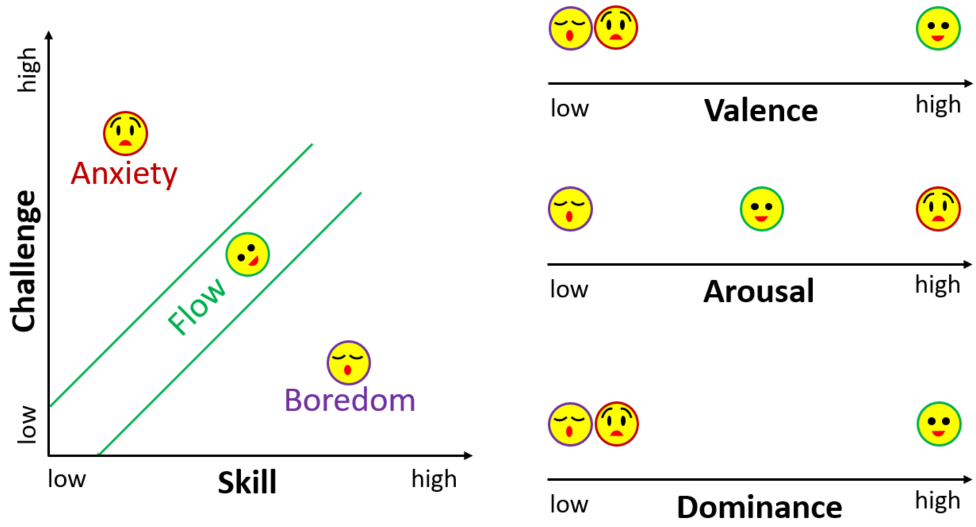


FIGURE 5.31: A simplified model of Flow mapped over challenges and skills (on the left) or over the dimensions of valence, arousal and dominance (on the right).

Assembly tasks in industrial settings typically involve repetitive and fixed procedures. As a result, workers gradually acquire the necessary skills to proficiently perform the task, leading to minimal variations in their individual skill levels over time. In such cases, the primary factor that influences the experience of Flow becomes the perceived level of challenge presented by the task itself (see Figure 5.31). This unique aspect of industrial tasks has led the project consortium to investigate how different perceived challenge levels evoke distinct user responses. Recognizing that Flow emerges when there is a balance between perceived challenge and skill, the goal is to adapt the task by adjusting the challenge level to facilitate Flow among cobot workers. For this purpose, the MindBot setup presented in Section 4.1 can be leveraged. In practice, by adjusting the production rate of the cobot, it is possible to obtain three distinct levels of challenge corresponding to the three commonly studied states in Flow research (Boredom, Flow, and Anxiety). Here, a summary of the experimental campaign and of the most relevant outcomes is reported, but the reader can refer to Prajod et al., 2024 for additional information.

Materials and methods

Once again, the workcell described in Section 4.1.1 and the collaborative assembly task presented in Section 4.1.2 are used. In order to have more flexibility in the production rate of the cobot, participants are asked to work on Sub_B while a number of pre-assembled copies of Sub_A are placed on the cobot table, as if they were produced by a secondary assembly line (not reproduced in this lab-based scenario). At a certain point during each production cycle, the cobot brings a pre-assembled part to the participant and holds it in a convenient position for the final joint activity of the production cycle (gears meshing). By simply tuning the time it takes for the robot to bring Sub_A to the participants, it is possible to change the overall production rate. With this approach, three distinct experimental conditions can be realized based on the production rate of the participant and the cobot:

1. *Slow condition*: The cobot performs a scanning motion over all the sub-assemblies using the camera on its wrist before picking one of them up and bringing it to the user. Overall, it takes around 55 seconds from the start of each production cycle for the cobot to get to the participant for the joint activity. The Slow condition represents a low level of challenge for the participants since they have plenty of time to finish their part of the assembly before the cobot comes for the joint activity. This leads to the participant waiting for the cobot and plausible experience of Boredom.

2. *Fast condition*: The cobot does not perform any scanning motion, it moves straight to the next sub-assembly to pick it up and bring it to the participant. Overall, it takes around 15 seconds from the start of each production cycle for the robot to get to the participant for the joint activity. The Fast condition is expected to be perceived by the participants as a high level of challenge since they do not have enough time to assemble before the arrival of the cobot. This leads to the cobot waiting for the participant and could elicit Anxiety in the participants.

3. *Adaptive condition*: The cobot performs the previously mentioned scanning motion until a researcher, acting as a Wizard of Oz, triggers it to bring one sub-assembly to the participant. In this case, there is no fixed timing for the cobot since the wizard triggers the cobot whenever the participant is close to finishing his/her part of the assembly. The Adaptive condition is designed to be the optimal level of challenge since the production rate of the cobot is tuned according to the participant's performance.

Three types of data are collected to evaluate how the participants respond to the administered experimental conditions. Upper-body videos are collected using the Logitech C920 Pro HD webcam placed in front of the participant (1920 x 1080, $25fps$). Additionally, ECG data is collected at $130Hz$ by asking participants to wear a Polar H10 chest band. Finally, the NASA-TLX (Hart and Staveland, 1988) questionnaire is administered on paper at the end of each session.

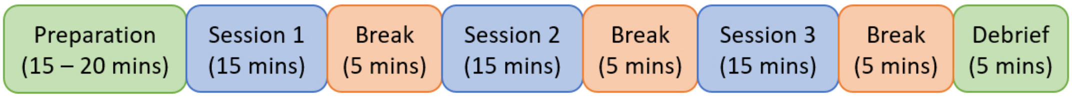


FIGURE 5.32: The steps making up the experimental protocol.

A total of 37 adult volunteers (8F and 29M) aged 18-48 years (mean = 29.03, SD = 7.08) have been recruited for the study. The within-subjects protocol design represented in Figure 5.32 is chosen. Every participant is administered all three experimental conditions with 5 minutes of break between consecutive sessions. Each condition lasts 15 minutes during which the participant keeps assembling gearboxes one after the other. The order in which the three conditions are administered is chosen randomly, in order to average out any side effect that may be caused by the sequence. During the break, the participants fill out the NASA-TLX questionnaire about the task load and experience pertaining to the completed session.

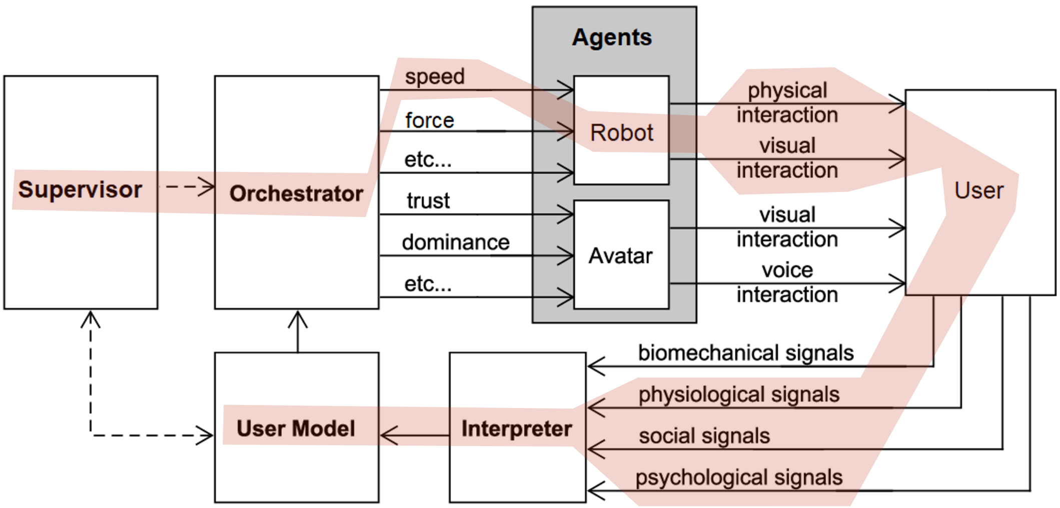


FIGURE 5.33: A highlight of the modules of the generalized human-driven architecture in play for this experimental activity.

The collected videos are used for emotion estimation purposes using a dedicated deep-learning model trained on the widely used AffectNet dataset (Mollahosseini, Hassani, and Mahoor, 2019). The face region is identified and cropped using MediaPipe Blaze face detection model (Bazarevsky et al., 2019). This represents the input for the trained model, capable of classifying images into seven discrete emotion classes: Neutral, Happy, Sad, Surprise, Fear, Disgust, and Anger, together with the two continuous values of Valence and Arousal. A point in favor of the reliability of this model is provided by Prajod, Huber, and André, 2022 and Prajod et al., 2021, where the authors use explainable AI heat maps to represent the results of the training: the model actually learns the so-called "facial action units" which are the same features used in systematic evaluation of facial expressions by human observers. That being said, there is definitely a need for more rigorous validation, currently hindered by the lack of a standard data labeling approach, which limits the level of reliability of the obtained results. For this reason, data obtained from the emotion estimation module will only be used to complement other sources of information with the goal of gaining a more comprehensive view of the status of the user.

Additionally, ECG data is pre-processed to obtain a cleaner signal and then leveraged to extract HRV features. All of these procedures have been carried out thanks to the use of SSI, which can be directly interfaced with the ROS network, as explained in Section 4.1.3. As a result, the whole system is built within the generalized human-driven architecture proposed in Section 3.2. Figure 5.33 highlights the specific modules of the architecture in play for this experimental campaign.

Analysis of collected data

A first analysis can be done on the responses of the participants to the NASA-TLX questionnaire. The mean response values (on a 20-points scale) for each condition are reported in Table 5.5.

TABLE 5.5: The average responses to the NASA-TLX questionnaires after each condition

Category	Slow	Fast	Adaptive
Mental demand	4.81	**6.35**	4.95
Physical demand	4.59	**6.84**	5.05
Temporal demand	4.73	**10.54**	6.08
Effort	5.16	**7.76**	5.65
Performance	**7.30**	6.81	6.81
Frustration	**5.35**	**5.35**	4.27

The Fast condition resulted in the highest Effort, Mental, Physical, and Temporal demands. The Slow condition scored lowest in these categories. If such a difference is expected in Temporal and Physical demands due to the design of the experimental conditions, it is interesting to see that the cobot production rate affected other categories of task load. Moreover, although the number of assemblies was the highest in the Fast condition and lowest in the Slow condition, the perceived Performance was highest for the Slow condition. Another notable observation is that the participants experienced lower frustration in the Adaptive condition.

Moving to the analysis of the collected emotion indicators, primary focus is given to the valence and arousal indexes, as continuous values provide a more dynamic estimation of emotions. Averaging over all participants, the mean valence levels (Slow: -0.025, Fast: -0.018, Adaptive: -0.023) are lowest for the Slow condition and highest for the Fast condition. The mean arousal values (Slow: 0.053, Fast: 0.074, Adaptive: 0.071) also follow a similar trend. However, from a statistical point of view, no significant difference in mean valence between conditions is detected. On the other hand, the Slow condition differed significantly in arousal from both the Fast ($p = 0.012$) and the Adaptive ($p = 0.015$) conditions, while no evidence of a significant difference in mean arousal between the Fast and the Adaptive conditions ($p = 0.884$) is found. However, the mean arousal values in all three conditions are in the range [0, 0.1]. These values are typically associated with a neutral emotional state, meaning that facial expressions are not a good indicator of the perceived challenge level.

Lastly, the extracted HR and HRV features are analyzed. To mitigate for individual differences, data is normalized and averaged to obtain the plot reported in Figure 5.34. In line with the trends of emotion estimation, the HR appears to be highest in the Fast condition (mean = 0.554), followed by the Adaptive condition (mean = 0.485), and the lowest in the Slow condition (mean = 0.402). A significant difference between the Slow condition and the other two conditions (Fast $p < 0.001$, Adaptive $p = 0.038$) is found, while only a trend-level difference between average heart rates of Fast and Adaptive conditions ($p = 0.056$) is detected. Beyrodt et al., 2023 observed that when a cobot is faster than the user, s/he tends to reappraise the situation and, over time, starts working at his/her own pace. This could be a plausible reason for the lack of significant difference obtained between the Fast and Adaptive conditions.

Overall, the analysis shows trends similar to the observations of other studies in the literature. As seen in Figure 5.34, HR increases with challenge level and HRV decreases with challenge level. The Adaptive condition resulted in

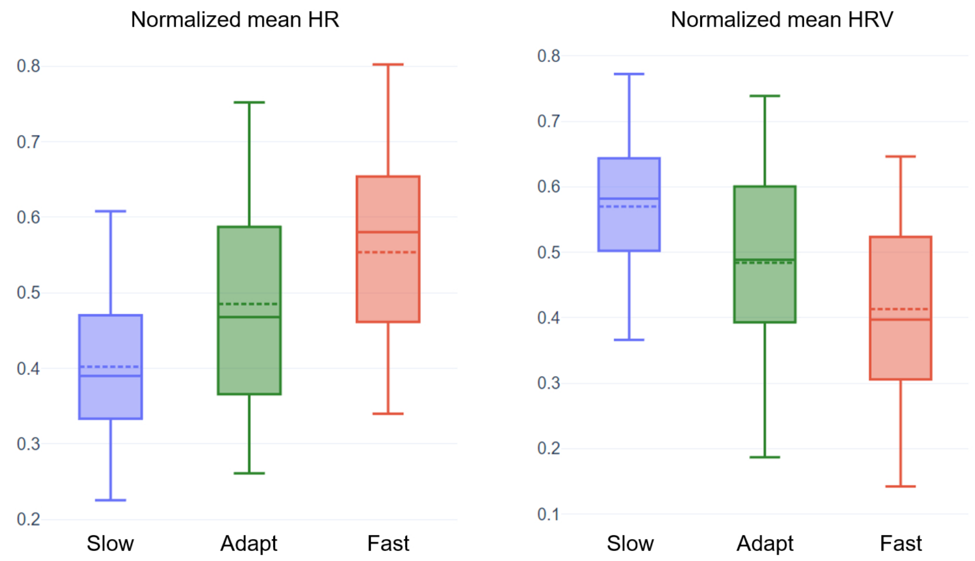


FIGURE 5.34: Box plots of normalized mean HR (left) and HRV (right) for the three experimental conditions.

a relatively moderate HR and HRV, which is expected in a challenge-skill balanced condition. Hence, HRV features could be good indicators of perceived challenge levels during human-robot collaboration tasks.

Challenge prediction

The significance test results of heart rate variability features are promising. On these basis, it was decided to train a simple feed-forward neural network to predict the challenge level experienced by the participants. If such a model can be effectively trained and validated, it could represent the basis for an automatically adaptive system on the lines of the gaze-based one described in Section 5.2.2. In fact, Flow is usually evaluated through questionnaires and therefore not suited for real-time tuning of automated systems. By evaluating and predicting the level of challenge as a direct translation of Flow experiences, it could be possible to introduce this psychological measure within the generalized human-driven control architecture.

Since the previous analyses repeatedly showed that no significant difference can be found between the Fast and Adaptive conditions, it was decided to train the model for a binary classification, discerning between the Slow condition and the others only on the basis of real-time HRV data. With this approach, the trained model achieves 0.707 in accuracy with an F1-score of 0.661. Although this result may not be notably high for a binary classifier, the model performs comparably to other flow detection models in literature.

As a final note, HRV seems to be a promising indicator of the level of experienced challenge and the consequent state of flow of the participant. However, the limited number of predictable classes and the corresponding obtained accuracy does not allow for an effective standalone implementation in the proposed architecture. A possible benefit of this result could be in the form of validation of other available measures. For instance, the well-performing gaze-based attention recognition model presented in Sections 5.2.1 and 5.2.2 could be expanded with some rule-based instructions. In fact, validation of the mentioned model highlighted difficulties in discerning real states of distraction. A dataset for action recognition and segmentation (see Section 5.2.3) was produced to mitigate this issue. With the same goal, this HRV-based challenge recognition model could enhance the overall capabilities of the system. An instance of distraction is often connected to a general state of boredom for the user, and boredom is usually a consequence of low challenge tasks. The ability to detect experiences of low challenge together with some action recognition data are therefore considered promising in making the system even more robust.

5.3.2 Synchronizing minds and machines: insights into cognitive and emotional factors in human-robot collaboration

Aligning with the goals of Industry 5.0, a relevant topic is the identification of which robot characteristics influence the emotional state and other psychological variables of the subjects who interact with it.

Section 5.3.1 already highlighted the need to consider the human emotional state in response to robot actions and features. There, the effect of different production rates was analyzed, inferring the user's emotional state mainly through objective physiological measures. However, subjective measures, collected through questionnaires, may offer further insights shedding light on how these control parameters influence the real emotional experience of a user interacting with the robotic device.

Additionally, among the many psychological variables in play, the so-called Locus of Control (Loc) is of particular interest for the present project. The LoC is the degree to which people believe they have control over events in their lives, rather than being influenced by external forces (Rotter, 1966). It is a one-dimensional construct characterized by two poles, internal and external, placed on the extremities of a continuum. People's attitudes are arranged along this continuum depending on how they attribute the cause of what happens to them. Individuals with an internal Locus of Control believe events are primarily a result of their actions (e.g., work performance depends almost entirely on their commitment and abilities). In contrast, those with an external LoC attribute events to external factors (e.g., work performance depends on external factors, including chance). Despite many studies reporting that an internal LoC is often associated with good physical health (Gale, Batty, and Deary, 2008; Arraras et al., 2002; Cobb-Clark, Kassenboehmer, and Schurer, 2014; Kesavayuth, Poyago-Theotoky, Zikos, et al., 2020), it would seem that even more important is the flexibility with which a person can adapt their thinking (external-internal) depending on the specific situational needs (Cheng et al., 2013). Regarding robotics, some authors have highlighted that people with an internal LoC have worse usage performance, as they struggle to leave control of the situation to an autonomous system (Takayama et al., 2011; Acharya et al., 2018). Personal experiences, including work experiences, can influence the Locus of Control. In particular, a change of LoC in a specific scenario when interacting/using a product, i.e., a cobot, is referred to as the Experiential Locus of Control (ELoC) (Jang et al., 2016). This concept is an extension of the classic Locus of Control construct and refers to the effect that one experience has on the Locus of Control relative to that specific experience.

To better understand these dynamics, a dedicated study was carried out with the data collected during the experimental campaign described in Section 5.3.1. Here, the main outcomes are reported, but the reader can refer to Mondellini et al., 2024 for additional details.

Materials and methods

The setup and assembly task presented in Sections 4.1.1 and 4.1.2, respectively, are leveraged once again. As mentioned, the data needed for this particular study has been collected concurrently with the experimental campaign described in Section 5.3.1. Therefore the same protocol, foreseeing three experimental conditions with three different levels of production rate (Slow, Fast and Adaptive), is followed (see Figure 5.32). Before the start of the interactive experience, each participant was administered the following questionnaires as baseline measure of their attitudes:

- *Internal Control Index* (Duttweiler, 1984) (ICI). This questionnaire, administered before the interaction with the cobot, consists of 28 items. For each item the participant provides his/her response on a 5-point Likert scale, where 1 corresponds to "rarely" and 5 to "usually". A high score (maximum 140) corresponds to a high internal Locus of Control level. The score can vary from 28 to 140.
- *Negative Attitudes Towards Robots Scale* (Nomura et al., 2006b). This psychometric scale measures negative attitudes towards robots through 14 items divided into three subscales. S1 relates to "negative attitudes toward situations of interaction with robots" (six items), S2 pertains to "negative attitudes toward the social influence of robots" (five items), and S3 addresses "negative attitudes toward emotions in interaction with robots" (three items). Each item is rated on a scale from 1 to 5 (1: strongly disagree - 5: strongly agree). The score can vary from 6 to 30 in S1, 5 to 25 in S2, and 3 to 25 in S3.
- *Perceived Stress Questionnaire* (PSS). This tool is designed to assess the degree to which individuals perceive their lives as unpredictable, uncontrollable, and overloaded—key components of the stress experience. The PSS is particularly effective for understanding how individuals appraise stress with daily life challenges rather than specific events. It includes 10 items (Cohen, Kamarck, and Mermelstein, 1983).

Moreover, between each condition, the number of completed assemblies is annotated as an indicator of the participant's performance together with the following subjective measures:

- *Experiental Locus of Control Questionnaire* (ELoC), This questionnaire is composed of 3 items from the "Internal Control Index" (Duttweiler, 1984), appropriately modified to evaluate the Experiential Locus of Control (Jang et al., 2016). The score can vary from 3 to 15.
- *Self-Assessment Manikin* (Bradley and Lang, 1994). It consists of a non-verbal pictorial assessment technique that directly measures Valence, Arousal, and Dominance associated with a person's emotive reaction to several stimuli. The participant responds on a 9-point Likert scale for each subscale (score 1-9). The Valence scale ranges from positive emotion (happy) to negative emotion (sad). On the Arousal scale, the score varies from high excitement to calmness. Low scores on the Dominance scale correspond to low control, and vice versa.

- *NASA Task Load Index* (NASA-TLX). This multidimensional assessment tool evaluates six key dimensions of workload: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration. Participants rate each dimension on a 20-point scale, capturing the subjective experience of task load and stress (Hart and Staveland, 1988). NASA-TLX has been extensively validated and is commonly applied in ergonomics and human factors research to evaluate task difficulty and workload in various settings, particularly those involving human-machine interactions;
- *Short Stress State Questionnaire* (SSSQ). This questionnaire is a concise self-report tool designed to assess immediate stress states in performance settings. It measures three distinct dimensions of stress: Distress, Engagement, and Worry. Each dimension captures specific psychological responses to stress: Distress reflects negative emotions and perceptions, Engagement indicates motivation and positive involvement, and Worry measures cognitive interference related to concerns or apprehensions (Helton, 2004);
- *International Positive and Negative Affect Schedule short-form* (I-PANAS-SF). This tool assesses the individual's positive and negative affect states. Respondents rate their experience of 5 positive and 5 negative emotions over a specified period, providing insight into their general affective state. This shorter version, derived from a 20-items PANAS version (Watson, Clark, and Tellegen, 1988), allows for a quicker yet reliable measurement of affect without sacrificing psychometric robustness (Thompson, 2007).

Analysis and results

Thanks to the rich dataset of collected subjective measures, a number of analyses can be carried out. First of all, two-way Friedman ANOVA tests are conduced to assess differences across the experimental conditions.

As expected, significant differences are found in performance (i.e., the number of assembled products) with the highest result for the Fast condition (median = 18, IQR = 4) and the lowest for the Slow condition (median = 13, IQR = 2). In line with this result, significant differences are also found in the perceived workload with the Fast condition leading to scores higher than both the Slow and Adaptive conditions ($p < .001$). Interestingly, positive affects in the Slow condition are statistically significantly lower than both the Fast and Adaptive conditions ($p < .001$), while no difference is found in terms of negative affects. These results are summarized by the box plots depicted in Figure 5.35.

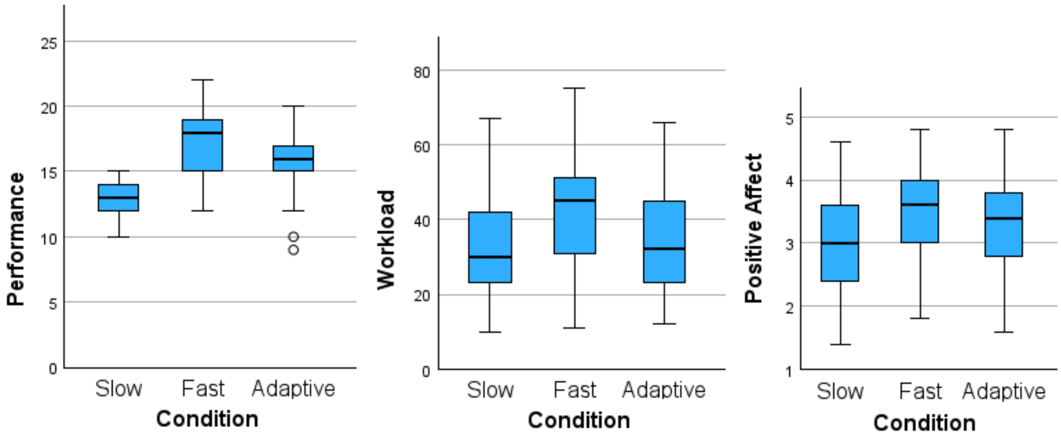


FIGURE 5.35: Box plots of performance (left), workload (center) and positive affect (right) for the three experimental conditions.

Overall it can be said that, in the proposed scenario, the production rate set for the cobot does not significantly influence users' emotional response and engagement levels. This result is different from what was found in Section 5.3.1. It is possible that, since the previous study mostly relied on objective physiological measures, other factors only detected by subjective questionnaires, such as individual differences, might play a more prominent role than expected. This highlights the importance of subjective variability in stress perception, which should be explored further in order to deploy effective personalized systems. Additionally, the statistical differences found for both performance and workload seems to be connected between each other. In fact, when participants have to be faster, more focused, and make fewer mistakes to keep up with the cobot, they perceive the condition as physically and cognitively more demanding. Again, subjective variability in stress perception may explain why this result does not reflect also in the scores related to the level of stress or in the NASA-TLX subscales. Finally, the statistically lower positive affect detected for the Slow condition highlights that such scenario leads participants to experience more frustration or less enjoyment, potentially due to the longer wait times and lower engagement in the task. In contrast, both the Fast and Adaptive conditions may foster higher levels of positive emotions due to the task's faster pacing or more engaging nature.

After this comparison between conditions, a detailed and complete statistical analysis has been performed. Related descriptives are reported in Tables from 5.6 to 5.10.

TABLE 5.6: Internal Reliability (Omega), Items deleted from the scale to enhance reliability, Kolmogorov-Smirnov (K-S) test for the distribution of scores, Median, Minimum and Maximum Scores and Interquartile Range

	Ω	**Deleted Items**	**K-S test**	**Median (Min-Max)**	**IQR**
BASELINE					
ICI	0.7	4, 5, 7, 8, 9, 17, 18, 19	.200	3.8 (2.5-4.65)	0.3
PSQ	.79	1	.200	1.78 (.56-3.11)	.95
NARS_1	.72	7, 8	<0.001	2 (1-4)	.88
NARS_2	.734	14	.021	2.5 (1-4.75)	1.13
NARS_3	.86	-	.044	2.33 (1-5)	1.83
C_S (Slow)					
Distress	0.9	-	.001	1.63 (1-3.88)	1.13
Engagement	0.81	-	.055	3.63 (2.38-5)	.94
Worry	.819	15,16	<.001	1.5 (1-3.83)	.83
Positive Aff	.85	-	.2	3 (1.4-4.6)	1.4
Negative Aff	.75	2	<.001	1 (1-3)	.5
eLoC	.53	1, 3	.069	3.25 (1.25-4.5)	1
C_F (Fast)					
Distress	.83	-	.011	1.5 (1-4)	.75
Engagement	.80	-	.2	3.88 (2.5-5)	.75
Worry	.79	15, 16	.136	2 (1-4)	1.33
Positive Aff	.87	-	.2	3.6 (1.8-4.8)	1
Negative Aff	.81	2	<.001	1.25 (1-3)	.63
eLoC	.64	1, 3	.262	3.5 (1.5-4.75)	1
C_A (Adaptive)					
Distress	.88	-	<.001	1.5 (1-3.5)	.5
Engagement	.83	-	.171	3.88 (2.38-5)	.88
Worry	.83	15, 16	<.001	1.67 (1-4.33)	1.42
Positive Aff	.84	-	.2	3.4 (1.6-4.8)	1.1
Negative Aff	.64	2	<.001	1.25 (1-2.75)	.38
eLoC	.69	1, 3	.038	3.5 (2-4.75)	1.25

TABLE 5.7: Distribution, Median, Minimum and Maximum Scores and Interquartile Range for the Baseline condition.

Phase	K-S test	Median (Min-Max)	IQR
BASELINE			
Valence	.003	4 (1-8)	2
Arousal	<.001	7 (3-9)	2.5
Dominance	<.001	7 (3-9)	2

TABLE 5.8: Distribution, Median, Minimum and Maximum Scores and Interquartile Range for the Slow condition.

Phase	K-S test	Median (Min-Max)	IQR
C1 (Slow)			
Valence	.90	4 (1-8)	2.5
Arousal	<.001	7 (3-9)	3
Dominance	<.001	8 (3-9)	2
Mental Workload	<.001	4 (1-15)	3.5
Physical Demand	<.001	4 (1-20)	3
Temporal Demand	.024	4 (1-14)	6
Effort	.001	4 (1-16)	5.5
Performance	<.001	5 (1-19)	7.5
Frustration Level	.003	4 (1-17)	5.5
Cognitive Workload	.198	30 (10-67)	19.5
Assemblies	<.001	13 (10-15)	2

TABLE 5.9: Distribution, Median, Minimum and Maximum Scores and Interquartile Range for the Fast condition.

Phase	K-S test	Median (Min-Max)	IQR
C2 (Fast)			
Valence	<.001	4 (2-8)	2
Arousal	<.001	7 (3-9)	3
Dominance	.004	7 (4-9)	2
Mental Workload	.023	5 (1-15)	7
Physical Demand	<.001	5 (2-16)	5
Temporal Demand	.029	12 (1-16)	8.5
Effort	.083	7 (1-16)	6.5
Performance	.172	8 (1-15)	6.5
Frustration Level	.003	4 (1-14)	6.5
Cognitive Workload	.2	45 (11-75)	22
Assemblies	.200	18 (12-22)	3

TABLE 5.10: Distribution, Median, Minimum and Maximum Scores and Interquartile Range for the Adaptive condition.

Phase	K-S test	Median (Min-Max)	IQR
C3 (Adaptive)			
Valence	.027	4 (2-8)	2
Arousal	.004	7 (3-9)	3
Dominance	.001	7 (3-9)	2
Mental Workload	.044	4 (1-14)	5
Physical Demand	<.001	4 (1-19)	3.5
Temporal Demand	.165	5 (1-14)	7.5
Effort	.003	4 (1-15)	4.5
Performance	.032	6 (1-16)	6.5
Frustration Level	.003	4 (1-11)	5
Cognitive Workload	.2	32 (12-66)	23
Assemblies	.006	16 (9-20)	2.5

To gain a better understanding of the collected data, simple and partial Spearman correlations are run between all the variables recorded at the baseline and after each condition, with the following results:

- *Slow condition*: Valence (the higher the score, the more negative the emotion) correlates with distress measured by the SSSQ ($\rho = .504, p = .014$). This indicates that higher distress is linked to more negative emotional states, which aligns with what is typically expected: people feeling more stressed tend to experience more negative emotions. Also, engagement correlates with positive affect ($\rho = .727, p < .001$), meaning that people that are more engaged in the task feel more positive emotions. A positive correlation between dominance and arousal ($\rho = .642, p < .001$) is also found. This suggests that a sense of control may be linked to a more relaxed state of mind, as reported in psychological studies (Hong et al., 2021).

- *Fast condition*: Total workload and arousal correlate negatively (ρ = -.439, p = .036). This relationship may indicate that increased workload leads to tension in participants, potentially as a response to heightened task demands. As reported by Carissoli et al., 2024 and Wixted and O'Sullivan, 2014, distress and cognitive workload can negatively affect employees' job performance and satisfaction, as well as their well-being. As for the Slow condition, engagement correlates positively with positive affect (ρ = .839, p $<$ .001). Interestingly, a positive correlation is found between distress and worry (ρ = .466, p = .025). Thus, the speed of the cobot may raise anticipatory concerns in people about task performance or potential errors, as previously reported by Arai, Kato, and Fujita, 2010.

- *Adaptive condition*: A positive correlation emerges between negative affect and distress (ρ = .60, p = .001). This finding aligns with prior research showing that negative emotional states amplify the perception of stress or discomfort (Fiori, Bollmann, and Rossier, 2015; Yoon et al., 2022). Finally, in line with the other conditions, engagement confirms its correlation with positive affect (ρ = .523, p = .006).

A dedicated discussion is needed for the results obtained regarding the Experiential Locus of Control. The strong positive correlation between engagement and ELoC found in the Slow ($\rho = .692, p < .001$) and Adaptive (ρ = .732, p $<$ .001) conditions (also in the Fast condition, but not when partial correlations are performed) points to an interesting association between participants' involvement in the task and their sense of control over the situation. This finding implies that when participants feel more engaged, they may perceive themselves as having a greater influence over the task outcomes or vice versa. This perceived control could enhance their motivation and sense of agency, ultimately fostering higher engagement in the activity. Thus, a higher Experiential Locus of Control in collaborative settings can make participants feel more active or autonomous. This aligns with well-known psychological theories of self-determination and intrinsic motivation (Baard, Deci, and Ryan, 2004; Gagné and Deci, 2005).

Psychology-in-the-loop.
Main take-aways:

Aligning with the goals of Industry 5.0, it is important to explore the influence of robot parameters on the participant's emotional and psychological state. In particular, considering industrial applications, **the effect of a changing collaborative production rate is addressed as one of the most perceivable and impacting factors**.

First, the use of objective measures is evaluated to establish a robust method for real-time estimation of the perceived level of challenge. In fact, if challenges and skills are balanced correctly it is possible to ease the user into an optimal state of Flow. Even though the affective state estimated through facial expression shows different trends as a response to the changing production rhythm, the associated values remain within the thresholds of a neutral state. Thus, **Valence and Arousal indexes are not good indicators of the perceived challenge level**. On the other hand, heart rate data make it possible to significantly distinguish between conditions, identifying the adaptive condition as the one with the best balanced user reaction. Hence, **HRV features represent a promising indicator for the level of perceived challenge**. Using this signal, a prediction model is also trained. However, results are not robust and informative enough for actual implementation in the generalized human-driven control architecture. At the current stage of development, **this kind of signals could be beneficial as an auxiliary measure providing additional information to discern unclear user states** (e.g., distraction and boredom).

Secondly, subjective measures are collected through a number of validated questionnaires. The different outcomes obtained with respect to the previous study highlight that **there is a need for a deeper understanding of subjective variability in stress perception**. In general, results show that faster production rhythms lead to higher perceived workloads and higher physical and cognitive demands, while slower production rhythms generate experiences of frustration and lack of engagement. Also, engagement is strongly correlated to both positive emotions and the Experiential Locus of Control. These results seem to suggest that **being able to personalize the collaboration rhythm with a cobot can generate positive interaction experiences by promoting engagement and a sense of control over the task**.

5.4 Integration and control of a virtual character

An analysis of human-human interactions reveals the importance for verbal and non-verbal communication. In every socially interactive scenario, motor correlates such as lip-syncing, head nods, deictic gestures and gaze movements are abundant and play a great role in expressing emotions and intentions and clarifying unexpressed details laying the ground for the actual content of the communication (Mavridis, 2015). Robots often do not offer any of these capabilities that are fundamental to build a natural and social interaction. From a conceptual point of view, a virtual avatar could act as a mediator between a robot and the user, promoting a more natural and social experience with what is often considered just a tool. In fact, software agents on a screen can easily move in lifelike ways and reproduce sets of actions that are not feasible for today's robots. On the other hand, physical embodiment and presence increases salience and importance of the entity compared to two dimensional entities (Kawamichi, Kikuchi, and Ueno, 2005). Starting from the hypothesis that this last statement can be considered true also for technology, studies demonstrate that physically co-located robots, moving in space and able to manipulate objects, are generally perceived as more anthropomorphic and more engaging (Kiesler et al., 2008). On the basis of the above observations, it appears that the integration of the physical capabilities of a robot with the verbal and non-verbal skills provided by a virtual character may represent a viable solution to enhance the perception of the system as a social entity. Depending on the perceived social role, this would also influence the experience of interaction of the user (Ray, Mondada, and Siegwart, 2008), promoting social engagement and overall well-being.

To explore this topic, the next sections address different aspects regarding the integration of a virtual character and a robot, considering both industrial and rehabilitation applications. The goal is to achieve a level of integration high enough to let the user perceive the two entities as one, as if the avatar visually represented the intelligence of the system with the robot incorporating its physical interaction capabilities. In fact, achieving such goal would allow the user to have more natural and social HRI experiences, specifically tuned on their characteristics and needs.

5.4.1 Towards social embodied cobots: the integration of an industrial cobot with a social virtual agent

Since the classic animation experiment presented by Wick et al., 2019, it is becoming more and more clear that humans have a strong tendency to impose narrative even on non-humanoid interactions. This is a promising starting point from which to build a system featuring both a robot and a virtual character, fused together as a single social entity. The goal is to be able to reach a level of integration high enough that the robot is perceived as the physical interface that allows the avatar to interact with the surroundings and, in turn, the avatar appears as the intelligence driving the actions and behaviors of the robot. A first step in this direction is to gain insights on the effect that the physical positioning of the robot and the display rendering the virtual character has on the perception that the user has of the system. For this purpose, two possible configurations are proposed to volunteers through an online survey. Here, the methodology and outcome of the study are reported, but the reader can refer to Lavit Nicora et al., 2023 for additional information.

Materials and methods

Once again, the setup and assembly task presented in Sections 4.1.1 and 4.1.2 respectively are leveraged for the purpose of this study. Figure 5.36 depicts two different configurations (co-located and non co-located) that are proposed here for they unique characteristics:

- *Co-located configuration*: A co-located configuration is realized by displaying the virtual character on a tablet screen and strapping the latter to the base of the cobot (left side of Figure 5.36). This way, the two entities are positioned in the same area of the workspace and move together (as the base of the robot rotates, the tablet and the avatar rotate with it). As a result, a sense of unity and embodiment is expected, consequently promoting the perception of the system as single social entity. Moreover, since the robot movements sometimes bring the avatar to positions where the latter is not visible, as would happen with a colleage that turns its back to the user, it is possible that the social role of a coworker is promoted.
- *Non co-located configuration*: A non co-located configuration is realized by putting some distance between the cobot and a large TV screen displaying the avatar (right side of Figure 5.36). As a result, a lower sense

of unity is expected since the two entities are further in space and do not share synchronized movements. Additionally, the TV screen is always within the field of view of the user and, in turn, the avatar is constantly overlooking the on-going task. Therefore, it is possible that, in this case, the social role of a supervisor is assigned to the system.

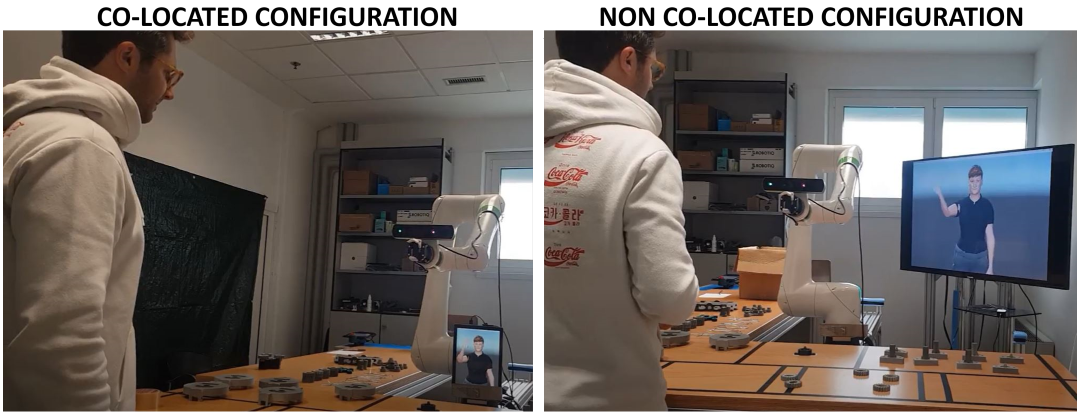


FIGURE 5.36: The configurations proposed to understand the effect of physical positioning of the two entities.

Exploiting these two variations of the workspace, first-perspective videos showing the same pattern of interaction between the user and the system are recorded. For each configuration, two instances of collaboration are represented in the videos:

- The user restocks the cobot table with components while the cobot is working on its part of the subassembly. Below, the related script is reported.

WORKER: Good morning! [*Worker speaks to the resting system.*]
SYSTEM: Good morning! [*Cobot wakes up and avatar waves at the worker.*]
WORKER: Let's get to work.
SYSTEM: OK. [*Avatar looks at the component and cobot moves to pick it.*]
WORKER: I have some parts for you. [*Worker looks at his hand full of components, places them on the table and then moves towards his side of the workcell.*]
SYSTEM: Thanks!

- The cobot holds its subassembly in a precise orientation while the user assembles it with the other subassembly by correctly meshing the gears. Again, the related script is reported below.

WORKER: Hey, I'm almost done. [*Worker is handling some parts and looks at the system.*]
SYSTEM: Here I am. [*Cobot brings the finished subassembly in front of the user, the user completes the assembly and retrieves the product.*]
SYSTEM: Thanks!
WORKER: Great! Thank you! [*Cobot moves back to its table while the avatar is looking in that direction. The worker puts the finished product in a box.*]

Then, the recorded videos are incorporated in an online questionnaire made up of four individual scales (5-point Likert, from "strongly disagree" to "strongly agree"): a social presence scale (the feeling of working with someone, rather than something), a collegiality social role scale (the feeling of working with a colleague, rather than a supervisor), a supervision social role scale (the feeling of working with a supervisor, rather than a colleague), and a unity scale (the feeling of perceiving the cobot and the avatar as a single entity). Additionally, a score (7-point Likert) is added to measure the ability to imagine oneself in the depicted situation. The scales are first piloted with 10 volunteers to make sure that their internal consistency is sufficient. After successful piloting (Cronbach Alpha between 0.7 and 0.95), 20 more volunteers are recruited for the actual study. Each participant is administered the above questionnaire integrated with either the video of the co-located interaction or the non co-located one. One week later, the participants are administered the same questionnaire again, but integrated with the video option they have not yet seen.

Analysis and results

After confirming that all scales are normally distributed, a comparison between the two groups (Group 1: the ones who have seen the co-located option first, Group 2: the ones who have seen the non co-located video first) demonstrates that the order of questionnaire administration does not affect the responses significantly. Moreover, the imagination scale shows an average moderate immersion ($M = 3.32$, $Median = 3$, $SD = 1.82$), meaning that the participants were able to project themselves in the situation depicted in the videos enough to consider all other responses reliable.

On these basis, additional statistical analyses can be carried out. First of all, differently from what was expected, no significant differences can be found between the two proposed configurations. In fact, both configurations achieve high levels of perceived social presence and unity, meaning that the relative position of the cobot and the avatar do not affect the experience of the user. However, some interesting correlations emerge, as from the scatter plots in Figure 5.37. In particular, significant correlations are found between social presence, collegiality and unity. Additionally, no correlation is found between social presence and supervision. This outcome points to an interesting direction: if the avatar is developed well enough to be perceived as a social entity, then also a sense of unity (between the avatar and the cobot) and collegiality arise. Even though further analyses, done in presence and with more participants, are required to confirm these results in a robust and generalizable way, they are extremely important as they confirm that the integration of a virtual character in the generalized human-driven control architecture would positively impact user experience. In fact, its presence would translate in the desired social role of a colleague, rather than a supervisor controlling user's behavior and performance.

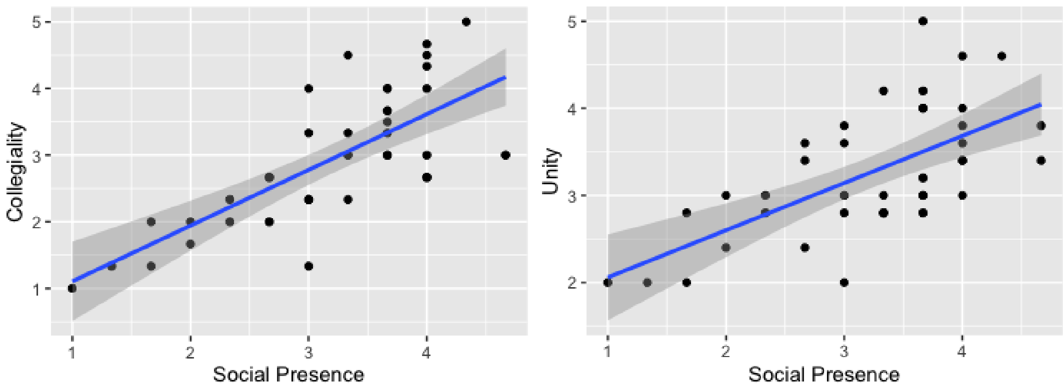


FIGURE 5.37: Plots of the correlation between social presence and collegiality (left) and social presence and unity (right).

5.4.2 Socially interactive agents as cobot avatars: developing a model to support flow experiences and wellbeing in the workplace

The study reported in Section 5.4.1 revealed that if a virtual character is convincing enough to be perceived by the user as an actual social entity, its presence within the robotic system is perceived positively and united with the robot itself. However, in order to achieve this goal it is necessary to base the behavior of the avatar on a proper emotional model of the user, designed to anticipate and counteract counterproductive emotional experiences during HRI.

With reference to Section 5.3.1, a pleasurable and effective state of deep engagement in a certain activity can be referred to as Flow. It has already been shown that a properly balanced skill-to-challenge ratio facilitates experiences of Flow, while a mismatch of these two dimensions can lead to either boredom/apathy or stress/anxiety and even shame. To react to the individual emotional experiences of boredom and anxiety in a way perceived as relevant to the user, these complex constructs need to be dissected first. In particular, a feeling of boredom can be generated from a situation of under-challenge (U-Boredom), the common definition of this psychological concept, but also from situations of over-challenge (O-Boredom). This condition, also referred to as self-focused boredom, can be seen as a defense mechanism against prolonged high-stress levels: individuals tend to subconsciously reduce negatively experienced and self-threatening emotions by entering a state of boredom (Nathanson, 1994). The regulation of negative emotions, however, can remove obstacles from experiencing Flow and even help reappraise the situational demand, consequently re-balancing the perceived skill-to-challenge ratio. Guidance towards said regulation can be beneficial but depends on the context, the individual's emotional experience, and subjective differences. Implicit guidance is often viewed as less obstructive (Heimbuch and Bodemer, 2017). In contrast, explicit guidance typically interrupts the process by giving prompts for action (Loksa et al., 2016). It is therefore important to design regulation guidance strategies that fit well with the specific application of interest and personalized to each single subject.

Starting from these concepts, the BASSF (boredom, anxiety, self-efficacy, self-compassion, flow) model, based on PAD (pleasure, arousal, dominance) dimensions dichotomization by Mehrabian and Russell, 1974, is introduced. The goal is to understand how the concept of Flow can be mapped onto the PAD space and then leveraged as input to the BASSF model to drive the behavior of a robot-avatar system. Here the main points of interest are reported, but the reader can refer to Beyrodt et al., 2023 for additional information.

Materials and methods

The proposed BASSF model uses the three dimensions of PAD to differentiate between every possible affect. These three dimensions are visualized within a three-dimensional space subdivided into eight octants depending on the value of each axis, with the result displayed in Table 5.11. To make the BASSF model decisions more transparent, adaptable to user needs, and testable, the

TABLE 5.11: The eight octants of the BASSF interventions space.

PAD octant	Affective state
+P+A+D	Flow
+P+A-D	Awe
+P-A+D	Relaxed
+P-A-D	Hopeful
-P+A+D	Hostile
-P+A-D	Anxious
-P-A+D	U-Boredom
-P-A-D	O-Boredom

mentioned differential causes of boredom are used and connected to stress and anxiety.

The setup and collaborative assembly task presented in Sections 4.1.1 and 4.1.2 respectively are operationalized for the purposes of this study. First, 30 minutes of standard collaborative assembly are foreseen: the cobot works on Sub_A while the subject assembles Sub_B and then they collaboratively join the two parts to obtain the final product. Since the robot takes around 50 seconds to complete Sub_A which is longer than the time needed by the participant for Sub_B (Slow phase), the task request is very underwhelming in this case, and a state of U-Boredom is expected. After that, a fake failure of the robot is introduced: the robot stops moving, the researcher comes in to fix the simulated issue and then asks the participant to speed up the production rhythm to make up for the lost time. Now, the robot is given preassembled copies of Sub_A and is, therefore, much faster than the user since it only has to pick them up and bring them in the collaborative assembly space (Fast phase). An additional 20 minutes is performed in this configuration, with the user struggling to keep up with the new robot pace. Since the participant does not know that the failure is simulated, and only hears the request to speed up his/her assembly operations, this second phase should cause a prolonged state of overwhelming time pressure. The reason behind the design of this second phase is the need to elicit strong emotional reactions in the user and, ideally, to study if the overwhelming task request is actually regulated with an O-Boredom response. Following these expected emotional reactions, the cobot-avatar interventions reported in Table 5.12 are activated during the experimental session at fixed times.

A total of 20 participants (12M-8F, 25-48 years old) have been recruited for the study. After the working phase presented above, the participants are asked to move to a separate room where the experimenter replays six recorded scenes. After each scene, the participant is administered a questionnaire developed

TABLE 5.12: Examples of possible cobot-avatar interventions.

Affect/ Intervention	Avatar verbal behavior	Avatar/Cobot nonverbal behavior	Theoretical justification
Self-Awareness vs U-Boredom	"Are you okay over there? Let me know if you need anything!"	A: Head tilted to the right, Bending hips/ C: Increase acceleration & velocity	Increase Self-Conscious- ness and Task-Awareness to reduce boredom via socio-cognitive conflict & increased challenge, while remaining caring (Chehayeb et al., 2021; Bambrah, Moynihan, and Eastwood, 2023).
Self-Efficacy vs Anxiety	"Look at that! We have already done so many pieces!"	A: Surprised Expression	Focus attention on the shared achievement to increase Skill-to-Challenge Ratio and ease the pressure by reminding them that they are a team (Lackas, 2021).
Self-Compassion vs O-Boredom	´´You are doing great! Everybody would be stressed at this speed"	Moderate zoom in: Short compassionate smile	Increases Self-Compassion to facilitate self-regulation and cognitive reappraisal (Lackas, 2021).

to collect subjective perceptions of pleasure, arousal, dominance, flow and self-efficacy. Four of these participants have also been involved in an additional semi-structured interview. The latter lasts around 30 minutes and tries to explore further the user feelings with respect to both the task and the avatar presence/interventions.

Analyses and results

Analyzing the data collected through the administered questionnaire, the self-efficacy dimension seems to struggle in achieving good prediction of Flow, probably due to the small sample size or to the scale used to measure it, which comprises only a single item. On the other hand, dominance achieves significant correlation with both self-efficacy and flow. This means that monitoring the dominance dimension opens up the possibility to infer the users' feeling about the on-going task, specifically in their ability to handle the proposed challenge with their own skill which, by definition, is a direct prediction of the Flow dimension. Figure 5.38 reports the PAD dimensions collected for each participant through questionnaires before and after the change of assembly pace.

Considering the qualitative in-depth analysis that can be done with the information collected during the semi-structured interviews, interesting results emerge. Starting from the Slow phase, all participants describe the task as relaxing, though acknowledging its repetitiveness. After the introduction of

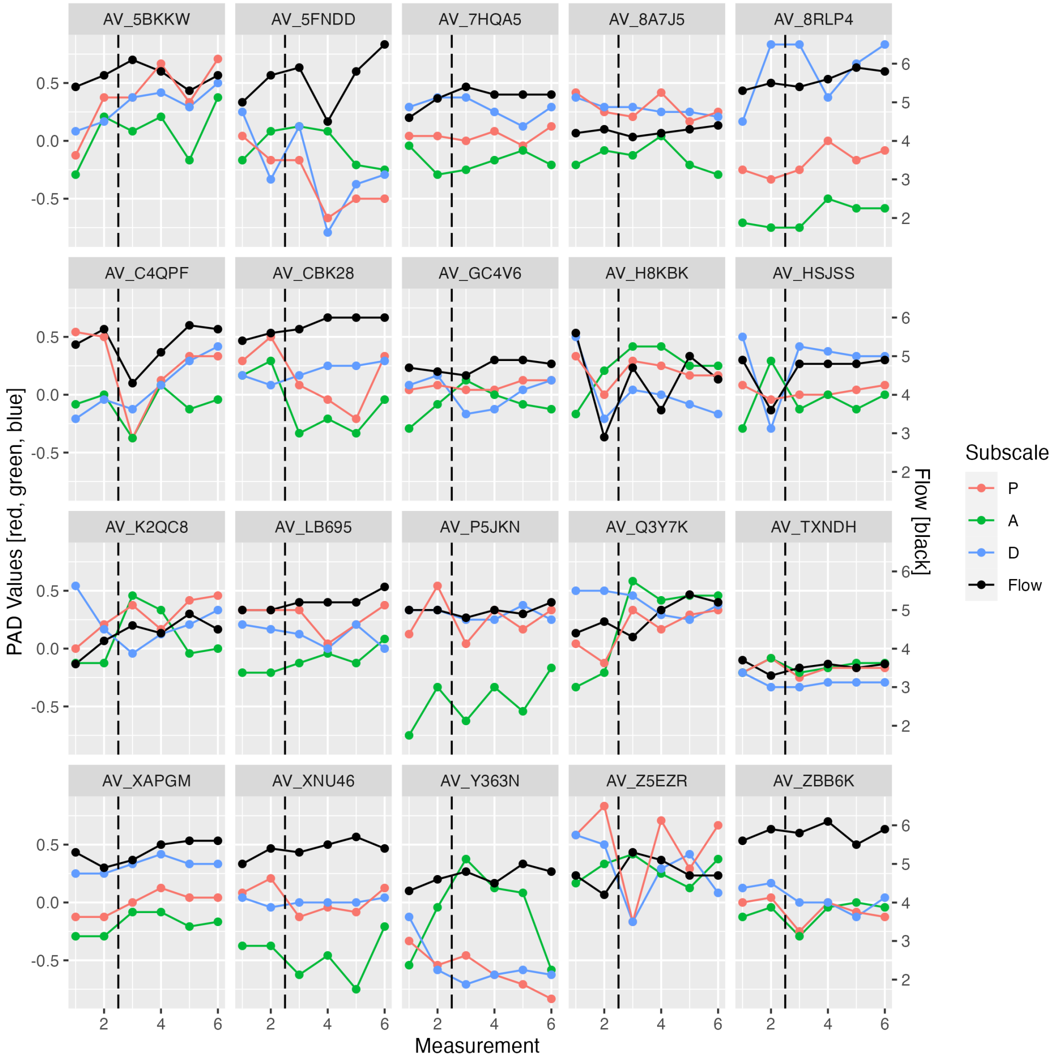


FIGURE 5.38: Flow and PAD values over time for each participant. The dotted vertical line represents the change from the slow to the fast phase.

the Fast phase, none of the participants is able to keep up with the robot, leading them to various levels of stress. All the interviewees, seem to have found their own way (e.g., reappraisal, avoidance, or disengagement) to deal with this feeling, which consequently stabilizes by the end of the session. However, the elicited strategies either caused a decrease in performance or left the subject with a negative feeling by the end of the experiment. This means that the way participants regulated their emotions is not optimal and affects both well-being and productivity, highlighting the need for proper guidance solutions.

To conclude, the findings suggest that the dominance dimension of the PAD model plays a crucial role in predicting flow and, therefore, should definitely be included as input for the generalized human-driven control architecture. Furthermore, this study highlights the importance of guidance in emotion regulation, as some strategies used by participants negatively impacted their well-being and productivity. This justifies and further strengthens the choice of introducing a supportive virtual character in the proposed architecture, but only if the latter is implemented with a robust behavior model driven by real-time reliable user experience measures.

5.4.3 Understanding and mapping pleasure, arousal and dominance social signals to robot-avatar behavior

The study reported in Section 5.4.2, worked on the design of a behavioral model promoting positive emotions for the users involved in human-robot interactions. However, in order for this model to be effective, reliable real-time measures of the user current experience are required. The experimental campaign in Section 5.4.2, based the evaluation of the proposed BASSF model on collected questionnaires which are not suitable for automatic adaptation purposes. On the other hand, recent advancements in modern artificial intelligence (i.e., neural-based machine learning) are leading automated estimation of human communicated emotions based on facial expressions analysis to levels that make it applicable for affective-aware applications (Toisoul et al., 2021). However, it is also known that, as soon as AI models are applied in real-time interactive applications, they often fail due to unpredictable conditions, such as variable lighting, poor-quality cameras, background noise, and unexpected user behaviors.

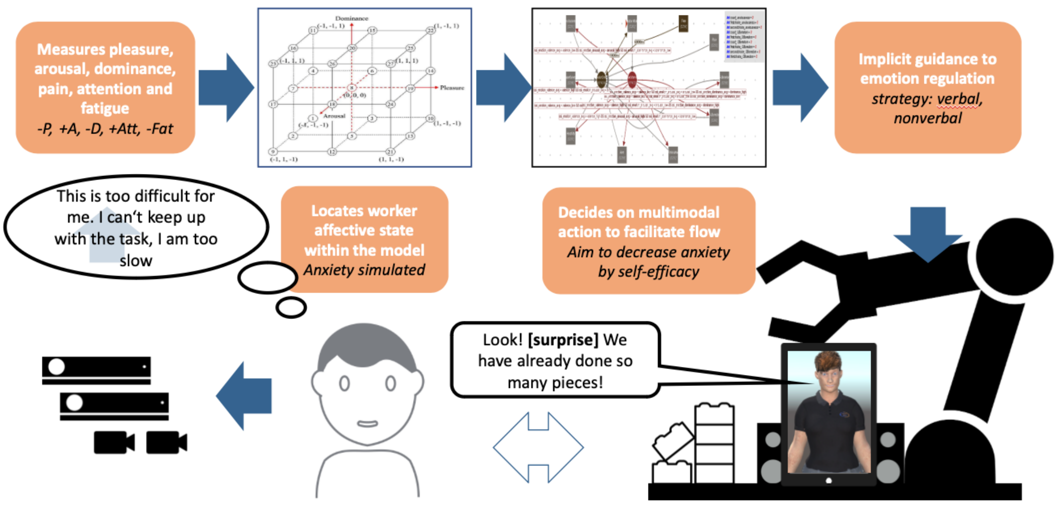


FIGURE 5.39: A top-level conceptualization of the models, sensors and interactions in play.

A possible desired deployment of such an interactive system in an industrial scenario is depicted in Figure 5.39. A worker performs a cyclical assembly task in collaboration with a cobot. The worker's face and body motion are analyzed by a set of AI modules to extract his/her pleasure, arousal and dominance (PAD) inferred values. These values are then leveraged to trigger the execution of a finite set of robot-avatar interventions, designed to promote worker's well-being and flow. However, the idea of using the collected information to trigger interventions as soon as some measure goes above or below a certain threshold, reveals itself to be unrealistic. The "raw" incoming PAD signals are jittery, with many unexpected spikes, the prediction range [0,1] is not fully covered, data distribution is not uniform nor normal, and signals are not centered. In addition, signals are often interrupted because the user is often too far, walks away from the camera field-of-view, the face is too rotated or the frame is too blurry. Unfortunately, in the research literature, such problems are often unaddressed, sometimes categorized as "technical details", and skipped in favor of the description of more theoretical aspects. However, details on the "tricks" injected by developers to put a system at work would often be of extreme importance for the reproducibility of previous research.

A solution to mitigate these effects is therefore sought for in this study. Here, the main results are reported, but the reader can refer to Nunnari et al., 2023 for further details.

Baseline data analysis

Once again, the setup and assembly task described in Sections 4.1.1 and 4.1.2 respectively are leveraged. Specifically, participants are instructed to complete as many products as they can in the given time, but without starting to work on the next gearbox before the previous one is finished. In fact, the robot assembly steps take sensibly longer to complete than the ones assigned to the user, resulting in long waiting times and therefore leading to boredom and frustration. Ultimately, the goal is to elicit significant emotional reactions within the short time of an experimental session, so that the collected data is rich enough to extract informative characteristics on their trend. A total of 14 participants (12M-8F, 25-48 years old) are recruited for the experimental campaign. Each one of them is first asked to assume a neutral facial expression (relaxed facial muscles and closed lips) and look into the webcam for

12 seconds. After that, 30 minutes of collaborative assembly start, with the avatar present in the workcell but not interactively.

The first point of interest comes from the analysis of the neutral face recordings from all participants. Interestingly, data clearly shows that the emotion recognition module does not provide centered values: different subjects lead to statistically different distribution profiles, as visible from Figure 5.40. Results make it clear that there can be no assumption about reference values for the neutrality of facial expressions: calibration is needed for each user, and assumptions on the emotion expressed by the face must be made relative to said calibration data.

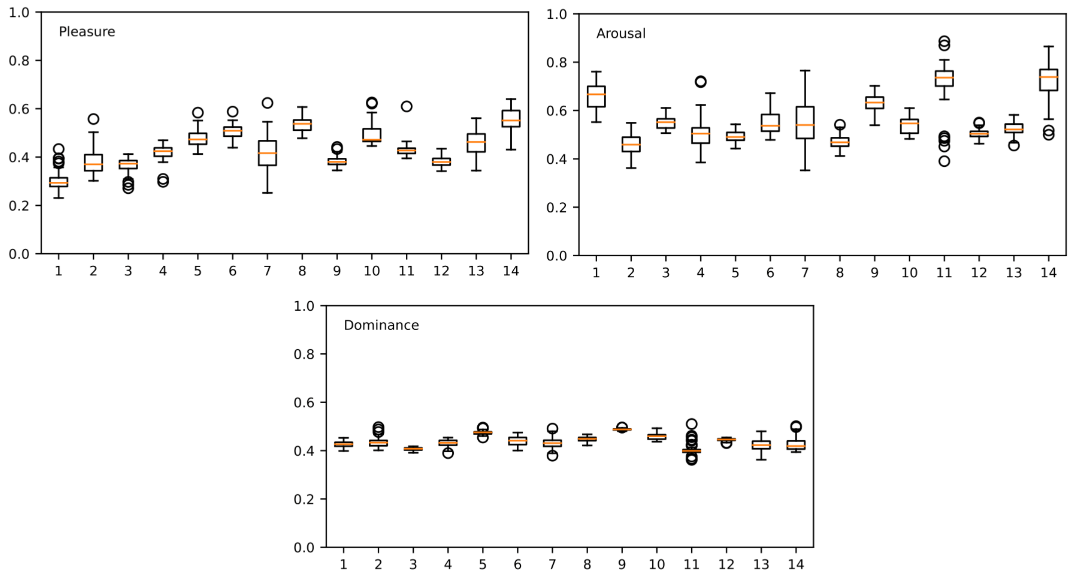


FIGURE 5.40: Box plots showing the distributions of pleasure (top-left), arousal (top-right) and dominance (bottom) collected from participants during the neutral expression phase.

Now, moving to the data collected during the 30 minutes of collaborative assembly, two main problems can be identified. First, data is very jittery, presenting strong spikes, especially when the user's face enters the camera field of view. This is of course not acceptable for a trigger-based system and can simply be addressed by filtering the incoming data with a median over the last second of collected samples. Additionally, the face is often not recognized, because of the user walking around, quick movements (blurry frames), or excessive head rotation. Hence, the continuous signal is full of "temporal holes". Again, this is not acceptable for the logic driving the systems and adds an additional risk: if a face is not visible for a while and then re-enters the camera view, the median might be computed between freshly received values and values received long before. A simple countermeasure can be implemented by introducing a time-aware buffer, so that values older than 1 second are removed from the median computation.

Another problem with the collected data distribution comes from the statistical rejection of the hypothesis of normality, as most of the data distributions present a strong skewness. As a consequence, the intervention triggering system cannot rely on thresholds symmetrical to the median and should be defined separately for positive and negative values. For this purpose, the data collected for each participant is first centered around its calibration mean and then split into positive and negative sets. The root mean squared error is then computed on the union of all positive and all negative sets separately. With this approach, it is possible to compute the generalized low and high threshold reported in Table 5.13 for each of the PAD dimensions. Then, at run time, the thresholds can be customized on each participant by computing them as deviation from their calibration median.

TABLE 5.13: Mean squared errors computed from our calibration data.

Dimension	above median	below median
Pleasure	$E_P^+ = 0.138$	$E_P^- = 0.098$
Arousal	$E_A^+ = 0.071$	$E_A^- = 0.134$
Dominance	$E_D^+ = 0.052$	$E_D^- = 0.020$

Mapping PAD signals to interventions

After gaining a better understanding of the features of the incoming PAD signals and of how they can be processed to render them usable in the application of interest, a model laying the basis for their exploitation is required. For this purpose, the affect model presented in Section 5.4.2 is chosen. The model aims at assisting workers reach the Flow state previously introduced in Section 5.3.1. In general, a mismatch between challenge and skill can cause a variety of unpleasant emotions such as anxiety, stress, and boredom, which are incompatible with Flow, when unregulated. The BASSF model aims to help directly and indirectly these kind of situations by influencing the aforementioned challenge-to-skill ratio of the worker (e.g., by manipulating their self-efficacy believes).

Now, given the six ad personam thresholds computed after calibration, when a triplet of PAD samples is received an intervention is triggered if each one of the three signals go above or below the corresponding threshold. This combination is linked to a so-called *activation code*, among the ones previously displayed in Table 5.11. For example `+P+A-D` denotes a combination of high pleasure, high arousal, but low dominance. On top of this, a multiplier $1/K$ is introduced in the computation of said thresholds to be able to control the sensitivity of the model. In fact, it is important to fine-tune how often the cobot and the avatar should intervene: too many interventions could distract the user from the task while too few would limit the interactivity and effectiveness of the system. By going back to the collected data and changing K, it is possible to find an optimized value that balances the intrusiveness of the system. Table 5.14 reports the results of this analysis.

TABLE 5.14: Average number of activations for P, A, D, and all signals.

K	P	A	D	ALL
2.0	557.14	657.29	479.29	10.79
2.1	484.71	577.93	429.14	4.07
2.2	418.21	518.00	389.50	2.29
2.3	362.57	473.50	351.43	1.21
2.4	313.29	427.07	314.93	0.21
2.5	267.79	381.07	280.29	0.00
2.6	223.57	323.57	250.86	0.00
2.7	187.86	259.86	223.50	0.00
2.8	159.93	187.93	201.29	0.00
2.9	127.71	118.29	179.36	0.00
3.0	102.93	70.71	161.71	0.00
3.1	78.14	43.14	144.50	0.00
3.2	56.14	33.14	127.64	0.00
3.3	43.64	30.79	111.07	0.00
3.4	32.00	27.86	99.07	0.00
3.5	22.29	26.29	87.07	0.00
3.6	14.07	23.71	77.21	0.00
3.7	7.86	21.71	66.50	0.00

As reported, as the sensitivity ($1/K$) decreases (i.e., K increases), the average number of activations for single dimensions goes from more than 450 to less than 60. However, when checking if all three dimensions surpass the thresholds simultaneously, a maximum average of 10 activations is recorded with $K = 2.0$, going down to only 0.21 with $K = 2.4$, and none after that. This is a clear sign that it is not possible to assume that the activation of the PAD channels happen simultaneously since the manifestation of emotions on people's face have different activation delay, persistence, and relaxation time, reflecting in a misalignment of the PAD signals. This means that the trigger based on activation codes should not refer to single sampling times, but to sampling windows. Again, by going back to the collected data it is possible to compute an optimal time window, equal to around 30 seconds, that balances well the delays existing between the three PAD dimensions. An example of the resulting behavior is reported below in Figure 5.41.

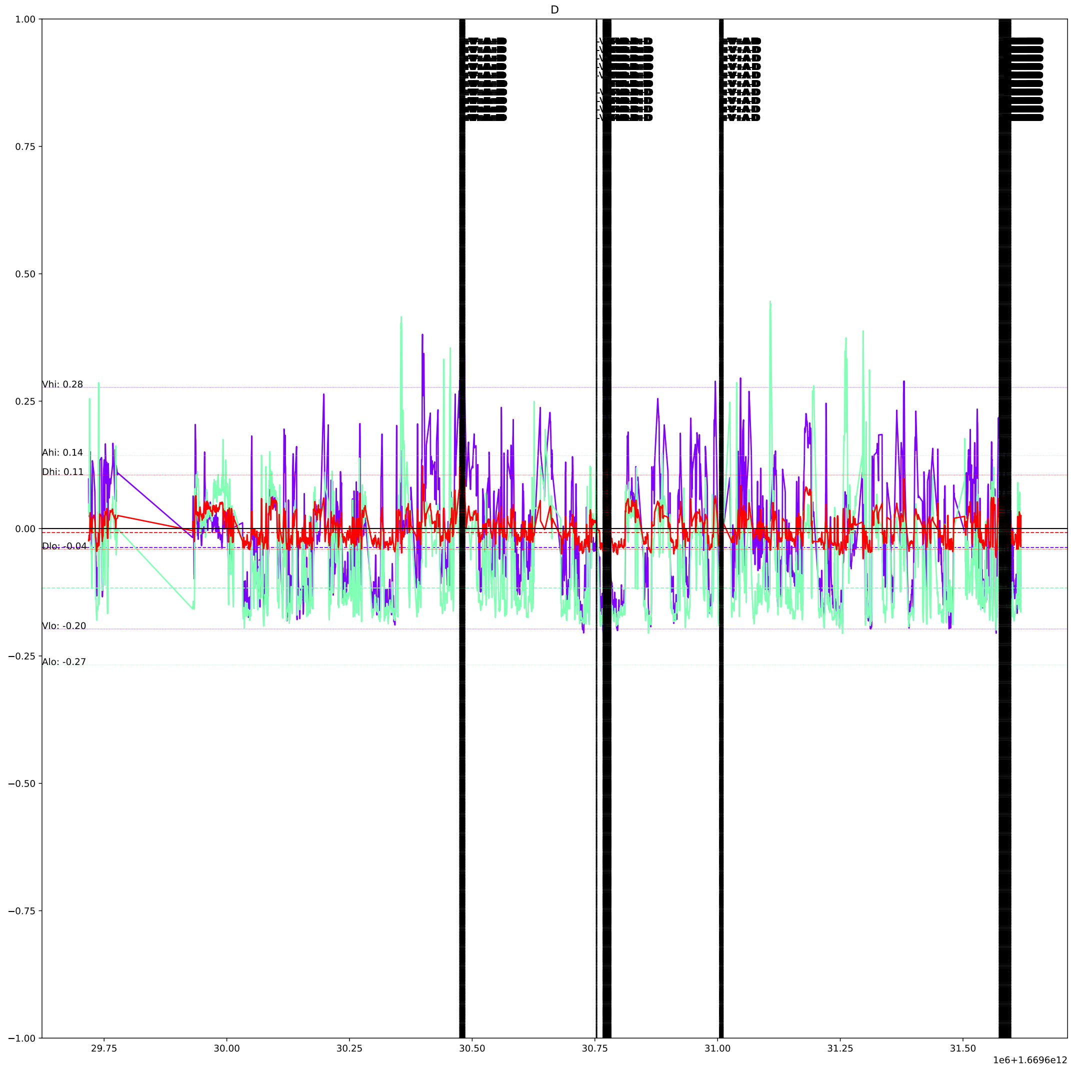


FIGURE 5.41: Example of PAD streams over 30 minutes and their activations (vertical lines), for a single user, with $W = 30$ seconds and $K = 2$.

A final consideration can be done regarding the time of interventions activation. A sudden intervention as soon as all the conditions mentioned above are satisfied might result in an annoying and distracting interruption during moments of high concentration. For this reason, the project consortium decided to synchronize the interventions with the assembly task: activation codes are cumulated during each production cycle and, if an intervention is planned, it is produced only after the joining phase (i.e., a moment of high concentration because of the collaborative meshing of the gears). Although the model and resulting behavior of the system is not formally validated through users' feedback within this study, preliminary piloting shows promising results.

5.4.4 Socially interactive agents for robotic neurorehabilitation training: conceptualization and proof-of-concept study

If up to now this section has focused on the integration of an avatar with an industrial collaborative robot, its application is promising also for other applications where social and empathic interactions are important. Above all, the rehabilitation field is of particular interest for the present project. In fact, robotic devices hold great potential in reducing the dependence on medical personnel during therapy but, at the same time, they generally lack the crucial human interaction and motivation that traditional in-person sessions provide. To mitigate this issue and inheriting the knowledge collected up to now for industrial application, the integration of an interactive socially-aware virtual agent into a neurorehabilitation robotic framework is explored. The primary objective is to test the feasibility of such a system in recreating the social aspects inherent to in-person rehabilitation sessions, fundamental to promote motivation, engagement and, ultimately, better therapy outcomes. Here, an overview of the Empathic Neurorehabilitation Trainer is reported together with a preliminary evaluation of the system, but the reader can refer to Arora et al., 2024 for more insights.

Materials and methods

To gain a baseline rationale, unstructured interviews are carried out with professionals in order to understand not only their approach to therapy but also needs and views regarding the use technological systems for their day-to-day activities. Based on the answers of 15 experienced therapists, the generalized human-driven control architecture presented in Chapter 3 and deployed for rehabilitation scenarios in Section 4.2 is considered promising for the development of the desired system. One primary research question therefore remains: how can the avatar be deployed effectively so that its behavior promotes engagement and motivation for patients without distracting them from the exercise to be carried out?

First of all, the social role that the virtual agent should have is defined: a coach motivating, informing and assisting the patient during his/her rehabilitation journey. Aligned with the view that drove the integration for industrial applications, the avatar and the robot should be synchronized well enough so that they are perceived as a single entity. Ideally, the agent should give the impression of helping the patient both vocally, through proper speech generation, and physically, through the assistive capabilities of the robotic device. Lydia, represented in Figure 5.42, is chosen for this purpose thanks to her speech, gesture and facial expression capabilities. This anthropomorphic design choice goes beyond aesthetics: it serves as a conduit for users to attribute human-like motivations and intentions to the agent, reinforcing feelings of warmth and approachability. In fact, the fundamental concept of trust should lay at the heart of every agent's design and implementation. Drawing inspiration from established principles of trust in human-human relationships, key elements of warmth and competence are therefore integrated into Lydia's behavior. For that, experienced psychologists are involved in the fine-tuning of the agent's verbal expressions, such as the ones reported as an example in Figure 5.42.

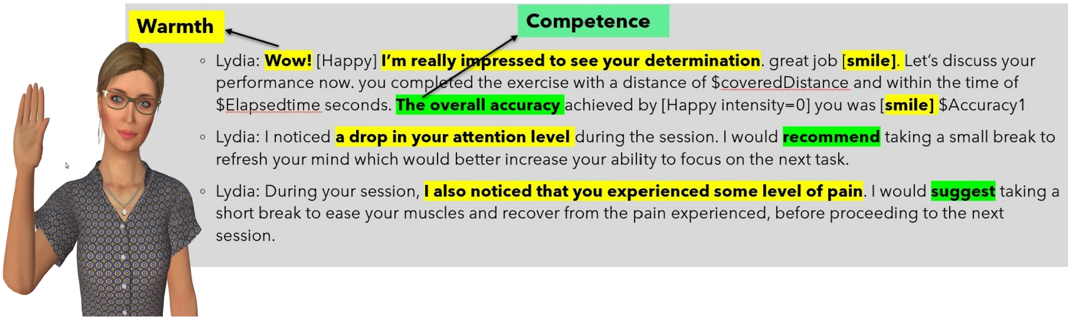


FIGURE 5.42: Lydia, the chosen virtual agent, and some speech examples presenting traits of both warmth and competence.

In the context of neurorehabilitation, metrics such as attention and pain are crucial. Hence, an empathic agent capable of identifying attention and pain contributes to the establishment of a rehabilitation environment that minimizes stress, essential to sustain patient motivation. Therefore, as explained in detail in Section 4.2.3, Lydia's behavior relies on the affective cues regarding stress, attention, and pain inferred by dedicated SSI pipelines. All the information is sent through ROS into VSM and, if the value of any of these social cues exceeds its threshold, the agent is triggered to empathetically inform the user about their current state.

In order to have a preliminary evaluation of the system, a pilot study is carried out with 18 healthy adult volunteers (12M and 6F, 22-33 years old). The GUI and training exercises presented in Section 4.2.2 together with the PlanArm2 rehabilitation robot described in Section 4.2.1 are leveraged for this purpose. In particular, the participants are required to move the end-effector along three ideal trajectories (a circle, an infinity symbol and a straight line), with the device set to assist them using the presented *tunnel controller*. Each exercise is repeated three times, with the user continuously monitored both in terms of performance (i.e., error from the ideal trajectory, total traveled distance and time to complete the task) and of social signals (i.e., facial expressions, gaze and heart-rate to infer stress, attention and pain). After each session, Lydia provides participants with a summary of their performance and inferred status, if different from optimal, offering suggestions to improve. At the end of the training, Lydia also communicates the session in which the participant demonstrated the best precision, with a comparative analysis of the three sessions. Finally, the participants are administered a post-training questionnaire, collecting further insights regarding their experience of interaction with the system.

Preliminary results

Starting from the collected performance data, Figure 5.43 shows the trend of the errors made by the participants over the three sessions. As represented, except for two clear outliers, a consistent downward trend suggests that the volunteers effectively adapted to the device and proposed exercise. This natural adaptation to the task also serves as a preliminary indication that the presence of the avatar did not negatively impact their performance, for instance by distracting them from the trajectory to be followed.

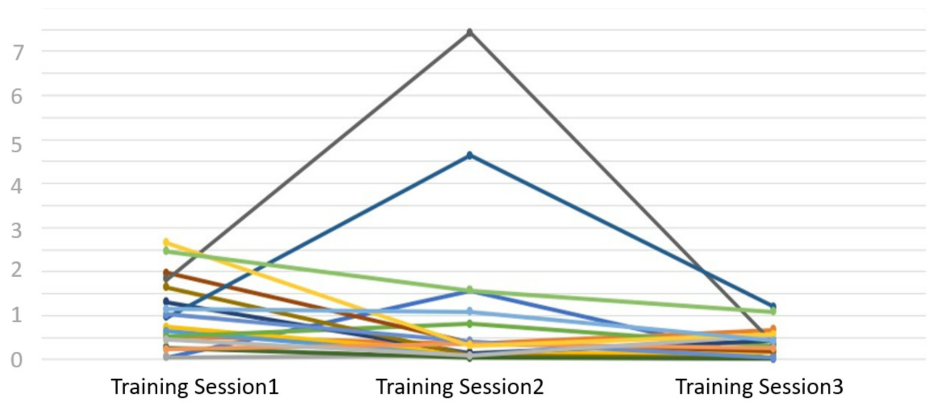


FIGURE 5.43: The error trend of all participants over the three experimental sessions.

Moreover, moving to the qualitative responses collected through questionnaires, it is of utmost importance to understand if the avatar is perceived as distracting or, in turn, as an engaging feature. Figure 5.44 reports the responses of the 18 participants to these questions. As represented, 94.4% of the respondents selected the lower scores with respect to the level of perceived distraction and the higher scores regarding the level of achieved engagement.

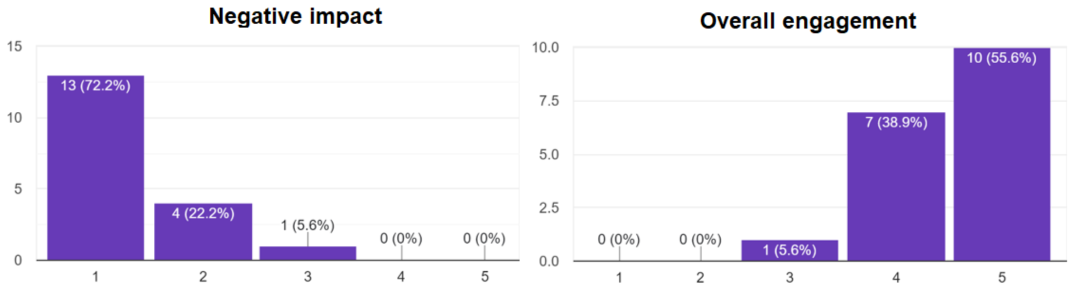


FIGURE 5.44: Participants' responses regarding the level of perceived distraction (left) and engagement (right).

Moreover, most of the participants report that Lydia is a likable virtual character, generally perceived as kind and coherent in its interventions with the proposed exercises. These additional insights are summarized in Figure 5.45. Overall, even though the obtained results are not statistically relevant due to the small sample size and therefore not generalizable to other settings, it is safe to say that this specific deployment of the generalized human-driven control architecture reaches the goals it was designed for. Of course, future studies should evaluate the system within a clinical setting and with actual patients to be able to derive an evaluation of its benefits also in therapeutic terms.

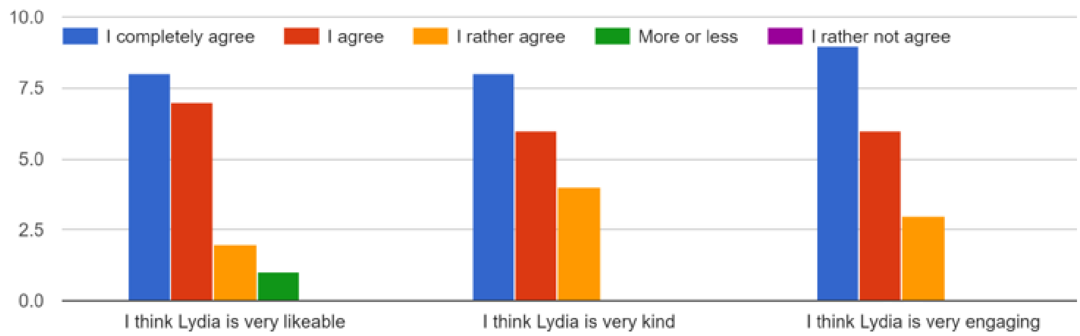


FIGURE 5.45: Participants' responses regarding their overall impression of the avatar behavior.

Integration and control of a virtual character.
Main take-aways:

The integration of the physical capabilities of a robot with the verbal and non-verbal skills provided by a virtual character may represent a viable solution to enhance the perception of the system as a social entity. Depending on the perceived social role, this would also influence the experience of interaction of the user, promoting social engagement and overall well-being.

Interestingly, preliminary results show that the relative positioning of the robot and the avatar do not affect the user perception of the system. In fact, whatever configuration is chosen, it seems that **the presence of the avatar alone, supposing that its behavior is convincing and robust, is enough to let the system embody the social role of a colleague**.

To achieve this goal, the BASSF behavior model is evaluated using subjective measured collected through questionnaires. Interestingly, **the dominance dimension shows the best results in the prediction of flow**. Also, the importance of guidance for emotion regulation is highlighted, strengthening the need for a properly modeled avatar in the system.

In order for the evaluated model to be integrated in interactive scenarios, however, reliable real-time measures of the user's experience are needed. These could be obtained by leveraging dedicated AI modules, if it wasn't for the noisy and uncalibrated nature of their output. **Post-processing strategies are therefore proposed and tested successfully with the aim of promoting Flow during HRI**.

Finally, the full implementation of a robot-avatar system is attempted also for neurorehabilitation applications. An Empathic Neurorehabilitation Trainer is deployed and preliminary evaluated through a dedicated pilot study. Results show that, on average, participants tend to improve their performance over time and **perceive the integrated virtual character as a positive and engaging feature, not causing any distraction during the exercise to be carried out**.

5.5 Comparing and supporting neurotypical and ASD subjects

The constantly growing paradigm of Industry 5.0 is paving the way for user-centered and user-oriented design of workplaces with the goal of transitioning to a more sustainable, inclusive and human-centric industry. Human-Robot Interaction is one of the concepts upon which this technological revolution is building. In this regard, research shows that a true understanding of the effects that HRI has on user experience should be deeply rooted in the analysis of human cognitive behavior (Hormaza et al., 2019). However, in the struggle for finding solutions personalized to the needs of each single user, there seems to be a lack of studies regarding the exploitation of these tools as an inclusion opportunity for vulnerable subjects (Hendricks, 2010).

Above all, this project focuses on people characterized by the Autism Spectrum Disorder as a condition that, often, leads to very specific needs in terms of behaviors and social relationships (American Psychiatric Association, 1994). A major part of the body of literature in this regard, deals with the use of robots as therapy tools for people with ASD, especially children. Their application in inclusive industrial settings, on the other hand, still needs to be deepened. In fact, the fixed and predictable routine with precise task assignment (Goris et al., 2020) that characterizes the collaborative work with a cobot could represent a great fit with the representative features of ASD. Social skills deficits (Weiss and Harris, 2001), a preference for predictability (Goris et al., 2020), difficulties in transitioning (Sterling-Turner and Jordan, 2007) and the need for concrete external feedback on personal performance (Larson et al., 2011) are relevant aspects that characterize this often overlooked condition. Starting from these considerations, the working routine required for industrial automated tasks may be beneficial to offer an important inclusion opportunity, specifically when considering the high-functioning part of the spectrum of the autism disorder (Gillberg, 1998).

Understanding the different needs and behavioral patterns of ASD and neurotypical (NT) users and finding solutions to ease positive interaction experience for all is the main focus of this section and of the studies presented below.

5.5.1 Behavioral patterns in robotic collaborative assembly: comparing neurotypical and autism spectrum disorder participants

Starting from the need of understanding the differences between ASD and NT operators when interacting with a cobot on industrial applications, the present work does not aim to build a new characterization theory, but rather to observe the behavioral manifestations in the two groups within a context that has been investigated very little so far. Given the innovative nature of this goal, an exploratory and observational approach is taken in order to better outline the needs of different users and use them as a starting point to provide an even more personalized experience through the proposed human-driven control architecture. Here, the main results are reported, but the reader can refer to Mondellini et al., 2023 for additional insights.

Materials and methods

The setup and collaborative assembly task presented in Sections 4.1.1 and 4.1.2 respectively are leveraged. A total 16 participants are involved, of which 8 NT (5 females and 3 males, 18-30 years old) and 8 diagnosed with high-functioning ASD (1 female and 7 males, 21-50 years old). The unbalance in the sex distribution towards males for the ASD group, is expected from literature (Loomes, Hull, and Mandy, 2017). It is important to note that none of the participants had prior experience working with an industrial cobot. Participants are asked to work on the task for 3.5 hours a day, for 5 consecutive days, in order to observe modifications in their performance and behavior during the overall experience (from Monday to Friday).

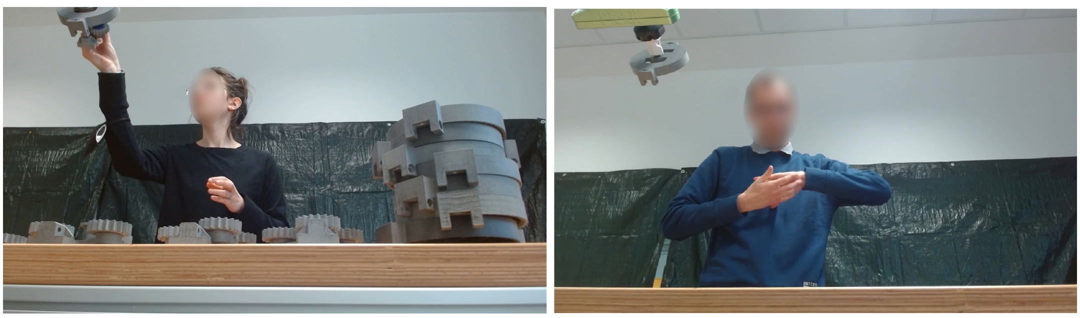


FIGURE 5.46: Webcam captures for the experimental sessions with NT (left) and ASD (right) subjects.

The Logitech C920 Pro HD webcam placed in front of the volunteers' workbench is used to record them during the experimental activities, as from the captures in Figure 5.46. Three sessions of approximately 10 minutes each are video-recorded during the first workday (beginning, middle, and end of the workday). Likewise, three additional videos are acquired during the last workday of the experiment. Thus, one hour of videos for each participant is available to be analyzed, for a total of 16 hours of videos. For this purpose, four different tools are used to collect robust measures representative of both predictable and unforeseen behaviors. Some of the chosen tools allow for the precise observation of predefined aspects of the collaboration, but are not suited for the analysis of long sessions (e.g., video-based annotations). Other tools, instead, have been selected for their good fit with long and unpredictable scenarios (e.g., live note-taking). Moreover, the different chosen measures allow for both a qualitative analysis of the observed behaviors and a quantitative comparison between the two mentioned groups. Below, the list of exploited tools is reported:

- *Observational grid*. To detect some predictable aspects related to well-being and performance, an observational grid is built. In particular, it was decided to note the observed manifestations related to 7 attitudes:

  1. <u>Manifestations of tiredness</u>. Body movements or facial expressions that convey to the observer that the participant is tired are of interest since they can give insights on how to improve the interaction and make the experience less demanding for the two groups.

  2. <u>Hand gestures</u>. All hand movements that are frequent but not useful for the task are noted as interesting features given the known stereotypical gestures of ASD subjects.

  3. <u>Assembly strategy</u>. A class encompassing how the participant assembles the planetary gearbox (e.g., using one or both hands, building several pieces at the same time).

  4. <u>Loading strategy</u>. Participants also have to periodically replenish the components buffers on the robot table. Knowing that subjects with ASD have rigidities in changing their behavior while working on a repetitive task, it is interesting to see how and when this step is carried out.

  5. <u>Regard for the cobot</u>. A class including reactions related to the behavior of the cobot (e.g. talking to it) but also no reactions (e.g. ignoring the robot).

6. Talk to someone. A researcher is always in the room for task supervision and conversations can take place.

7. Other manifestations. Other behaviors that cannot be categorized in the other classes but that still contribute to describing the moment are noted here.

- *Unstructured notes*. Additional data is collected in the form of unstructured note-taking to make sure that the loss of specific behavioral occurrences is minimized. The researcher supervising the session is therefore also in charge of annotating, on a dedicated document, all those interesting behavioral nuances that may be missed by the video-recording sessions. Using the collected notes, informative cards, called "Personas" are built for each participant considering 5 categories: task challenges and strengths, work organization, quotes, recurrent behaviors and emotional expressions. The reasoning behind the choice of these classes is similar to observational grid, but generalized in order to capture a wider range of occurrences.

- *Video annotations*. The NOVA (NOn Verbal Annotator) tool is chosen for the purpose. The recorded videos are run through the software and manually annotated using labels related to the behavior of both the robot and the participant. A first set of labels is dedicated to the ongoing activity phases (gathering parts, assembling, collaborative joining). To that, a dedicated label for waiting instances is added, also characterizing concurrent specific behaviors (e.g., distraction, gaze, talking).

- *Performance analysis*. One piece of information missing from the data collected using the above tools is the quantitative performance achieved by each participant. Therefore, for every day of the experimental week, the supervising researcher noted on an Excel sheet the start and end time of the session, any occurring stop of the activity (e.g., robot failure, participant taking a break), and the total number of assembled gearboxes per day. Using this data it is possible to extract uptimes, downtimes and a performance index computed as the ratio between total number of completed gearboxes and total uptime.

Thanks to this rich dataset, it is possible to run a series of analysis in order to gain further insights on repeating patterns, needs and behaviors. However, it is important to remember that the small sample size limits the chance of generalizing results. In fact, this work only aims to explore possible points of interest to be studied further before being translated into actual applications inside the proposed human-driven control architecture.

Main results

For brevity, here only a comparison between the two groups is reported, as some interesting differences emerge from the observations and analyses made on the collected dataset.

From a qualitative point of view, a greater number of manifestations are recorded for the ASD group both in terms of tiredness/boredom and regarding stereotyped movements/gestures. Also, in terms of facial expressions ASD participants show more variability than the NT ones while, analyzing their gaze behavior, the tendency to look towards the cobot when it is time to start collaborating seems to be significantly reduced. This is a first interesting result since, with reference to the study reported in Section 5.2.2, this outcome indicates that the effectiveness achieved with NT subjects by the gaze-based attention triggering system may not replicate with ASD operators. This is proof of the fact that solutions validated for specific groups of people may not be generalizable to all population, highlighting the need for the importance of personalization in HRI. Continuing with the comparison, the NT group seems to have a faster adaptation to the task, by gradually changing the sequence, timing and positioning of their actions to achieve a better performance over time. This adaptation is slower or totally lacking for participants with ASD, specifically in their way of interacting with the cobot. For most of the instances where the robot is waiting for the operator to initiate the joining phase, ASD participants did not show any urgency to collaborate, de-prioritizing this step in favor of their assembly scheme and with a consequent loss in performance. Moreover, the analyses show that ASD users prefer to maintain a certain distance from the robot throughout the sessions. This is evident by looking at the strategies elicited by the two groups regarding the replenishing of components on the robot table: if NT participants gather these components as soon as they are needed regardless of the robot actions, the ASD group tends to perform that action when the robot is stopped in the collaborative joining position.

The quantitative data collected regarding the interaction seems to confirm some of the differences mentioned above between the two groups. With reference to the left side of Figure 5.47, the lack of urgency in attending the robot observed for ASD participants reflects in the statistical distribution of

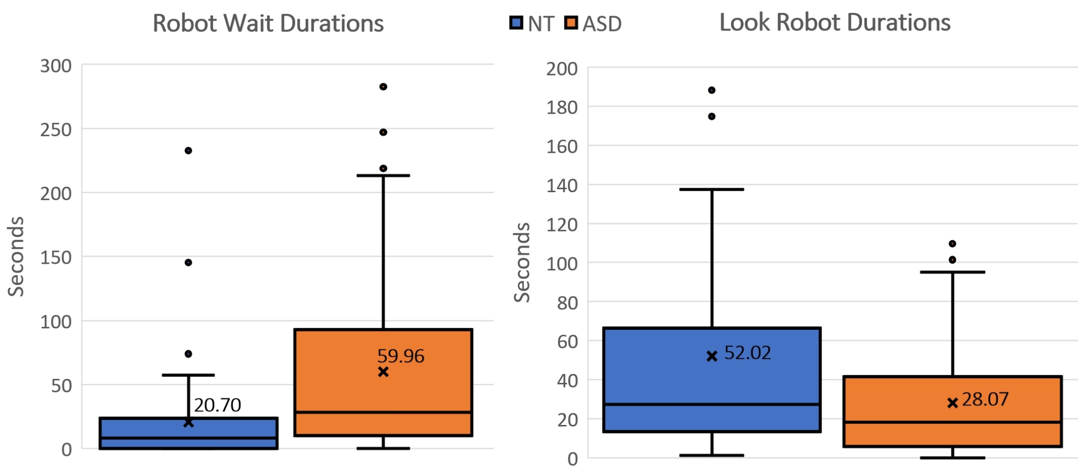


FIGURE 5.47: ASD vs NT differences regarding robot waiting times (left) and gazes directed towards the robot (right).

robot waiting times. As shown, the average robot wait per video recording is 20.70*s* for neurotypical subjects, while the ASD group accounts for almost three times that, resulting in 59.96*s* of wait time per video. The differences in gaze patterns are also confirmed by the quantitative analysis. As reported by the box plot in the right side of Figure 5.47, NT generally spend more time looking at the robot. Moreover, further analyses show that the duration of gaze contact with the robot is sensibly shorter for the ASD group, replicating previous results obtained by Damm et al., 2013.

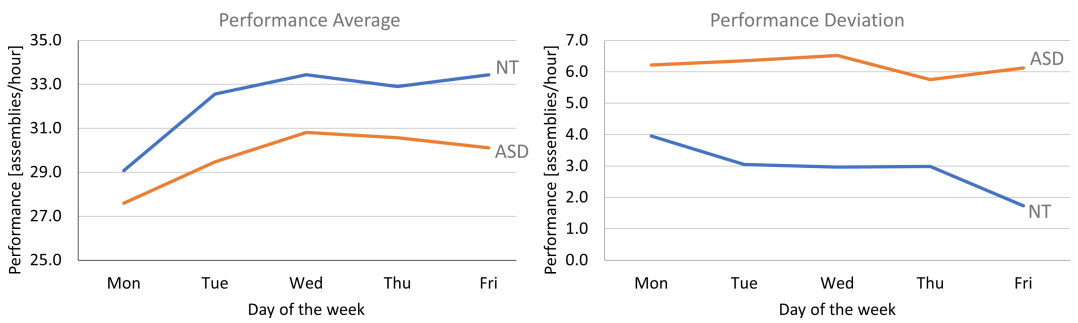


FIGURE 5.48: ASD vs NT differences regarding performance average (left) and deviation (right).

In terms of performance (expressed as *gearboxes/hour*), results are summarized in Figure 5.48. The NT groups clearly shows a trend of increasing performance (+15%) and a tendency to converge towards a common top result (SD from 3.95 to 1.73 over the week). The ASD group also achieves an increase in performance (+9%), even though more moderate, but, on the other hand, results are quite spread apart and do not seem to converge to a shared optimum (SD oscillating between 5.75 and 6.52 over the week). If, on average, data seem to confirm what is highlighted by the qualitative observations in terms of performance (i.e., a slower rate of adaptation and multi-tasking for the ASD group, resulting in a lower number of completed products), a surprising result emerges through a subject by subject analysis. In fact, both the best and worst performers among all participants belong to the ASD group, signaling that the mentioned inclusion opportunity could be beneficial also in terms of production, if the right task is assigned to the right person and personalized accordingly.

A final interesting note should be taken into consideration. Even though, on average, the two groups seems to have significant differences in terms of gaze, gestures, adaptation to the task and resulting performance, it is always important to remember that true personalization cannot happen through generalization. Each individual subject, even though considered within a group as a clustering strategy to simplify the implementation issues, has his/her own needs and approaches to daily life. Therefore, general rules drawn for homogeneous groups should only represent the starting point from which smart systems adapt their behavior with the aim of promoting positive and natural interactions.

5.5.2 Biomechanical analysis on neurotypical and autism spectrum disorder people during human-cobot interaction

Sections 5.1.2 and 5.1.3 provide evidence that workers' physical assessments are of utmost importance to enhance well-being and reduce work related musculoskeletal disorders. For this reason, if up to now the comparison between ASD and NT individuals in industrial HRI scenarios has focused on behaviors and performance, biomechanical aspects should not be left out. In fact, Section 5.5.1 started to highlight differences between the two groups also in terms of manifestations of tiredness and stereotyped movements which could impact their biomechanical assessment. Moreover, research demonstrates that ASD people often show reduced motor performance and motor coordination of both upper and lower limbs (Fournier et al., 2010; Bennett et al., 2021), increasing the risk of physical fatigue. To further explore these topics in a semi-realistic industrial environment, this study provides a detailed comparison between ASD and NT volunteers in terms of biomechanical measures collected while working on a collaborative assembly task.

Materials and methods

The data necessary to perform the desired analyses is collected during the extensive experimental campaign already described in Section 5.5.1. In particular, skeleton data is acquired using the two redundant Azure Kinect cameras and following the same protocol used before: three acquisitions lasting 10 minutes each are carried out at the beginning of the experimental week (start, middle and end of the workday) and the pattern is repeated for the last day, in order to capture variability both through the day and through the week.

Leveraging the model already presented in Section 5.1.2, the collected skeleton data is used for the computation of kinematic and dynamic parameters. Additionally, since participants are free to move around the workcell for assembly purposes, the covered distance is computed as the 3D Euclidean distance of the spine-chest joint from one frame to the consecutive one. This joint is chosen as one representative of the center of mass (the actual center of mass cannot be inferred as the tracking of the legs is not permitted due to the occlusion caused by the workbench). Torque, power and energy are again computed as in Section 5.1.2 and normalized by dividing by the weight and the height of the participant, in order to allow inter-individual comparison (Saadatian, Sahebozamani, and Karimi, 2023). Further elaborations of the biomechanical parameters are made to identify whether instant or medium-term fatigue/effort is taking place during the task. For this purpose, threshold values of torque and energy are identified (Lorenzini et al., 2023; Yu et al., 2019). The torque threshold is defined as the shoulder torque needed to maintain the shoulder elevated at 90° in a static configuration. On this basis, events of instantaneous fatigue are detected as the times in which the torque exceeded the 80% of the threshold. The energy threshold is, instead, defined as the shoulder work required to perform repetitive reaching movements lasting about 1 second each in a one minute session. An index for medium-term fatigue is then computed as the percentage of the energy in time windows of 1 minute with respect to the energy threshold.

Analyses and results

No significant differences are found between the ASD and NT groups in terms of covered distance around the workcell, even though ASD participants show a tendency of moving less ($p = 0.05$). The maximum normalized torque seems to decrease between the first and last days of the work week for both groups ($p = 0.016$), even though the number of events of instantaneous fatigue remain almost unchanged. A similar decreasing trend ($p = 0.009$) is found when considering the maximum normalized power. Interestingly, this measure also identifies a significant difference between the two groups ($p = 0.012$), with the ASD group reaching higher values. The same can be said for the normalized mean energy ($p = 0.022$), with the ASD accounting for significantly higher levels of expended energy. Figure 5.49 can be used as reference for the data highlighting the mentioned differences.

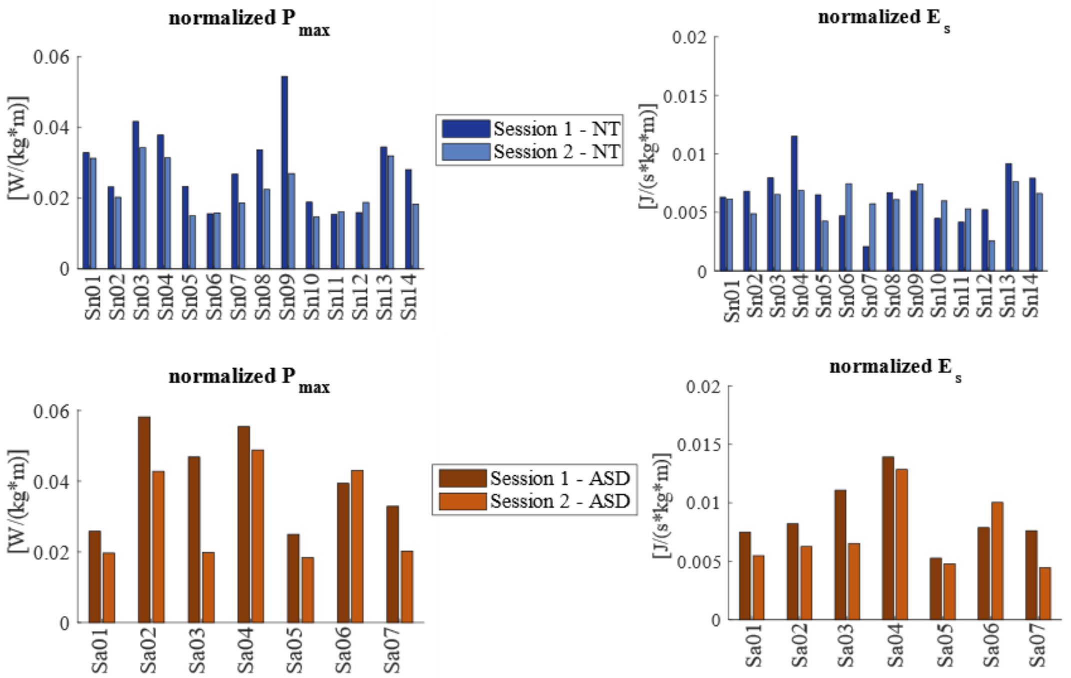


FIGURE 5.49: Barplots for the ASD and NT groups regarding normalized power (left) and energy (right).

Overall, both groups showed an adaptation process that may have taken place to reduce the effort during industrial work. However, results also suggest that the ASD group is subject to higher risks of fatigue due to slightly less efficient movement performance. This result is in line with the findings reported in Section 5.5.1, suggesting that ASD workers account for more manifestations of fatigue and therefore highlighting the need for a careful definition of both the workspace and work schedule. For instance, since ASD operators seem to get tired more rapidly, more frequent breaks could be considered. Other than that, no clearly observable contraindications are found, strengthening the opportunity of inclusion that HRI and Industry 5.0 may offer to vulnerable subjects.

5.5.3 Design and testing of (A)MICO: A Multimodal feedback system to facilitate the Interaction between Cobots and human Operators

The studies presented in Sections 5.5.1 and 5.5.2 highlight a number of differences between ASD and NT operators regarding industrial HRI scenarios. However, these differences do not pose any contraindication towards the involvement of vulnerable subjects if the workplace, assigned task and schedule are designed properly. In fact, Section 5.5.1 highlighted that the characteristics of neurodivergent operators may even be beneficial, leading to higher performances overall. Creating accessible environments where everyone can join in and have the same experience is one of the main concepts promoted by the approach called 'Design for All', a term introduced to indicate "a design for human diversity, social inclusion and equality" (All Europe, 2004) and to emphasize the importance of guaranteeing the dignity of all users. This concept aims to promote the development of smart solutions, well harmonized with their surroundings and usable by all indiscriminately (Ielegems, 2014).

Neurodivergent people face several difficulties in finding and maintaining their job due to the lack of support, also caused by the overall organization of resources and environmental factors such as stigma (Unger, 2002). Research focusing on solutions facilitating the employment of neurodivergent workers, such as those with ASD may provide employers with the appropriate tools and knowledge on inclusive recruitment (Nicholas et al., 2019), reducing barriers to employment. One of those barriers, can definitely be identified in problems of communication. This aspect is of great importance in human-robot interaction, as also highlighted by several scales developed to analyze the level of anxiety towards robots, the process and the factors involved (Nomura et al., 2006a). Robots should be designed with intuitive communication and cooperation modalities for human operators. Sciutti et al., 2018 describe this process as "humanization" of human-robot interaction, which in this context does not refer to the choice of an anthropomorphic appearance, but to the development of a code for mutual understanding between the two agents. This aspect acquires a greater importance especially when designing accessible robotic systems.

Starting from these considerations and the actual socio-cultural context, the present study aims to design a solution to make the interaction between collaborative robots and operators more intuitive and accessible for all neurotypical and neurodivergent operators. In particular, a multi-modal feedback system is proposed, featuring a combination of visual and acoustic signals, to reinforce and integrate the information transmitted by the cobot regarding the activity in progress. Here, the most relevant details about the implementation and testing of such device are reported, but the reader can refer to Dei et al., 2024 for additional details.

Design of the prototype

To design a positive experience with the cobot, the information must be conveyed as intuitively and simply as possible, so that it can be understood by everyone, reducing the stress and anxiety caused by not being in control of the situation. On this, several studies confirmed that visual and auditory signals are the most immediate modalities for individuals to interact with robots (Su et al., 2023). In addition, making the cobot activity more transparent allows for a faster and more efficient collaboration (Gross and Krenn, 2024). Moreover, considering that many workplaces are not designed with accessibility functions, the developed solution must be adaptable to a preexisting system and its feedback should be customizable onto a wide variety of applications.

Starting from the above requirements, the first prototype named (A)MICO (A Multi-modal device to improve inclusive Interaction between Cobots and Operators) is realized as represented in Figure 5.50. The device has a cylindrical shape with a properly sized footprint so that it can be placed in workstations with limited available space, but still transmit the information with enough clarity. (A)MICO is composed of a base with a speaker, used to produce acoustic signals. On top of the base, a hollow semi-transparent cylinder is mounted, equipped with five RGB LED strips. This design makes it possible to provide visual feedback both in terms of colors and 2D graphical patterns (to make it accessible also for color-blind users). Being displayed on a rounded surface, no additional movements of the user are required to see the feedback. Overall, the aesthetic of the device is inspired by that of the signal towers often found in industrial settings. In addition, the rounded shape recalls the lines of the cobot links, making it more emotional and affective, while allowing the possibility to fix the device on the robot itself, thanks to its hollow center.

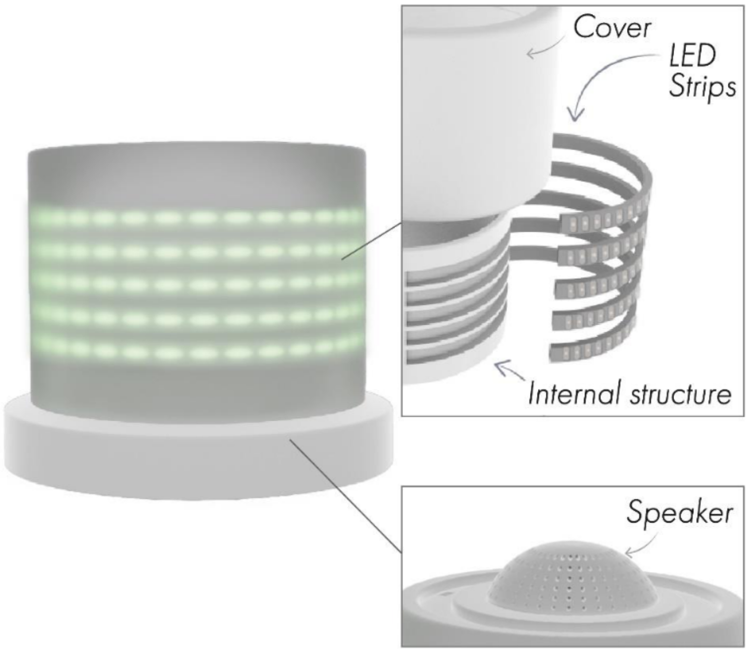


FIGURE 5.50: A rendering for the design and internal structure of the first prototype of (A)MICO.

The device is controlled via dedicated firmware running on an Arduino Nano board. The board is connected to the main PC controlling the robotic workcell via Bluetooth connection. A dedicated software module runs on the main computer and interfaces the device with the rest of the architecture using ROS. Thanks to this approach, it is possible to translate the actions and status of the cobot into lighting and acoustic patterns in real time. Moreover, this guarantees flexible and multiplatform use, as more and more robot manufacturers develop controllers compatible with the ROS communication middleware.

At this point, a co-design phase involving five individuals (1F and 4M) with high-functioning autism is organized. The volunteers, selected for their peculiar communication issues and need for explanations and guidance, are recruited among the ones that participated in the study presented in Section 5.5.1 in order to make sure that they have a proper understanding of the system and of the difficulties that they may face interacting with it. In fact, the goal of this phase is to identify the type of information the device should be able to communicate and how this transfer should be realized. Each participant is therefore administered an interactive questionnaire integrating videos of different combinations of visual and acoustic feedback produced by the device, as in Figure 5.51.

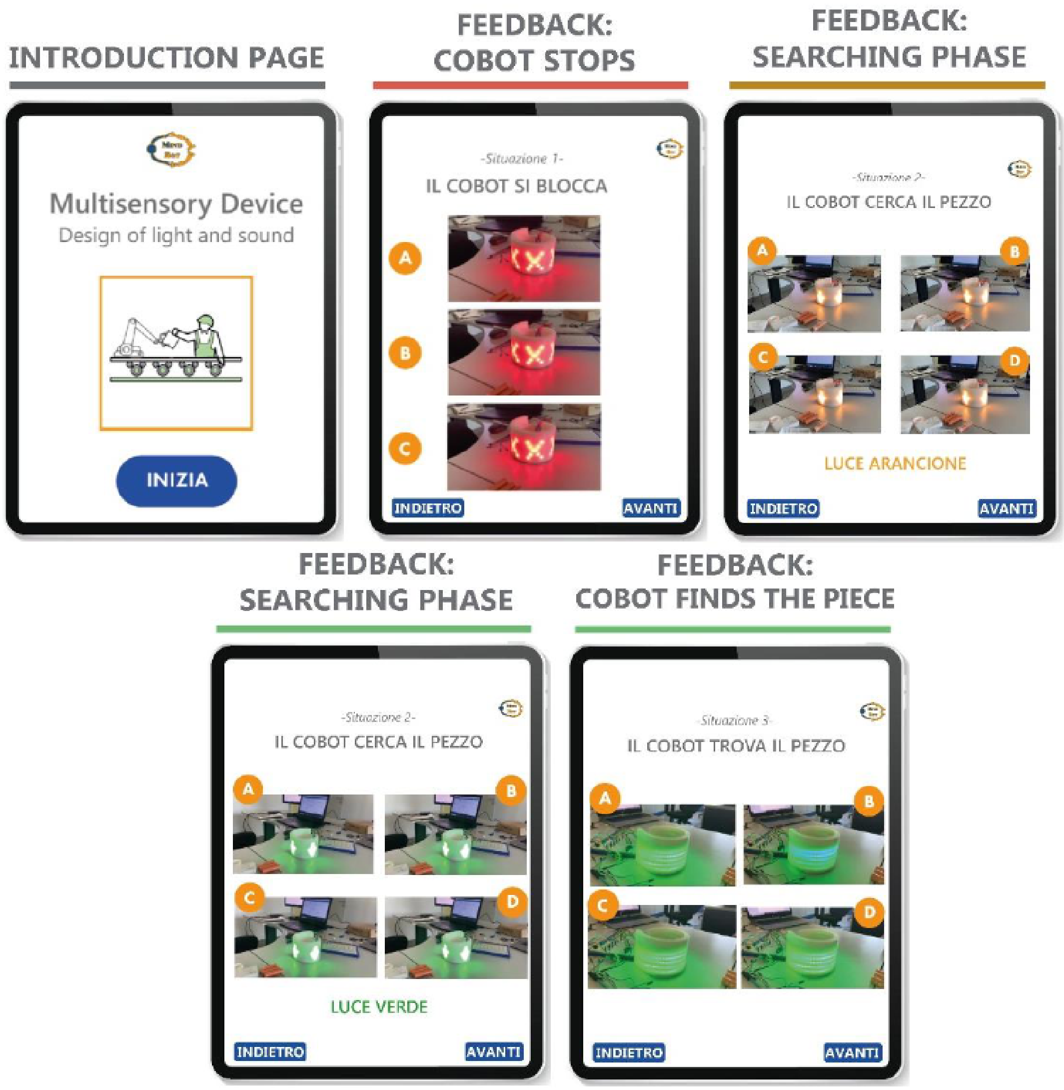


FIGURE 5.51: The administered questionnaire exploring possible feedback during three critical assembly phases.

Based on previous experience, three critical situations are presented in the questionnaire:

1. The cobot stops because of a system error. Manual intervention of the supervising researcher is required to restart the system.
2. The cobot stops because the detection camera fails in finding the next component. The operator is supposed to bring new parts on the workbench or reorganize them so that they are well visible to the camera.
3. The cobot is waiting over the buffer of components while the detection camera succeeds in finding the next part. This small pause in the movements of the robot may look like the detection has failed, when in reality the camera has succeeded. The operator should not intervene

and move the components around in this phase because the robot is about to proceed picking up the next one.

For each of these situations, four options are proposed so that the participants can vote their preferred one. A dedicated space is also left for the responders to express additional concerns/suggestions regarding the use of (A)MICO in the setup. Based on the outcomes of this co-design phase, four distinct multimodal feedback strategies are implemented. Figure 5.52 provides a summary of their characteristics in terms of color, 2D pattern and associated acoustic signal.

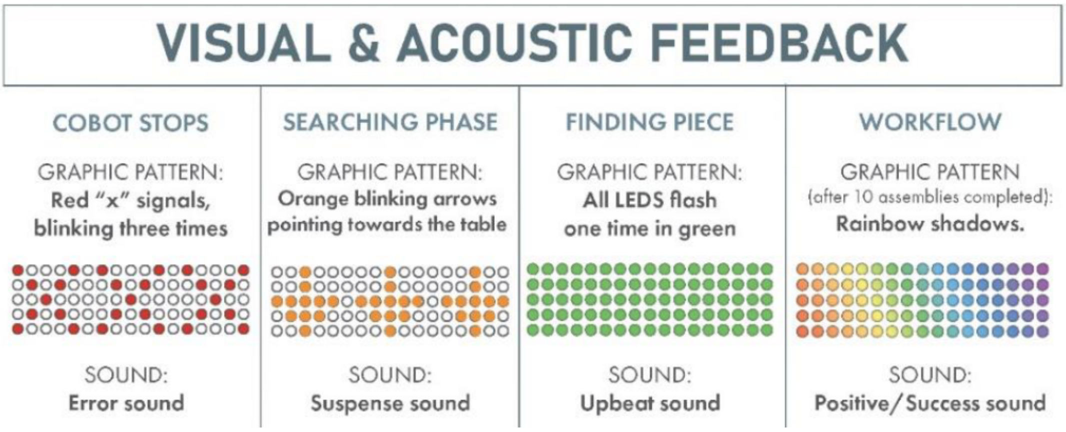


FIGURE 5.52: The four feedback strategies implemented to respond to the main needs identified by the participants in the co-design phase.

Three of the shown strategies address the critical situations mentioned above, but a fourth one is also added. In fact, responses collected in the "open comments" section of the questionnaire highlight the need for a feedback on the current performance during the task. For this purpose, an additional combination of visual and acoustic feedback is implemented as a signal sent to the user every 10 completed products. Indeed, especially for people with ASD, having frequent feedback on the ongoing work is a valuable way for reducing stress and for having a better awareness of time passing.

Testing and evaluation

To evaluate the potential of the device, the setup reported in Section 4.1.1 is integrated with (A)MICO and the collaborative assembly task described in Section 4.1.2 is once again proposed. Twelve new participants with ASD (2F and 10M) are involved. First, each subject is asked to work on the task for around 10 minutes without the help of (A)MICO. After that, additional

10 minutes are spent on the same task, but with the system integrating the proposed feedback strategies. Note that, apart from clarifications regarding the task, the researcher did not provide any explanation about the meaning of the different feedback strategies since the goal is to understand their effectiveness and intuitiveness in delivering the message. At the end of the two assembly phases, participants are asked to also take part in a semi-structured interview of around 15 minutes. The aim is to investigate quality of experience, clarity of the feedback strategies and achieved assembly proficiency regarding the specific proposed task. Here, the main results are reported:

- *Perception of the device*. Figure 5.53 shows that half of the participants have a correct understanding of the purpose of the device, even though it was not explained by the researcher during the activity. This is a positive result considering that the selected ASD participants have a reflective reasoning process, rather than intuitive. Given more time to familiarize with the task and with the functionalities of the device, the design of (A)MICO seems promising in quickly communicating its purpose to a wide cohort of individuals. Moreover, a similar proportion of the experimental group expresses a preference to work with (A)MICO rather than without. Considering those who do not find the device helpful, no expression of annoyance is annotated during the interviews since the reason for their dislike is mainly based on the fact that they did not pay attention to it during the task. This is another positive result since ASD individuals are often hypersensitive to external stimuli.

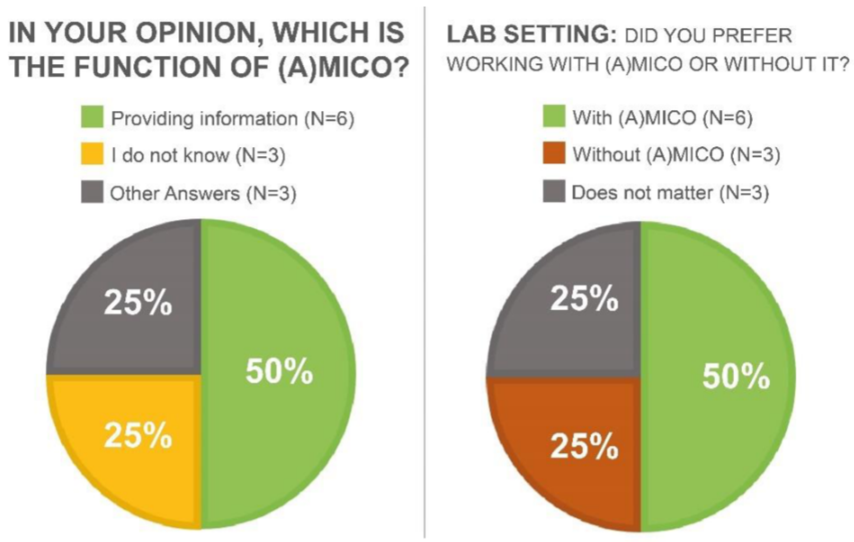


FIGURE 5.53: The answers collected during the semi-structured interviews regarding the perception of the device.

- *Understanding of the feedback*. Investigation of the level of comprehension regarding the proposed feedback strategies is of utmost importance for this study. Each participant is shown a video of the stimuli and asked to explain their meaning. Figure 5.54 summarizes the obtained results. The first stimulus on the left, representing an error state is the one achieving the highest level of correct understanding. On the contrary, the feedback related to the searching phase (the cobot waiting for the detection camera to find the next component) is the one that leads to a wider variety of wrong interpretations. The two remaining stimuli shown on the right, are intended as indicators of a positive situation and only a few participants considered them as conveying a negative message. Even though further studies and design iterations are needed to improve the effectiveness of the device (for some feedback more that others), the collected preliminary results are promising. In fact, inferring complex information from a combination of visual and acoustic stimuli is not an easy process, especially for a target group characterized by difficulties in abstract reasoning.

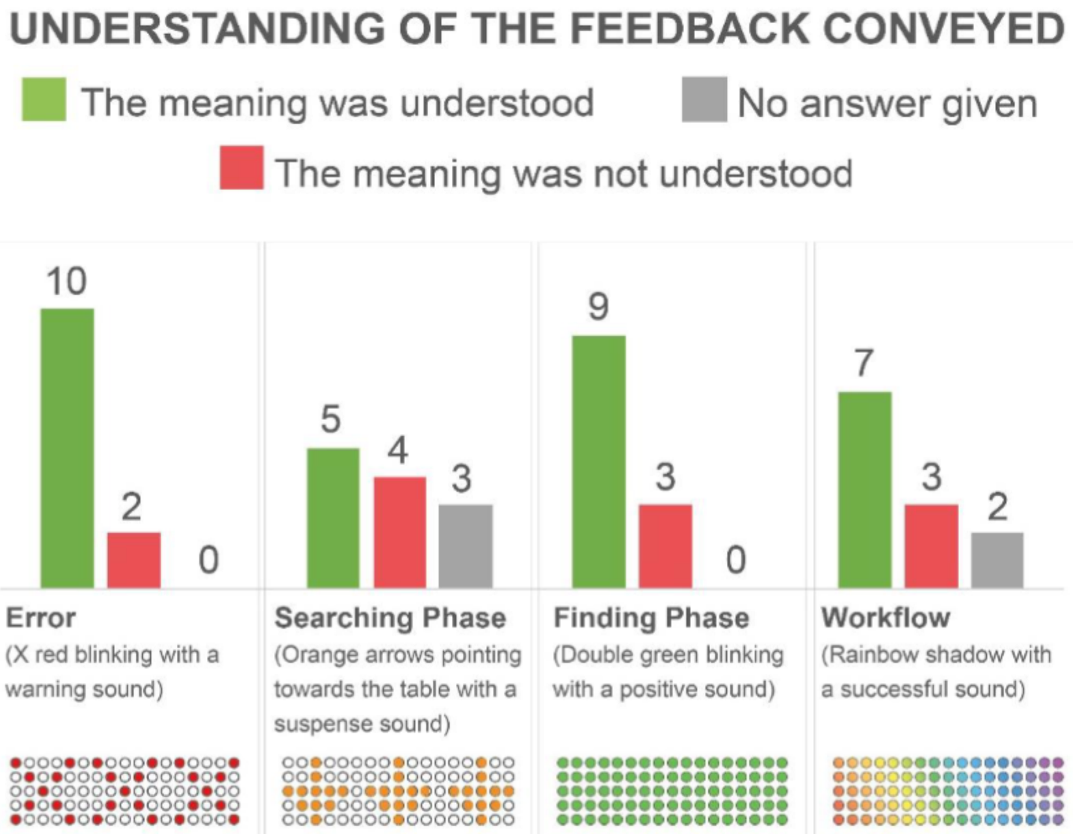


FIGURE 5.54: The answers collected during the semi-structured interviews regarding the understanding of feedback strategies.

- *General opinion on user experience*. A wide variety of considerations emerge in terms of user experience, highlighting the heterogeneity of sensitivities of users with ASD. If no relevant comments address the proposed visual patterns, different suggestions can be considered for the auditory part. For instance, some users would like (A)MICO to convey messages through actual speech synthyesis, while others would find it helpful to have some kind of musical rhythm helping them maintain the workflow and reduce the perception of tiredness. Regarding the positioning of the device, some users suggest to have it closer to them, while others would prefer to have it mounted on the robot arm, so that attention can be paid to both entities simultaneously.

Even though longer tests with a bigger experimental population would be required to have robust conclusions regarding the acceptability and effectiveness of (A)MICO in assisting ASD during HRI, promising results are achieved. However, it is important to remember that autism is a spectrum, meaning that its manifestations are varied and heterogeneous. For this reason, the functionalities of the proposed multisensorial feedback system should be expanded to provide easy customization for each single subject, adapting to their needs and sensibility.

Comparing and supporting neurotypical and ASD subjects.
Main take-aways:

The fixed and predictable routine with precise task assignments that characterizes the collaborative work with a cobot could represent a great inclusion opportunity for individuals with ASD, in line with the goals of Industry 5.0. However, it is important to remember that **systems designed to satisfy the needs of neurotypical users may not be optimal for neurodivergent ones**. Therefore, dedicated studies are required to understand the main differences and propose effective and personalized solutions.

A first observational study comparing ASD and NT users in the context of HRI reveals that **the two groups have significant differences in terms of gaze, gestures, adaptation to the task and overall resulting performance**. Moreover, focusing on the biomechanical aspects of the interaction, ASD individuals elicit slightly less efficient movements indicating higher risks of fatigue. However, all the observed differences can be addresses through a proper organization of the task, schedule and workplace. With this approach, **no contraindication for the employment of ASD workers remain: results even show that performance levels above the NT average can be achieved**.

One way to promote this inclusion opportunity is to develop new assistive solutions aligned with the concept of Design for All. Following this paradigm, a co-design session involving ASD individuals is organized to understand how a cobot can communicate with the user in an intuitive and predictable way. As a result, **a prototype of the (A)MICO multisensorial feedback system is realized and tested achieving promising results**. Further studies are required to test the effectiveness of the device over longer periods of interaction and to define the best strategy to customize the provided feedback for each specific application and user.

Chapter 6

Conclusions

The introduction of robots in every field of application is rising exponentially. It is well known that robots are exceptional solutions when it comes to repetitive operations or physically demanding and dangerous tasks since they can substitute their human counterpart, freeing them from intense labor in favor of activities requiring problem-solving and flexibility skills. However, a question arises: are we fully taking into account the needs and peculiarities of the people that will be interacting with them? Given the impact that the capillary diffusion of robots may have on our society, it is of utmost importance to study the effects that these devices may have on the users in order to understand them better and leverage this newly acquired knowledge to improve their effectiveness even further.

In fact, robotic devices are often designed to take over tasks traditionally performed by humans, risking the loss of social nuances inherent to human nature and sensitivity. To safeguard well-being, it is crucial to adopt a user-centered approach during the design and implementation of these solutions. The ultimate aim should be to replicate human-human interaction (HHI) within human-machine contexts while ensuring safety and performance. This is particularly vital in areas requiring close interaction such as work (industrial sector) and healthcare (rehabilitation sector). A comprehensive review of the state of the art, reported in Chapter 2, revealed that current advancements in sensing and computing are opening new paths for the introduction of human-centered design and socially-aware control strategies. High-level information related to the experience of the interacting user can be collected from biomechanical, physiological, social and psychological signals. However, further research is still required to better understand how this multimodal data can be fused together and effectively leveraged in the control system of robotic devices in order to offer more natural and social

HRI experiences. Additionally, the use of robotics as promoters of inclusivity for vulnerable groups, such as neurodivergent users, is going largely unexplored, highlighting the need for further studies.

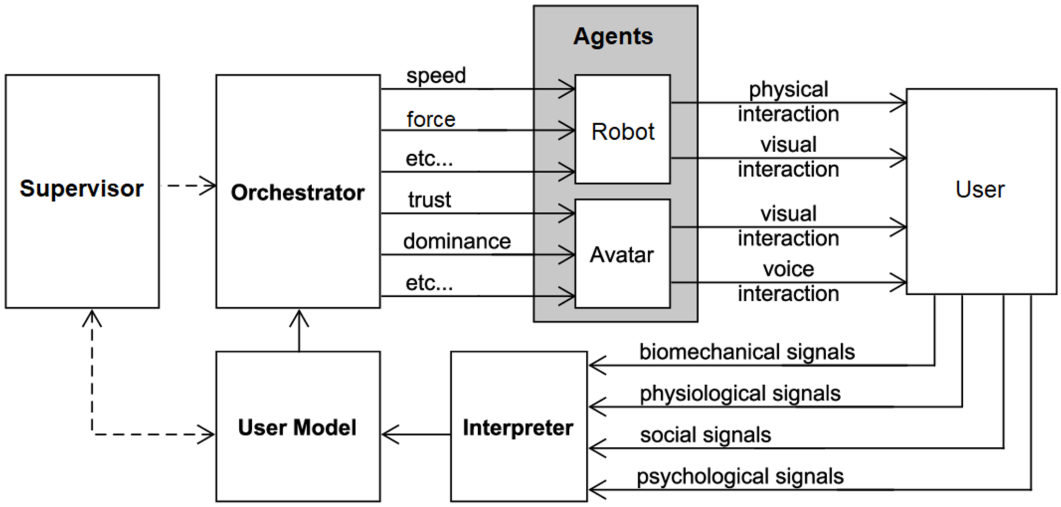
Depending on one's culture, education and experiences, a certain approach to social interactions takes place whenever two individuals interact with each other. Therefore, the first step of the present project consisted in a simplified schematization of the above mechanism to understand it better and reproduce it within the control system of automated machines. Section 3.1 went into the details of this topic, highlighting three main points of interest.

1. The real state of a person often does not coincide with the one that is expressed to the outside because of an innate *Social filter*. By merging psychological, affective and emotional expressions with measurable physiological and biomechanical reactions, an interactive entity should attempt to gain an understanding of the user's state as close as possible to the actual one. A wide array of sensors collecting heterogeneous raw signals from the user is therefore key for a successful technological reconstruction of the social mechanism.

2. One's ability to interpret the state of others is referred to as *Social monitor*. Within day-to-day interactions, these capabilities are deeply rooted among the intricate laws of human sensibility and social awareness. Exploiting existing knowledge and the power of AI, one of the biggest challenges of the present project is, in fact, achieving reliable inference of user experience.

3. Once a proper understanding of the state of a person is achieved, new challenges arise in trying to change the ways of interaction in order to offer a positive, social and natural experience. This is can be represented by a so-called *Social controller*, an expert system embedding rule-based instructions coming from both existing knowledge and the outcomes of novel experimental campaigns.

A generalized human-driven control architecture has therefore been conceptualized on the basis of the above observations. Section 3.2 and the rest of Chapter 3 presented the choices behind this architecture and the role of each foreseen module. In this regard, three additional points of interest should be highlighted:

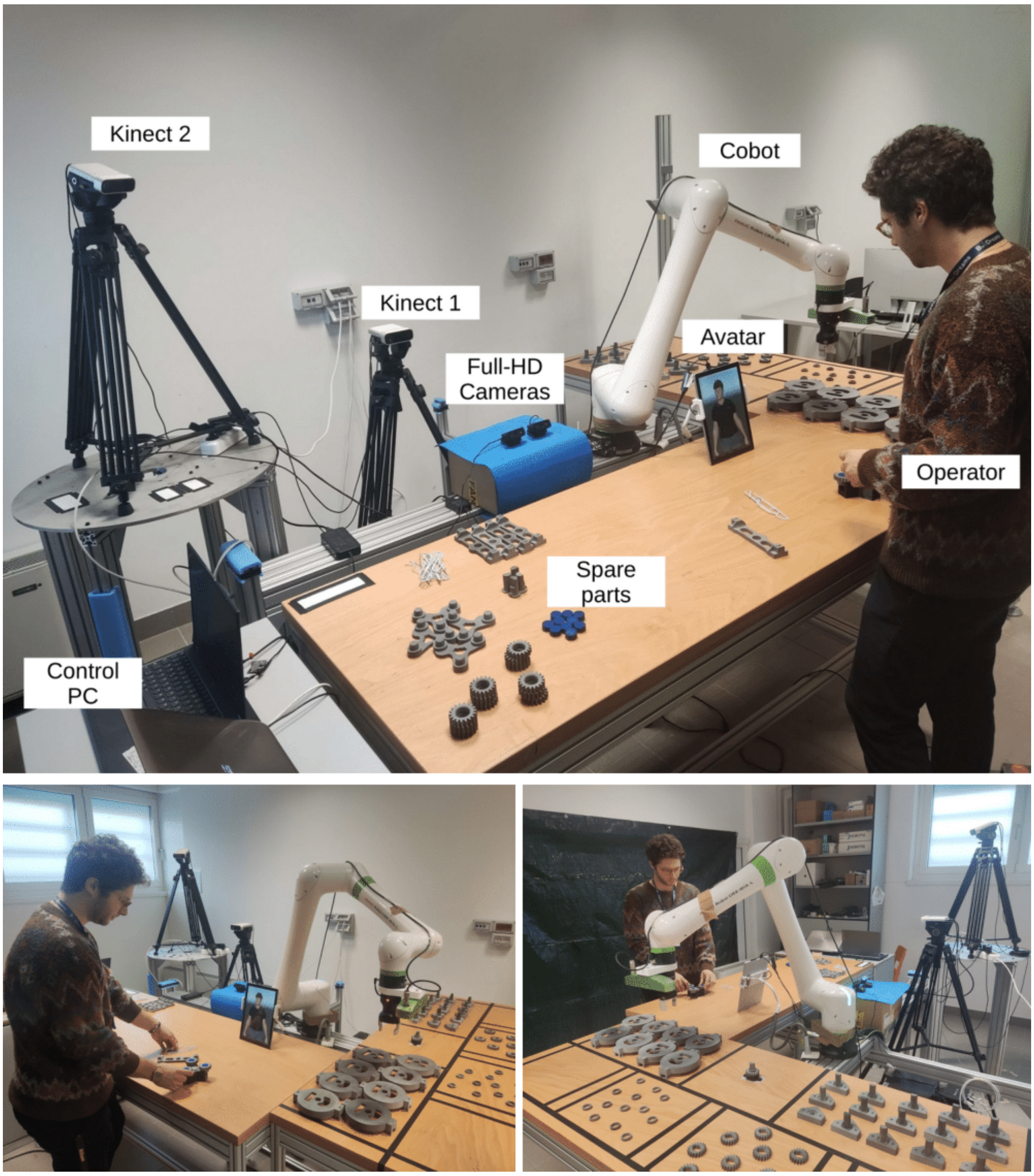


1. The interactive technological system is composed of both a robot and a virtual character, integrated together as a single entity in order to provide physical, visual and auditory interaction capabilities similar to what is experienced in human-human interactions.

2. Biomechanical, physiological, social and psychological raw signals are collected from the users through non-invasive techniques in order to gain a complete understanding of their state, ideally bypassing the uncertainties produced by social filtering. High-level interpretation of these signals and mapping upon a personalized user model are key steps to fuse this information together and achieve a robust basis for the decision-taking process of the control system.

3. The tuning of the behavior of the system is realized through an intuitive programming interface so that no specific programming skills are required for the deployment of the system. With this approach, high automation levels can be achieved providing a natural experience since the users are not required to manage the system itself, but simply to interact with it as they would with a human counterpart. However, a supervisor is still foreseen in the architecture to make sure that irreplaceable human judgment can be injected within the control loop whenever deemed necessary.

Leveraging the proposed architecture, two use-case scenarios have been setup for this project. The first application of interest can be referred to as Mind-Bot, a mental-health friendly workcell for collaborative manufacturing. The setup featured a Fanuc CRX10iA/l cobot and an interactive virtual avatar, coherently orchestrated using VSM. A wide array of sensors was added to the demonstrator and interfaced with the rest of the architecture using SSI, while ROS was chosen to build the communication framework necessary for dispatching information and commands. A 3D printed gearbox was also ideated specifically as assembly task for the purposes of this study, with characteristics aiming to promote close collaboration by design. Section 4.1 presented all the details of this deployment.



Secondly, an Empathic Neurorehabilitation Trainer was realized to test the feasibility of the proposal in a rehabilitation scenario, surely one where physical and cognitive interactions play a relevant role in the effectiveness of the deployment. The setup incorporated the PLANarm2 prototype, a robotic device developed within STIIMA-CNR to assist upper-limb planar movements even in domestic environments. To provide a visual representation of the warmth and competence of the system, a virtual character was again introduced and controlled using the same tools mentioned above. A GUI, realized using Unity3D, was added in this case to provide visual feedback of the exercises to be carried out. Together with the array of portable sensors, selected to be feasible for future home-therapy, lab-based and more invasive equipment is also interfaced with the architecture solely as a validation tool for the effectiveness and acceptability of the system. All the details regarding the implementation of the system and the available rehabilitation tasks have been addressed in details in Section 4.2.

Thanks to these two demonstrators and to the approval obtained by the involved ethical committees, a number of experimental campaigns have been carried out. The goal here was multifaceted:

- Verify the technical feasibility of the deployed systems.
- Explore specific aspects regarding the experience of human-robot interaction.
- Test the effectiveness of possible solutions aiming to make HRI more natural and social.
- Collect a list of "lessons learned" to construct a guideline that can be referenced for future expansions on this work.

Section 6.1 provides a summary of the main outcomes of these studies and tries to extrapolate the main take-away messages achieved by this project. After that, Section 6.2 attempts to draw the guidelines for future further developments in the field, highlighting both technical and organizational aspects. Finally, Section 6.3 is dedicated to the identification of the limitations of the present project and of a series of interesting research lines, worth exploring in future works.

6.1 Main take-aways

Chapter 5 was dedicated to the experimental campaigns carried out for the purposes of this project. Each section addressed one specific topic and was completed by a summary sheet reporting the main outcomes and a brief discussion. Overall, the findings highlight the multifaceted nature of human-robot interaction and emphasize the importance of personalized, adaptive and socially aware robotic systems to optimize both performance and user experience. In particular, this work investigated how various factors including biomechanics, gaze behavior, psychological measures, and the integration of virtual characters affect the perception and effectiveness of human-robot collaborations, with significant implications for both industrial and neurorehabilitation settings. A growing need for a holistic approach to HRI is underlined, not only to optimize task performance but also to prioritize user well-being, engagement, and inclusivity.

Assessing physical well-being through biomechanical monitoring

One of the most significant aspects of HRI in industrial settings is ensuring that the collaboration between humans and robots does not contribute to physical strain, fatigue, or discomfort. This thesis highlights the importance of biomechanical assessments in both laboratory and real-world production environments, underlining the limitations of studies conducted in controlled settings that do not account for the complex dynamics of live work environments. In contrast to laboratory conditions, real-world industrial settings are far more dynamic and unpredictable, presenting challenges in terms of sensorization, ergonomics, task design, and physical demands on the user. The results presented throughout Section 5.1 emphasize the importance of non-invasive, real-time assessment tools to monitor and improve user ergonomics. Technologies like the Microsoft Azure Kinect, which offer markerless motion tracking, are particularly beneficial in industrial applications due to their ability to operate with minimal setup and maintain the natural flow of interaction. The use of such technologies in collaborative robotic systems allows for continuous, real-time adaptation to the user's movements and preferences. For instance, this thesis demonstrated that robots can adjust their poses dynamically to minimize strain and improve overall user comfort, ensuring that both the robot and the human can collaborate efficiently and without causing physical harm or discomfort.

Furthermore, the analysis of neuromotor rehabilitation tasks reveals the broader potential of robotics in healthcare. The introduction of robotic assistance in rehabilitation scenarios was shown to improve exercise performance, though it did not significantly alter the underlying structure of muscular synergies. This demonstrates that robotic systems can support the physical recovery of users while respecting natural movement patterns, which is crucial for long-term rehabilitation and preventing further injury. The biomechanics of rehabilitation should continue to be a priority in future studies, as understanding the interplay between robot assistance and human movement can lead to more personalized and effective rehabilitation protocols.

Leveraging natural social cues

Gaze has been shown to be a crucial social cue in human-human interaction, serving to communicate intent, establish attention, and facilitate collaboration. In Section 5.2, gaze-based attention recognition was successfully integrated into the generalized human-driven control architecture, demonstrating its potential as a tool for enhancing social interaction with robots. In fact, the development of a gaze-based attention recognition model revealed that gaze is a key indicator of collaborative intent, with over 80% of collaborative instances preceded by gaze directed towards the robot. This supports the hypothesis that humans (specifically neurotypical subjects) instinctively use gaze as a non-verbal communication tool, even when interacting with robots. The developed model achieved an 82% F1 score, demonstrating its potential for real-time applications in collaborative robotics. Furthermore, integrating gaze behavior into the generalized human-driven control architecture enhanced the robot's ability to respond to user intentions, leading to a success rate of 89% in triggering collaboration.

However, some of the results presented in Section 5.5 demonstrate that this gaze-based robotic behavior, successfully leveraged with neurotypical subjects, may not be usable when considering neurodivergent populations. In particular, ASD individuals were observed to have different gaze patterns, rarely looking towards the robot and for shorter periods of time. These results highlight even more the need for personalized interaction strategies in HRI: broader implications lie in the potential for robots to understand and engage with humans on a more intuitive and human-like level. As robots become more integrated into daily life, the ability to interpret personalized non-verbal cues will be increasingly important in fostering seamless and natural interactions. This could result in more adaptive, efficient, and pleasant collaborative experiences, where robots are not just tools but active participants in social and professional tasks.

Supporting emotion regulation and mental well-being

The role of psychological elements in human-robot collaboration is another area where this research can offer interesting insights. As robots become more embedded in workplace environments, it is crucial to monitor not only their physical but also their emotional impact on users. Previous research has shown that emotional well-being and cognitive load play a significant role in user performance and engagement. This thesis builds on this idea by exploring how physiological signals, such as heart rate and facial expression analysis, can provide valuable feedback for the robot, enabling it to adjust its behavior to better support the user's emotional state. Interestingly, while facial expressions are often used as a common tool for emotional state analysis, the findings presented throughout Section 5.3 indicate that heart rate variability (HRV) measures are more robust and reliable indicators of perceived challenge and user engagement. HRV was found to distinguish between conditions where the user felt challenged and those where they were in a more balanced state, pointing to its potential as a real-time feedback signal for robot behavior adjustment. Moreover, subjective measures (e.g., self-reported questionnaires) further confirmed that personalized adjustments to the robot's collaboration rhythm could positively influence user experience, reducing workload and enhancing engagement.

Given the growing recognition of the importance of emotional and cognitive states in user experience, these findings suggest that robots could be designed not just to perform tasks but also to promote emotional regulation and mental well-being. For example, by monitoring HRV and other psychological signals, a robot could adjust the difficulty level of a task to prevent frustration, boredom, or stress and ultimately improving user satisfaction and task performance. Further studies could refine these measures and explore their integration into real-time control systems, with promising opportunities also for ASD users for whom the monitoring of physiological signals may represent a solution overcoming the higlighted differences in behavior and emotional/introspective expressions. A promising direction could be the development of hybrid models that combine physiological data with subjective self-reports (if applicable) to more accurately assess the user's emotional state. Such models could be employed across a range of high-stakes applications, from industrial environments to healthcare settings, where understanding and supporting the emotional needs of the user is essential.

Enhancing social interactions with virtual characters

The integration of virtual characters into human-robot systems offers another fascinating opportunity to enhance social interaction and user experience. The use of avatars or virtual characters in conjunction with physical robots has the potential to create a more cohesive and engaging user experience. This thesis investigates how the social role of the robot can be enhanced through the addition of a virtual character. In particular, Section 5.4 demostrated that the presence of an avatar, provided its behavior is convincingly modeled, can help the system embody a social role, such as a the one of a colleague or a collaborator. This is especially relevant not only in industrial applications, where this approach can promote better performance and emotional well-being, but also in contexts like neurorehabilitation, where the social engagement of the user can significantly impact their recovery process.

The BASSF behavior model, used to guide the avatar's actions, highlights the importance of emotional regulation in promoting optimal user engagement and task performance. By focusing on dimensions like dominance and guidance, the model demonstrated how avatars can be designed to encourage positive emotional states, supporting Flow during collaborative tasks. These findings align with broader research in the field of social robotics, which suggests that robots with socially aware avatars can create more natural, engaging, and supportive interactions with users. Moreover, incorporating virtual characters into robots could expand the scope of robot capabilities in neurorehabilitation. The Empathic Neurorehabilitation Trainer developed in this thesis demonstrated that avatars could enhance user performance and engagement during rehabilitation exercises without causing distractions. This suggests that virtual characters, when integrated thoughtfully, can support users not only by guiding physical tasks but also by providing emotional support, further promoting engagement and progress.

Promoting inclusion for neurodivergent users

A particularly innovative aspect of this research was presented in Section 5.5, where HRI is explored as a tool for promoting the inclusion of individuals with the Autism Spectrum Disorder in collaborative work environments. As the goals of Industry 5.0 emphasize human-centric and inclusive production systems, this research highlights the potential for cobots to offer unique opportunities for neurodivergent individuals, particularly those with ASD. Both quantitative and qualitative analyses revealed significant differences between ASD and neurotypical users in terms of gaze, gestures, and task adaptation. These differences, however, could be mitigated through thoughtful task structuring, careful scheduling, and a personalized approach to collaboration.

Interestingly, individuals with ASD exhibited slightly less efficient movements during collaboration, which could indicate higher risks of fatigue. However, these differences did not preclude their ability to perform well, especially when tasks were tailored to their needs. By designing systems that are flexible and responsive to the specific challenges of neurodivergent users, it is possible to create work environments that are both inclusive and optimized for performance and where individuals with ASD could not only meet but exceed the performance of their NT counterparts. This aligns with the principles of "Design for All," which advocates for the creation of systems that are accessible to diverse user groups. The development of assistive technologies like the (A)MICO multisensory feedback system represents a promising step toward achieving this goal. By providing intuitive, predictable communication between the robot and the user, this system can improve the user experience and facilitate more effective collaboration. Further studies should focus on refining these systems and testing them over extended periods to ensure their effectiveness in real-world environments.

This thesis contributes to the growing body of research on human-robot interaction by exploring **the integration of biomechanical, behavioral, emotional, and psychological measures into a unified robotic framework**. By employing non-invasive technologies, leveraging gaze behavior as a social cue, monitoring physiological and psychological responses, and incorporating virtual avatars, this work **contributes to the development of more personalized, adaptive, and socially aware robotic systems**. Furthermore, including neurodivergent individuals in the design process further enriches this vision, aligning with the inclusive goals of Industry 5.0 and **demonstrating that robots can promote both performance and well-being in a diverse range of users**.

The findings presented here underscore the importance of personalization and social awareness in HRI, showing that robots can and should be designed to not only perform tasks but also engage with users in ways that promote emotional well-being, inclusion, and task success. As robots continue to play an increasingly central role in various industries, future work should further explore the integration of these insights into real-world scenarios, such as collaborative workcells or neurorehabilitation applications, ensuring that collaborative robots can serve as effective, supportive, and inclusive partners for all users.

6.2 Guidelines for replication

The contents and outcomes of this thesis required a significant effort in overcoming both organizational and technical issues. Building upon the lessons learned up to now could allow future studies and replications to avoid unnecessary mistakes and optimize the research flow. For this purpose, here a list of guidelines is reported based on the experience collected during the doctoral program.

Organizational guidelines

- Considering realistic use-case scenarios that could significantly benefit from the study outcomes is key to promote the impact of research outside the laboratory walls. In fact, a review of the state of the art revealed that the results obtained in controlled environments are often not representative of the dynamism and complexity of real-life applications.

- When designing research questions, study protocols and experimental campaigns, it is important to extend as much as possible the range of involved users in order to mitigate the risk of biased generalization. In fact, some of the obtained outcomes may be applicable to the studied populations but not as much for other overlooked groups. As a notable example, the use of gaze as a promoter of natural HRI was demonstrated to be very effective for neurotypical subjects but unfeasible for ASD individuals.
- A rich and multidisciplinary research consortium is key to tackle complex issues from a heterogeneous set of points of view. In fact, as an author with technical backgrounds, the collaboration with colleagues coming from diverse backgrounds proved itself necessary to avoid overlooking important aspects that would have been deprioritized otherwise.
- A complete knowledge of the requirements dictated by the need of an ethical approval can help in speeding up the clearance process. In the case of this project, the support of a clinical partner was fundamental both for the involvement of ASD subjects and for the addresses rehabilitation scenarios.

Technical guidelines

- At a research level, selecting robotic devices that can be freely controlled at a low-level can speed up the integration of promising solutions into working demonstrators. For instance, the implementation of all the software interfaces needed to use the Fanuc CRX10iA/l cobot in the MindBot project required quite some time compared to the steps needed to integrate the PLANarm2 prototype in the Empathic Neurorehabilitation Trainer.
- Using tools that ease the process of sharing and replaying collected data opens up the possibility of decentralizing development, calibration, testing and analyzing processes, a relevant aspect when working with international teams from different laboratories.

- Building local networks can help sharing computational loads onto different machines, avoiding unnecessary bottlenecks and achieving better system performances overall. Of course, a distributed systems requires a tool capable of effectively exchanging information and commands. In this project, ROS was chosen as state-of-the-art solution for robotic applications. Future studies could even consider a migration to the newer ROS2 version, which is becoming more and more stable and provides powerful solutions to ease the implementation process.
- When selecting sensors, one should consider that some are only applicable in controlled environments. Consequently, this kind of tools could be leveraged for evaluation and validation purposes but not for their implementation as real-time data providers since their use would not be feasible in a realistic environment.
- Considering physiological data, the above consideration can be extended even further. Collecting EMG signals requires long setup times, experience for a precise positioning and often becomes invasive for the subject. On the other hand, EDA sensors are often easy to install, but the data they produce is heavily affected by movements, hindering the replicability in real-life applications. Extremely interesting insights can be obtained by analyzing the brain activity of individuals. However, most EEG sensors present both of the above limitations: they require long setup times and do not allow the subject to freely move around.
- For easy installation, reduced invasiveness and replicability in realistic environments, one should prefer to base real-time adaptive systems onto data related to heart activity, as available devices (e.g., chestbands) can be worn quickly and autonomously by the subject with minimal worries related to noise disrupting the collected data. Similar considerations can be done regarding marker-less vision systems, both in terms of RGB and depth cameras. In fact, provided that their positioning is redundant and robust enough for the specific application, they can be effectively leveraged to extract useful information such as biomechanical assessments, facial expressions, gaze behaviors, etc.
- Self-reported questionnaires are also useful for a complete representation of the user's status. However, one should remember that they are asynchronous tools that could represent a distraction for the subject if administered during the task under analysis.

6.3 Limitations and future work

When working on experimental campaigns, often some limitations must be accepted. In the case of the present thesis, some of the analyses carried out for the MindBot project required participants to work on the task for extensive periods of time and over multiple days. Thanks to this approach a rich and varied dataset can be collected for each subject but, on the other hand the size of the analyzed population must be limited. The reason is twofold: first, there is the organizational complexity of recruiting participants willing to reschedule their day-to-day activities in order to take part in such a study and second, the need for supervision during this kind of experimental activities represents a significant time burden for the researchers and for the lab availability as well. Similarly, when working with ASD individuals, recruitment is a problematic task. Collaboration with dedicated associations can ease the process but still, the reachable population is limited and must be further filtered depending on the portion of the spectrum suitable for the purpose of the study. As a consequence, some of the obtained results could be rendered statistically more robust if, over time, data related to bigger and bigger populations were to be collected.

Similar considerations can be done also for the activities related to the Empathic Neurorehabilitation Trainer. Due to ethical limitations, the involvement of patients was not foreseen for the planned experimental campaigns. However, in order to test the effectiveness of the system on the outcomes of actual therapies, future works should transition from the lab to the clinics or, even better, to the homes of patients in order to validate the feasibility of the solution also for domestic and unsupervised environments.

Overall, the generalized human-driven control architecture proposed in this thesis proved itself effective in promoting natural and social human-robot interaction experiences both for industrial and rehabilitation use-cases. In this sense, it would be interesting to quantitatively evaluate how easily and effectively the proposed demonstrators could be repurposed for different application or even modified to accommodate different robots and sensors. Moreover, many aspects of human-robot interaction still need to be explored in detail. The outputs of this thesis provide a series of tools that can ease future studies, but research cannot end here. The answer to one research question naturally opens up many more and only a continuous push can provide the knowledge that is needed to keep improving automatic systems by providing growing levels of awareness and personalization.

Appendix A

MindBot Consortium

FIGURE A.1: A picture taken during one of the MindBot General Assemblies depicting the group of researchers involved in the project.

- *IRCCS – Associazione La Nostra Famiglia 'Istituto Scientifico Eugenio Medea'*. Coordinator of the project and expert for the involvement and monitoring of ASD participants (MEDEA, 2024).
- *Università degli Studi di Milano*. Contributing with its psychology department for the experience questionnaire collection and analysis (UMIL, 2024).
- *Consiglio Nazionale delle Ricerche*. The STIIMA institute (to which the author of the present thesis is affiliated) was in charge of providing the technical skills necessary to implement the overall software and hardware architecture and in particular to realize the foreseen adaptive robot control (CNR, 2024).

- *BioRICS NV BE*. A company based in Belgium responsible for the mental energy interpreter module based on their proprietary Mindstretch app (BIORICS, 2024).
- *Deutsches Forschungszentrum fur Künstliche Intelligenz Gmbh*. The german research center for artificial intelligence, involved as expert regarding avatar control and cognitive user modeling (DFKI, 2024).
- *Sveuciliste u Rijeci, Filozofski Fakultet u Rijeci*. The University of Rieka, responsible for the analysis and assessment of the organizational impact that the presented system may have on small and medium enterprises (FFRI, 2024).
- *KUKA Deutschland Gmbh*. Involved to perform a baseline analysis of work environment and organizational specifications together with the identification of companies willing to test the system in their premises (KUKA, 2024).
- *Universitaet Augsburg*. Its Affective Computing Research group was the main responsible for the social and affective signals interpretation layer of the control architecture (UAU, 2024).
- *Republic of Croatia – Ministry of Labour, Pension System, Family and Social Policy*. Collaborating with the University of Rieka for an employment assessment of the current industrial scenario (MRMSOSP, 2024).

Appendix B

MindBot ROS Communication Pipelines

Nodes	Topics	Services
/mindbot_service_requester This node is created by the VSM module and is used to communicate with the RobotControl module via services	----------------------	/mindbot/robot/action_done (server) /mindbot/robot/set_ctrl_state (client) /mindbot/robot/set_ctrl_mode (client) /mindbot/robot/set_detection (client) /mindbot/robot/set_gripper_action (client) /mindbot/robot/set_joint_target (client) /mindbot/robot/set_tcp_target (client) /mindbot/robot/set_max_tcp_acc (client) /mindbot/robot/set_max_tcp_vel (client) /mindbot/robot/set_min_clearance (client)
/mindbot_topic_subscriber This node is created by the VSM module and is used to read the topics published by the RobotControl	/mindbot/robot/ctrl_state (subscribe) /mindbot/robot/ctrl_mode (subscribe) /mindbot/robot/tcp_state (subscribe)	----------------------

/crx10ial This node is created by the RobotControl module and is responsible for the direct integration with the controller of the Fanuc robot using the RMI interface	/mindbot/robot/joint_states (publish, 20Hz) /mindbot/robot/robot_state (publish, 20Hz) /tf (publish, 40Hz) /tf_static (publish, latching) /rmi_controller/result (publish, 20Hz) /rmi_controller/goal (subscribe)	/controller_manager/load (server) /controller_manager/unload (server)
/mindbot_manager_node This node is created by the RobotControl module and is responsible for the dispatching of all the commands to the correct hardware modules (robot, gripper, detection camera)	/mindbot/robot/ctrl_state (publish, 1Hz) /mindbot/robot/ctrl_mode (publish, 1Hz) /mindbot/robot/tcp_state (publish, 15Hz) /human/body_frames (subscribe)	/mindbot/robot/action_done (client) /mindbot/robot/set_ctrl_state (server) /mindbot/robot/set_ctrl_mode (server) /mindbot/robot/set_detection (server) /mindbot/robot/set_gripper_action (server) /mindbot/robot/set_joint_target (server) /mindbot/robot/set_tcp_target (server) /mindbot/robot/set_max_tcp_acc (server) /mindbot/robot/set_max_tcp_vel (server) /mindbot/robot/set_min_clearance (server)
/pickit This node is created by the RobotControl module and is responsible for the direct integration with the controller of the Pickit3D detection camera	/tf (subscriber) /tf_static (subscriber)	/pickit/product/load (server) /pickit/setup/load (server) /pickit/check_for_object (server)
/ssi This node is created by the Analysis of Social and Affective Cues module and is used to publish the inferred data to the ROS framework	/ssi/emotion (publish, 10Hz) /ssi/pain (publish, 10Hz) /ssi/attention (publish, 10Hz)	---------------------

/kinect This node is created by the Image Processing module and is responsible for sharing all the information related to the operator's skeleton	/kinect_0/skeletons (publish, 12Hz) /kinect_1/skeletons (publish, 12Hz)	---------------------
/biomech This node is created by the Biomechanical module and is used to publish the computed indexes to the ROS framework.	/kinect_0/skeletons (subscribe) /kinect_1/skeletons (subscribe) /human/body_frames (publish, 12Hz) /human/fatigue (publish, 2.5Hz) /human/fatigue_index (publish, 2.5Hz) /biomech/joints (publish, 2.5Hz)	---------------------

Appendix C

Ethical Clearance

Given the study protocol provided for the project under the name "Modeling perception of human-robot interaction: toward natural and social HRI experiences" together with all the accompanying documentation, the approval granted from the ethical committee of STIIMA–CNR is reported below (in italian).

The present study does not have any diagnostic or clinical goal. However, robot prototypes developed for rehabilitation purposes are used in the experimental campaigns as tools useful to test the users' experience of interaction in a setting resembling that of a neurorehabilitation session. Moreover, the study foresees the involvement of participants characterized by ASD as a particular group of interest for the study given the specific needs and behaviors that they may elicit when interacting with a robotic device. Given the aspects highlighted above, the approval of the ethical committee of a clinical partner was also requested and granted to complement the positive feedback already obtained by STIIMA–CNR. In particular, Dr. Eng. Fabio Storm, a researcher for IRCCS Eugenio Medea La Nostra Famiglia collaborating on the project, presented the outline of the study to the committee and obtained the approval reported below (in italian).

PARERE DI *ETHICAL CLEARANCE*

PREMESSA

La Commissione per l'Etica e l'Integrità della Ricerca, considerate le proprie competenze attribuite dal Decreto del Presidente del CNR del 23 settembre 2019 – prot. n. 0065527/2019 e s.m.i., dando seguito a una richiesta di valutazione etica da parte dell'**Istituto dei Sistemi e Tecnologie Industriali Intelligenti per il Manifatturiero Avanzato** (STIIMA-CNR), ha preso visione e analizzato natura, obiettivi e modalità di svolgimento dello studio **"Modeling perception of human-robot interaction: towards natural and social HRI experiences"**. Sono partner dello studio: l'Istituto di Tecnologie Biomediche del CNR (ITB-CNR); l'Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Eugenio Medea-Associazione "La nostra famiglia" (MEDEA); l'azienda privata AuticonM; l'università tedesca di Augusta (Lehrstuhl für Menschzentrierte Künstliche Intelligenz); la partnership no profit pubblico-privato Deutsches Forschungszentrum fuer Kunstliche Intelligenz GmbH.

DESCRIZIONE

Il progetto ha lo scopo di identificare e quantificare alcuni aspetti chiave dell'interazione tra uomo e robot al fine di rendere tali interazioni sempre più naturali, sociali e sicure. In particolare, il progetto si propone di elaborare modelli generali di interazione uomo-robot per lo sviluppo di algoritmi di controllo mirati all'ottimizzazione dell'esperienza dell'utente. Tali modelli sono generalizzati a partire da gruppi di utenti omogenei, ovvero adulti neuro-tipici e adulti con una diagnosi di Disordine dello Spettro Autistico (ASD) ad alto funzionamento (ovvero privi di limitazioni di tipo cognitivo).

Lo studio intende contribuire all'inserimento dei robot nella vita quotidiana non solo quali ausilii nell'attività umana, ma anche quali dispositivi con una capacità di interazione sociale in grado di simulare quanto più possibile l'esperienza tipica delle relazioni umane. Lo studio prevede tre fasi sperimentali principali, per una durata complessiva di almeno tre anni. Al fine della conduzione delle attività sperimentali è arruolato un campione minimo di 60 partecipanti, cui potranno essere aggiunti ulteriori soggetti in funzione del soddisfacimento della soglia statistica richiesta per alcuni parametri dello studio.

A ognuno dei partecipanti sono proposte una o più esperienze di interazione diretta con un dispositivo robotico oppure attraverso schermo o tablet. Nel corso dell'interazione con i dispositivi robotici sono raccolti segnali cineto-dinamici, biomeccanici, sociali, fisiologici e psicologici degli utenti (parametri quantitativi in grado di definire lo stato complessivo dell'utente) e questionari al fine di valutare la qualità della relazione percepita[1]. I dati raccolti ed elaborati nella prima fase sono successivamente analizzati al fine di sviluppare modelli generalizzati di interazione uomo-robot che potranno essere impiegati per la previsione delle reazioni di un 'utente tipo' dal punto di vista sociale, fisiologico e psicologico in risposta al variare del comportamento del dispositivo robotico e delle caratteristiche del *task* proposto. Nella fase conclusiva, tali modelli generalizzati sono impiegati per lo sviluppo di algoritmi di controllo del comportamento del dispositivo robotico in

[1] Tali dati possono includere: Dati di posizione, velocità e forza scambiata con il dispositivo; Registrazione audio/video; Eye tracking; Elettroencefalogramma (EEG); Elettromiografia (EMG); Attività elettrodermica (EDA); Elettrocardio-gramma (ECG); Frequenza cardiaca (BVP); Saturazione ossigeno (SpO2); Respirazione (PZT). Tra i parametri valutati nei questionari vi sono: Fatica percepita; Performance cognitive; Stato emotivo; Usabilità e comfort; Attitudine/rapporto con la tecnologia; Ansia legata all'uso della tecnologia.

funzione di alcuni parametri inferiti dallo stato attuale dell'utente, al fine di ottenere un livello di interazione ottimale tra uomo e robot. A tal fine sono impiegate tecniche di *machine learning* grazie alla collaborazione con il partner Deutsches Forschungszentrum fuer Kunstliche Intelligenz GmbH. Tutti i dati acquisiti durante le sessioni sperimentali sono acquisiti in forma anonima e criptata e sono conservati fino al raggiungimento delle finalità di progetto, raggiunte le quali l'intero database sarà eliminato in modo definitivo.

ETHICAL CLEARANCE

La Commissione,

valutato che:

- Il progetto si pone in continuità con il precedente progetto "MindBot-Mental Health promotion of cobot Workers in Industry" in merito al quale la Commissione per l'Etica e l'Integrità nella Ricerca ha rilasciato un parere di *Ethical Clearance* in data 20 luglio 2021 (prot. n. 0051763/2021);
- le attività di ricerca svolte nel precedente progetto non hanno evidenziato criticità etiche e lo studio in esame è condotto con metodologie analoghe e prevede il coinvolgimento dello stesso gruppo di ricerca;
- l'arruolamento prevede il coinvolgimento tra i partecipanti anche di personale CNR non strutturato e che tuttavia non vi sono rapporti gerarchici diretti tra i partecipanti e il responsabile scientifico del progetto, e non vi sono comunque circostanze tali da generare il rischio di induzione del consenso;
- tutti i dispositivi commerciali utilizzati nello studio sono dotati delle apposite dichiarazioni di conformità e marcatura CE[2], e che per i due dispositivi prototipali (PLANarm2 e PhiCube) è stata effettuata una valutazione del rischio secondo i parametri definiti dalla norma UNI CEI EN ISO 14971: 2012;
- il progetto non è diretto alla validazione dei dispositivi PLANarm2 e PhiCube in ambito clinico né vengono effettuate valutazioni di tipo clinico dei partecipanti sulla base dei dati raccolti;
- i risultati dei questionari e delle analisi dei dati sono elaborati, pubblicati e diffusi in forma anonima e aggregata;
- è fornita ai partecipanti adeguata informativa per l'espressione del consenso, compresa l'esplicitazione del diritto a revocarlo senza che derivi alcuna conseguenza;
- è resa ai partecipanti specifica informativa sul trattamento dei dati personali;
- è ragionevolmente esclusa la possibilità che i risultati siano impiegati per scopi che esulino dalle finalità del progetto;
- i partner internazionali hanno formalmente dichiarato che: i) le proprie attività di ricerca si svolgono nel rispetto della normativa locale vigente e sono compatibili con la normativa UE; ii) non si prevedono criticità etiche particolari relative alle modalità di arruolamento dei partecipanti; iii) sono adottate specifiche misure tecniche e organizzative per garantire la riservatezza e la protezione dei dati personali; iv) in caso di criticità etiche emergenti o di mutamenti normativi locali, ne sarà data tempestiva informazione al Coordinatore di progetto e tramite questi al Coordinatore della Commissione per l'Etica e l'Integrità nella Ricerca,

preso atto:

[2] I dispositivi commerciali utilizzati nello studio sono il: (i) HasoMed RehaStim FANUC CRX10iA/L; (ii) Robotiq 2F Adaptive Gripper; (iii) Robotiq Hand-E Gripper; (iv) QBRobotics SoftHand Research Gripper e sono dotati delle seguenti dichiarazioni di conformita *Council Directive* 93/42/EECe (i); ISO 10218-1 (ii); EN ISO 12100:2010 (iii-iv); ISO/TS 15066 (v).

- delle dichiarazioni: i) del Direttore di STIIMA-CNR attestante la conformità del trattamento dei dati personali svolta nell'ambito del progetto alle disposizioni di cui alle normative europee e nazionali applicabili dichiarazione resa in data 26 aprile 2022; ii) del Responsabile scientifico del progetto con cui si impegna ad aderire alle Regole deontologiche per i trattamenti a fini statistici o di ricerca scientifica pubblicate ai sensi dell'art. 20, comma 4, del d.lgs. 10 agosto 2018, n. 101- 19 dicembre 2018, resa in data 29 luglio 2022;
- della validazione del documento di valutazione preliminare di impatto (DPIA) da parte del Direttore di Istituto (4 novembre 2022) a seguito della valutazione positiva espressa dal Responsabile Protezione Dati del CNR e dal Corrispondente presso il Dipartimento DIITET;
- del parere favorevole del Comitato Etico dell'I.R.C.C.S. Eugenio Medea, sez. Scientifica dell'Associazione "La Nostra Famiglia" in merito allo studio "GiocAbile" che impiegava il dispositivo PhiCube di cui al prot. N. 89/21-CE,

<u>precisato inoltre che</u>:
- il presente parere etico non solleva in ogni caso i ricercatori dalla responsabilità legale connessa al trattamento dei dati personali e alla conduzione del progetto;
- il valore autorizzativo del presente parere etico è subordinato all'approvazione del progetto da parte del Comitato Etico dell'IRCCS Eugenio Medea (MEDEA), in particolare in ragione dell'impiego dei dispositivi prototipali summenzionati,

<u>con la richiesta che</u>:
- sia trasmesso alla Commissione il parere autorizzativo non appena rilasciato dal Comitato Etico dell'IRCCS Eugenio Medea (MEDEA);
- in caso di arruolamento di ulteriori partecipanti oltre al numero attualmente previsto di 60, ne sia data semplice comunicazione alla Commissione per l'Etica e l'integrità nella Ricerca;
- al paragrafo "Soggetti" del protocollo sperimentale (che prevede che ogni volontario possa ritirarsi dallo studio) sia eliminata la richiesta di una dichiarazione sui motivi del ritiro o che la procedura sia riformulata nel senso che resti nella facoltà del partecipante rispondere o meno alla richiesta;
- nel foglio informativo, al paragrafo "Comitato etico" sia corretta la denominazione in "Commissione per l'Etica e l'Integrità nella Ricerca del CNR";
- sia esplicitato anche nell'informativa sul trattamento dei dati personali che lo studio non ha in alcun modo finalità diagnostiche o cliniche;
- nella conduzione del progetto e nella pubblicazione dei risultati vengano rispettati i principi di integrità nella ricerca di cui alle *Linee guida per l'integrità nella ricerca* della Commissione per l'Etica e l'integrità nella Ricerca del CNR[3] nonché allo *European Code of Conduct for Research Integrity* di ALLEA[4];
- nella pubblicazione dei risultati e nella loro diffusione sia adottato uno stile espositivo improntato alla chiarezza, onestà, obiettività, rigore e trasparenza.

<u>Tutto ciò considerato</u>, <u>richiamando le richieste più sopra evidenziate</u>, la Commissione per quanto di propria competenza, <u>approva lo studio in esame</u>.

Per la Commissione per l'Etica e l'Integrità nella Ricerca,

il Coordinatore

[3] https://www.cnr.it/it/ethics.
[4] https://allea.org/code-of-conduct.

IRCCS EUGENIO MEDEA
LA NOSTRA FAMIGLIA
CURA RIABILITAZIONE E RICERCA DALLA PARTE DEI BAMBINI

Sede Centrale
Via don Luigi Monza, 20
23842 Bosisio Parini (LC)
tel. +39 031 877111
fax +39 031 877499
medea@lanostrafamiglia.it
segreteria.scientifica@pec.emedea.it
www.emedea.it

Prot. N. 08/23 - CE

Bosisio Parini, l'1 febbraio 2023

Gentilissimo ing. **Fabio Storm**

E pc: al Direttore Scientifico
al Direttore di Polo
al Coordinatore dell'Area di ricerca
al Referente Amministrativo della Ricerca
alla Segreteria Scientifica Centrale

PROTOCOLLO di STUDIO:

TITOLO:	**Modeling perception of human-robot interaction: towards natural and social HRI experiences**
SPERIMENTATORE PRINCIPALE:	**Ing. Fabio Storm**
PROMOTORE:	CNR STIIMA– via Previati 1/E, Lecco (LC)
TIPOLOGIA:	PROGETTO DI RICERCA
FINANZIAMENTO:	Ricerca Spontanea
IDENTIFICATIVO E VERSIONE:	Id. 993 - Versione 1.0 del 22.04.2022

Il Comitato Etico dell'I.R.C.C.S. Eugenio Medea, sez. Scientifica dell'Associazione "La Nostra Famiglia", si è riunito il giorno 26 gennaio 2023, in via telematica, con la seguente composizione:

COMPONENTI	FUNZIONI	P/A
Dr. Paolo AROSIO	*Presidente*	PRESENTE
Dr.ssa Graziella UZIEL	*Vicepresidente e clinico*	PRESENTE
Dr.ssa Sara GALBIATI	*Clinico e Sostituto Permanente del Direttore Sanitario Centrale*	PRESENTE
Dr.ssa Eleonora MAINO	*Clinico*	ASSENTE GIUSTIFICATA
Dr. Maurizio RAVERA	*Medico di medicina territoriale*	PRESENTE
Dr. Agostino SILVA	*Pediatra*	ASSENTE GIUSTIFICATO
Dr.ssa Eliana RULLI	*Biostatistico*	PRESENTE
Dr. Marco POZZI	*Farmacologo*	PRESENTE
Dr. Giovanni BORIN	*Farmacista del SSR*	PRESENTE
Avv. Anna Paola MANFREDI	*Esperto in materia giuridica*	ASSENTE GIUSTIFICATA
Dr. Andrea LAVAZZA	*Esperto di bioetica*	PRESENTE

Dr.ssa Francesca VILLANOVA	*Rappresentante area professioni sanitarie*	PRESENTE
Sig.a Anna Lisa NOVATI	*Rappresentante dell'associazionismo dei genitori*	PRESENTE
Ing. Giuseppe ANDREONI	*Esperto in dispositivi medici*	ASSENTE GIUSTIFICATO
Prof. Giuseppe BORSANI	*Esperto in genetica*	PRESENTE
Ing. Paola GRIGIONI	*Ingegnere Clinico*	PRESENTE
Dr.ssa Maria Teresa BASSI	*Direttore Scientifico Centrale*	PRESENTE
Dr. Massimo MOLTENI	*Direttore Sanitario Centrale*	ASSENTE GIUSTIFICATO
Dr.ssa Cristina LONGONI	*Farmacista della Struttura*	PRESENTE
Dr.ssa Noemi MISCIOSCIA	*Segreteria*	PRESENTE

VERIFICATA la presenza del numero legale ed esaminata la documentazione di seguito elencata:

N.	TIPOLOGIA DOCUMENTO	VERSIONE e DATA
1	LETTERA AL CE	*28.11.2022*
2	DOCUMENTO SULLE ATTIVITÀ DI TRATTAMENTO DEI DATI PERSONALI PER LE ATTIVITÀ DI RICERCA	/
3	PARERE DI ETHICAL CLEARANCE - COMMISSIONE PER L'ETICA E L'INTEGRITÀ DELLA RICERCA	*23.11.2022*
4	PROTOCOLLO SPERIMENTALE PER LO STUDIO	*Versione n.1.0 del 22.04.2022*
5	MODELING PERCEPTION OF HUMAN-ROBOT INTERACTION: TOWARDS NATURAL AND SOCIAL HRI EXPERIENCES	/
6	ALLEGATO 1 - FOGLIO INFORMATIVO	/
7	ALLEGATO 2 - INFORMATIVA PER IL TRATTAMENTO DEI DATI PERSONALI	/
8	ALLEGATO 3 - CONSENSO INFORMATO	/
9	ALLEGATO 4 - DATA PROCESSOR AGREEMENT	/
10	ALLEGATO 5 - NOMINA DESIGNATO AL TRATTAMENTO	/
11	ALLEGATO 6 - Dichiarazione di Impegno alla Riservatezza	/
12	ALLEGATO 7 - Dichiarazione privacy PI	*29/07/2022*
13	ALLEGATO 8 - Dichiarazione privacy Direttore	*29/07/2022*
14	ALLEGATO 9 - Elenco questionari	/
15	ALLEGATO 10 - Statement of compliance DFKI	*25.08.2022*

16	ALLEGATO 11 - Statement of compleance UA	*25.08.2022*
17	ALLEGATO A - Brochure FANUC CRX10iAL	/
18	ALLEGATO B - Dichiarazione di conformità Robotiq 2F Adaptive Gripper	/
19	ALLEGATO C - Dichiarazione di conformità Robotiq Hand-E Gripper	/
20	ALLEGATO D - Brochure QBRobotics SogtHand Gripper	/
21	ALLEGATO E - Manuale d_uso PLANarm2	/
22	ALLEGATO F - Manuale d_uso PhiCube	/
23	ALLEGATO G - Analisi dei rischi PLANarm2	*Versione 1.0*
24	ALLEGATO H - Analisi dei rischi PhiCube	*Versione 1.0*
25	ALLEGATO I - Protocollo Polimi PLANarm2	*22.05.2020*
26	ALLEGATO L - Protocollo GiocAbile PhiCube	*Versione 1.0 del 23/06/2021*
27	ALLEGATO M - Approvazione Polimi PLANarm2	*Parere n. 10/2020 del 28/05/2020*
28	ALLEGATO N - Approvazione GiocAbile PhiCube	*Prot. N. 89/21-CE del 12.11.2021*
29	ALLEGATO O - Manuale d'uso RehaStim	/

A seguito di valutazione e tenuto conto del parere di PARERE DI ETHICAL CLEARANCE espresso dalla Commissione per l'Etica e l'Integrità della Ricerca del CNR del 23.11.2022, il Comitato Etico ha espresso **PARERE FAVOREVOLE** allo svolgimento dello studio, rilevando che il presente protocollo è rispettoso dei principi etici dell'Istituzione e della rilevante normativa vigente.

Il parere sopra espresso si intende limitato esclusivamente alle versioni citate ed alla documentazione presentata. Ogni variazione della stessa deve essere obbligatoriamente sottoposta al parere di questo Comitato Etico.

Al Responsabile dello studio si rammenta quanto segue:

- comunicare per iscritto la DATA DI INIZIO E DI CONCLUSIONE dello studio, come pure della sua eventuale SOSPENSIONE o CONCLUSIONE ANTICIPATA con l'indicazione dei relativi motivi;
- comunicare per iscritto eventuali PROROGHE alla chiusura dello studio;
- condurre il Progetto secondo le MODALITÀ indicate;
- NON introdurre VARIAZIONI al protocollo senza che il Comitato Etico competente abbia

espresso parere favorevole;

- inviare, alla fine della ricerca, una RELAZIONE FINALE dello studio;
- ottemperare alle eventuali RACCOMANDAZIONI richieste dal Comitato Etico competente e a darne comunicazione per iscritto.

Lo studio dovrà essere eseguito secondo i principi etici fissati nella Dichiarazione di Helsinki; tutte le fasi dello stesso dovranno inoltre essere predisposte, attuate e descritte seguendo i principi della Buona Pratica Clinica.

Colgo l'occasione per porgere cordiali saluti.

Il Presidente
Dr. Paolo Arosio

Bibliography

Abboudi, Rochel Lieber et al. (1999). "A biomimetic controller for a multi-finger prosthesis". In: *IEEE Transactions on Rehabilitation Engineering* 7.2, pp. 121–129.

Acharya, Urja et al. (2018). "Inference of user qualities in shared control". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 588–595.

Acı, Çiğdem İnan, Murat Kaya, and Yuriy Mishchenko (2019). "Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods". In: *Expert Systems with Applications* 134, pp. 153–166.

Ahlstrom, Christer, Katja Kircher, and Albert Kircher (2013). "A gaze-based driver distraction warning system and its effect on visual behavior". In: *IEEE Transactions on Intelligent Transportation Systems* 14.2, pp. 965–973.

Aisen, Mindy Lipson et al. (1997). "The effect of robot-assisted therapy and rehabilitative training on motor recovery following stroke". In: *Archives of neurology* 54.4, pp. 443–446.

Alghamdi, M., Noura Alhakbani, and Abeer Al-Nafjan (2023). "Assessing the Potential of Robotics Technology for Enhancing Educational for Children with Autism Spectrum Disorder". In: *Behavioral Sciences* 13. DOI: 10.3390/bs13070598.

All Europe, Design for (2004). *The EIDD Stockholm Declaration 2004 - EIDD - DfA Europe — dfaeurope.eu*. https://dfaeurope.eu/what-is-dfa/dfa-documents/the-eidd-stockholm-declaration-2004/. [Accessed 17-12-2024].

American Psychiatric Association, APA (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV*. Vol. 4. American psychiatric association Washington, DC.

Arai, Tamio, Ryu Kato, and Marina Fujita (2010). "Assessment of operator stress induced by robot collaboration in assembly". In: *CIRP annals* 59.1, pp. 5–8.

Argyle, Michael, Mark Cook, and Duncan Cramer (1994). “Gaze and mutual gaze”. In: *The British Journal of Psychiatry* 165.6, pp. 848–850. URL: https://doi.org/10.1192/S0007125000012253.

Arora, Rhythm et al. (2024). “Socially Interactive Agents for Robotic Neurorehabilitation Training: Conceptualization and Proof-of-concept Study”. In: *Frontiers in Artificial Intelligence* 7, p. 1441955.

Arraras, JI et al. (2002). “Coping style, locus of control, psychological distress and pain-related behaviours in cancer and other diseases”. In: *Psychology, Health & Medicine* 7.2, pp. 181–187.

Artemiadis, P. and K. Kyriakopoulos (2010). “EMG-Based Control of a Robot Arm Using Low-Dimensional Embeddings”. In: *IEEE Transactions on Robotics* 26, pp. 393–398. DOI: 10.1109/TRO.2009.2039378.

Atman Uslu, Nilüfer, Gulay Öztüre Yavuz, and Yasemin Koçak Usluel (2023). “A systematic review study on educational robotics and robots”. In: *Interactive Learning Environments* 31.9, pp. 5874–5898.

Aygun, Ayca et al. (2022). “Investigating Methods for Cognitive Workload Estimation for Assistive Robots”. In: *Sensors (Basel, Switzerland)* 22. DOI: 10.3390/s22186834.

Baard, Paul P, Edward L Deci, and Richard M Ryan (2004). “Intrinsic need satisfaction: a motivational basis of performance and weil-being in two work settings 1”. In: *Journal of applied social psychology* 34.10, pp. 2045–2068.

Bambrah, Veerpal, Andrew B. Moynihan, and John D. Eastwood (Feb. 2023). “Self-focused but lacking self-knowledge: The relation between boredom and self-perception”. en. In: *Journal of Boredom Studies* 1.1. URL: https://www.boredomsociety.com/jbs/index.php/journal/article/view/15 (visited on 04/19/2023).

Bänninger-Huber, Eva and Felix Steiner (1992). “Identifying microsequences: A new methodological approach to the analysis of affective regulatory processes”. In: *“Two Butterflies on My Head...” Psychoanalysis in the Interdisciplinary Scientific Dialogue*. Springer, pp. 257–276.

Barchard, K. et al. (2020). “Measuring the Perceived Social Intelligence of Robots”. In: *ACM Transactions on Human-Robot Interaction (THRI)* 9, pp. 1 –29. DOI: 10.1145/3415139.

Bazarevsky, Valentin et al. (2019). “Blazeface: Sub-millisecond neural face detection on mobile gpus”. In: *arXiv preprint arXiv:1907.05047*.

Bemporad, Alberto (2020). “Global optimization via inverse distance weighting and radial basis functions”. In: *Computational Optimization and Applications*.

Bemporad, Alberto and Dario Piga (2019). “Active preference learning based on radial basis functions”. In: *CoRR* abs/1909.13049. arXiv: 1909.13049.

– (2021). “Global optimization based on active preference learning with radial basis functions”. In: *Machine Learning* 110, pp. 417–448.

Benchmark International, . (May 2024). *Robotics Industry Report*. URL: https://www.benchmarkintl.com/insights/2024-robotics-industry-report/.

Bennett, Hunter J et al. (2021). “Inter and intra-limb coordination variability during walking in adolescents with autism spectrum disorder”. In: *Clinical Biomechanics* 89, p. 105474.

Benrud-Larson, Lisa M and Stephen T Wegener (2000). “Chronic pain in neurorehabilitation populations: prevalence, severity and impact”. In: *NeuroRehabilitation* 14.3, pp. 127–137.

Berg, Andrew, Edward F Buffie, and Luis-Felipe Zanna (2018). “Should we fear the robot revolution?(The correct answer is yes)”. In: *Journal of Monetary Economics* 97, pp. 117–148.

Beyrodt, Sebastian et al. (2023). “Socially Interactive Agents as Cobot Avatars: Developing a Model to Support Flow Experiences and Well-Being in the Workplace”. In: *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, pp. 1–8.

Bickmore, Timothy, Amanda Gruber, and Rosalind Picard (2005). “Establishing the computer–patient working alliance in automated health behavior change interventions”. In: *Patient Education and Counseling* 59.1, pp. 21–30.

BioRICS (2021). *Mindstretch*. Retrieved November 21, 2024. BioRICS. URL: https://mindstretch.biorics.com/.

BIORICS (2024). *BioRICS NV BE*. https://www.mindbot.eu/consortium/biorics/. [Accessed 30-10-2024].

Blank, Amy A et al. (2014). “Current trends in robot-assisted upper-limb stroke rehabilitation: promoting patient engagement in therapy”. In: *Current physical medicine and rehabilitation reports* 2, pp. 184–195.

Blender (2024). *Blender*. URL: https://www.blender.org/.

Block, Alexis E et al. (2021). “The six hug commandments: Design and evaluation of a human-sized hugging robot with visual and haptic perception”. In: *Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction*, pp. 380–388.

Bogue, Robert (2012). “Robots for space exploration”. In: *Industrial Robot: An International Journal* 39.4, pp. 323–328.

Bogue, Robert (2017). “Domestic robots: Has their time finally come?” In: *Industrial Robot: An International Journal* 44.2, pp. 129–136.

Bonarini, Andrea et al. (2008). “Stress recognition in a robotic rehabilitation task”. In: *Robotic helpers: user interaction, interfaces and companions in assistive and therapy robotics, a workshop at ACM/IEEE HRI*, pp. 41–48.

Boucher, Jean-David et al. (2012). “I Reach Faster When I See You Look: Gaze Effects in Human–Human and Human–Robot Face-to-Face Cooperation”. In: *Frontiers in Neurorobotics* 6. DOI: 10.3389/fnbot.2012.00003.

Bradley, Margaret M and Peter J Lang (1994). “Measuring emotion: the self-assessment manikin and the semantic differential”. In: *Journal of behavior therapy and experimental psychiatry* 25.1, pp. 49–59.

Brambilla, Cristina et al. (2023a). “Azure Kinect performance evaluation for human motion and upper limb biomechanical analysis”. In: *Heliyon* 9.11.

Brambilla, Cristina et al. (2023b). “Biomechanical Assessments of the Upper Limb for Determining Fatigue, Strain and Effort from the Laboratory to the Industrial Working Place: A Systematic Review”. In: *Bioengineering* 10.4, p. 445.

Breazeal, Cynthia, Kerstin Dautenhahn, and Takayuki Kanda (2016). “Social robotics”. In: *Springer handbook of robotics*, pp. 1935–1972.

Broekens, J, M Heerink, H Rosendal, et al. (2009). “Assistive social robots in elderly care: A review. Gerontechnology, 8 (2), 94–103”. In: *Zugriff am* 11, p. 2012.

Bull, Peter and Gerry Connelly (1985). “Body movement and emphasis in speech”. In: *Journal of nonverbal behavior* 9, pp. 169–187. URL: https://doi.org/10.1007/BF01000738.

Cai, Laisi, Dongwei Liu, and Ye Ma (2021). “Placement recommendations for single kinect-based motion capture system in unilateral dynamic motion analysis”. In: *Healthcare*. Vol. 9. 8. MDPI, p. 1076.

Cancrini, Adriana et al. (2022). “The effects of robotic assistance on upper limb spatial muscle synergies in healthy people during planar upper-limb training”. In: *Plos one* 17.8, e0272813.

Canete, Angelika, Javier Gonzalez-Sanchez, and Rafael Guerra-Silva (2024). “Exploring Cognition and Affect during Human-Cobot Interaction”. In: *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. DOI: 10.1145/3610978.3641082.

Caporaso, T., S. Grazioso, and G. Gironimo (2022). "Development of an Integrated Virtual Reality System with Wearable Sensors for Ergonomic Evaluation of Human–Robot Cooperative Workplaces". In: *Sensors (Basel, Switzerland)* 22. DOI: 10.3390/s22062413.

Carbonero, Francesco, Ekkehard Ernst, and Enzo Weber (2020). "Robots worldwide: The impact of automation on employment and trade". In.

Carissoli, Claudia et al. (2024). "Mental workload and human-robot interaction in collaborative tasks: A scoping review". In: *International Journal of Human–Computer Interaction* 40.20, pp. 6458–6477.

Cary, Mark S (1978). "The role of gaze in the initiation of conversation". In: *Social Psychology*, pp. 269–271. URL: https://doi.org/10.2307/3033565.

Castelnuovo, Gianluca et al. (2016). "Psychological treatments and psychotherapies in the neurorehabilitation of pain: evidences and recommendations from the Italian Consensus Conference on Pain in Neurorehabilitation". In: *Frontiers in Psychology* 7, p. 115.

Ceseracciu, Elena, Zimi Sawacha, and Claudio Cobelli (2014). "Comparison of markerless and marker-based motion capture technologies through simultaneous data collection during gait: proof of concept". In: *PloS one* 9.3, e87640.

Chehayeb, Lara et al. (2021). "Individual Differences and the Function of Emotions in Socio-Emotional and Cognitive Conflict: If an Agent Shames you, will you still be Bored?" In: *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, pp. 1–8.

Cheng, Cecilia et al. (2013). "Cultural meaning of perceived control: a meta-analysis of locus of control and psychological symptoms across 18 cultural regions." In: *Psychological bulletin* 139.1, p. 152.

Cheng, Chao et al. (2023). "Recent advancements in agriculture robots: Benefits and challenges". In: *Machines* 11.1, p. 48.

Chitta, Sachin (2016). "MoveIt!: an introduction". In: *Robot Operating System (ROS) The Complete Reference (Volume 1)*, pp. 3–27.

Chitta, Sachin et al. (2017). "ros_control: A generic and simple control framework for ROS". In: *Journal of Open Source Software* 2.20, pp. 456–456.

Christensen, Julien M et al. (n.d.). "Man-Systems Integration Standards NASA-STD-3000, Volume II Revision B, July 1995". In: ().

Christiernin, Linn Gustavsson (2017). "How to describe interaction with a collaborative robot". In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 93–94.

Cicirelli, Grazia et al. (2015). "A kinect-based gesture recognition approach for a natural human robot interface". In: *International Journal of Advanced Robotic Systems* 12.3, p. 22.

Cieza, Alarcos et al. (2020). "Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019". In: *The Lancet* 396.10267, pp. 2006–2017.

CNR (2024). *Consiglio Nazionale delle Ricerche*. https://www.mindbot.eu/consortium/cnr/. [Accessed 30-10-2024].

Cobb-Clark, Deborah A, Sonja C Kassenboehmer, and Stefanie Schurer (2014). "Healthy habits: The connection between diet, exercise, and locus of control". In: *Journal of Economic Behavior & Organization* 98, pp. 1–28.

Cohen, Sheldon, Tom Kamarck, and Robin Mermelstein (1983). "A global measure of perceived stress". In: *Journal of health and social behavior*, pp. 385–396.

Commission, European (2017). *Eurobarometer — europa.eu*. https://europa.eu/eurobarometer/surveys/detail/2160. [Accessed 18-09-2024].

Corona-Acosta, Ileana P. and Eduardo Castillo-Castaneda (2015). "Dimensional Synthesis of a Planar Parallel Manipulator Applied to Upper Limb Rehabilitation". In: *Multibody Mechatronic Systems*. Ed. by Marco Ceccarelli and Eusebio Eduardo Hernández Martinez. Cham: Springer International Publishing, pp. 443–452. ISBN: 978-3-319-09858-6.

Cramer, Steven C. et al. (Apr. 2011). "Harnessing neuroplasticity for clinical applications". In: *Brain* 134.6, pp. 1591–1609. ISSN: 0006-8950. DOI: 10.1093/brain/awr039. URL: https://doi.org/10.1093/brain/awr039.

Crocker, T. et al. (2013). "Physical rehabilitation for older people in long-term care". In: *Cochrane Database of Systematic Reviews* 2013.2. cited By 111. DOI: 10.1002/14651858.CD004294.pub3.

Csikszentmihalhi, Mihaly (2020). *Finding flow: The psychology of engagement with everyday life*. Hachette UK.

Csikszentmihalyi, Mihaly (2000). *Beyond boredom and anxiety.* Jossey-bass.

Csikszentmihalyi, Mihaly, Reed Larson, and Suzanne Prescott (2014). "The ecology of adolescent activity and experience". In: *Applications of Flow in Human Development and Education*. Springer, pp. 241–254.

Csikszentmihalyi, Mihaly and Judith LeFevre (1989). "Optimal experience in work and leisure." In: *Journal of personality and social psychology* 56.5, p. 815.

Damm, Oliver et al. (2013). "Different gaze behavior in human-robot interaction in Asperger's syndrome: An eye-tracking study". In: *2013 IEEE RO-MAN*. IEEE, pp. 368–369.

Damon Kohler Rodrigo Queiro, Ernesto Corbellini (2019). *rosjava*. Retrieved November 21, 2024. ROS. URL: https://github.com/rosjava.

Dao, Le Anh et al. (2023). "Experience in Engineering Complex Systems: Active Preference Learning with Multiple Outcomes and Certainty Levels". In: *arXiv preprint arXiv:2302.14630*.

Das, Sanjib Kumar and Suman Mukhopadhyay (2014). "Integrating ergonomics tools in physical therapy for musculoskeletal risk assessment and rehabilitation–a review". In: *International Journal of Engineering & Scientific Research* 2.10, pp. 136–155.

d'Avella, Andrea et al. (2006). "Control of fast-reaching movements by muscle synergy combinations". In: *Journal of Neuroscience* 26.30, pp. 7791–7810.

Deal, Kathleen Holtz (2007). "Psychodynamic theory". In: *Advances in Social Work* 8.1, pp. 184–195.

Dei, Carla et al. (2024). "Design and testing of (A) MICO: a multimodal feedback system to facilitate the interaction between cobot and human operator". In: *Journal on Multimodal User Interfaces*, pp. 1–16.

Desideri, L. et al. (2019). "Emotional processes in human-robot interaction during brief cognitive testing". In: *Comput. Hum. Behav.* 90, pp. 331–342. DOI: 10.1016/J.CHB.2018.08.013.

DFKI (2024). *Deutsches Forschungszentrum fur Künstliche Intelligenz Gmbh*. https://www.mindbot.eu/consortium/dfki/. [Accessed 30-10-2024].

Dillon, Roberto, Ai Ni Teoh, Denise Dillon, et al. (2022). "Voice Analysis for Stress Detection and Application in Virtual Reality to Improve Public Speaking in Real-time: A Review". In: *arXiv preprint arXiv:2208.01041*.

Ding, B. et al. (2014). "Path Control of a Rehabilitation Robot Using Virtual Tunnel and Adaptive Impedance Controller". In: *2014 Seventh International Symposium on Computational Intelligence and Design*. Vol. 1, pp. 158–161. DOI: 10.1109/ISCID.2014.204.

DiPietro, Janice D. et al. (2019). "Computer- and Robot-Assisted Therapies to Aid Social and Intellectual Functioning of Children with Autism Spectrum Disorder". In: *Medicina* 55. DOI: 10.3390/medicina55080440.

Duchetto, Francesco Del, Paul E. Baxter, and Marc Hanheide (2020). "Are You Still With Me? Continuous Engagement Assessment From a Robot's Point of View". In: *Frontiers in Robotics and AI* 7. DOI: 10.3389/frobt.2020.00116.

Dumas, Raphaël, Rachid Aissaoui, and Jacques A de Guise (2004). "A 3D generic inverse dynamic method using wrench notation and quaternion algebra". In: *Computer methods in biomechanics and biomedical engineering* 7.3, pp. 159–166.

Dumas, Raphaël, Laurence Cheze, and J-P Verriest (2007). "Adjustments to McConville et al. and Young et al. body segment inertial parameters". In: *Journal of biomechanics* 40.3, pp. 543–553.

Duttweiler, Patricia C (1984). "The internal control index: A newly developed measure of locus of control". In: *Educational and psychological measurement* 44.2, pp. 209–221.

Ekman, Paul (1992). "An argument for basic emotions". In: *Cognition & Emotion* 6, pp. 169–200. URL: https://api.semanticscholar.org/CorpusID:11771973.

Ekman, Paul, Wallace V Freisen, and Sonia Ancoli (1980). "Facial signs of emotional experience." In: *Journal of personality and social psychology* 39.6, p. 1125.

Eldardeer, O, G Sandini, and F Rea (2020). "A Biological Inspired Cognitive Model of Multi-sensory Joint Attention in Human Robot Collaborative Tasks". In.

Ellwood, Charles A (1901). "The theory of imitation in social psychology". In: *American Journal of Sociology* 6.6, pp. 721–741.

Emery, Kim and Julie N Côté (2012). "Repetitive arm motion-induced fatigue affects shoulder but not endpoint position sense". In: *Experimental brain research* 216.4, pp. 553–564.

Falerni, Matteo Meregalli et al. (2024). "A framework for human–robot collaboration enhanced by preference learning and ergonomics". In: *Robotics and Computer-Integrated Manufacturing* 89, p. 102781.

Fanuc (2021). *Collaborative Robot CRX-10iA/L - Fanuc — fanuc.eu*. https://www.fanuc.eu/it/en/robots/robot-filter-page/collaborative-robots/crx-10ial. [Accessed 08-11-2024].

FFRI (2024). *Sveuciliste u Rijeci, Filozofski Fakultet u Rijeci*. https://www.mindbot.eu/consortium/ffri/. [Accessed 30-10-2024].

Fiore, S. et al. (2013). "Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior". In: *Frontiers in Psychology* 4. DOI: 10.3389/fpsyg.2013.00859.

Fiori, Marina, Grégoire Bollmann, and Jérôme Rossier (2015). "Exploring the path through which career adaptability increases job satisfaction and lowers job stress: The role of affect". In: *Journal of Vocational Behavior* 91, pp. 113–121.

Flanders, Martha, Stephen I Helms Tillery, and John F Soechting (1992). "Early stages in a sensorimotor transformation". In: *Behavioral and Brain Sciences* 15.2, pp. 309–320.

Forecast, Statista Market (2024). *Robotics - Worldwide | Statista Market Forecast — statista.com*. https://www.statista.com/outlook/tmo/robotics/worldwide#revenue. [Accessed 18-09-2024].

Fournier, Kimberly A et al. (2010). "Motor coordination in autism spectrum disorders: a synthesis and meta-analysis". In: *Journal of autism and developmental disorders* 40, pp. 1227–1240.

Frère, Julien and François Hug (2012). "Between-subject variability of muscle synergies during a complex motor skill". In: *Frontiers in computational neuroscience* 6, p. 99.

Fritsch, Frederick N and Ralph E Carlson (1980). "Monotone piecewise cubic interpolation". In: *SIAM Journal on Numerical Analysis* 17.2, pp. 238–246.

Fullagar, Clive, Antonella Delle Fave, and Steve Van Krevelen (2018). "Flow at work: The evolution of a construct". In: *Current Issues in Work and Organizational Psychology*. Routledge, pp. 278–299.

Gagné, Marylène and Edward L Deci (2005). "Self-determination theory and work motivation". In: *Journal of Organizational behavior* 26.4, pp. 331–362.

Gale, Catharine R, G David Batty, and Ian J Deary (2008). "Locus of control at age 10 years and health outcomes and behaviors at age 30 years: the 1970 British Cohort Study". In: *Psychosomatic Medicine* 70.4, pp. 397–403.

Gallardo, Jhair et al. (2024). "Human Emotion Estimation through Physiological Data with Neural Networks". In: *2024 19th Annual System of Systems Engineering Conference (SoSE)*, pp. 153–159. DOI: 10.1109/SOSE62659.2024.10620945.

García-Cossio, Eliana et al. (2014). "Cortex integrity relevance in muscle synergies in severe chronic stroke". In: *Frontiers in human neuroscience* 8, p. 744.

Gaschler, Andre et al. (2012). "Social behavior recognition using body posture and head pose for human-robot interaction". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2128–2133. DOI: 10.1109/IROS.2012.6385460.

Gassend, Blaise (2014). *Dynamyc Reconfigure Package*. Retrieved November 21, 2024. ROS. URL: https://wiki.ros.org/dynamic_reconfigure.

Gatzounis, Rena et al. (2017). “Taking a break in response to pain. An experimental investigation of the effects of interruptions by pain on subsequent activity resumption”. In: *Scandinavian Journal of Pain* 16.1, pp. 52–60.

Gebhard, Patrick, Gregor Mehlmann, and Michael Kipp (2012). “Visual SceneMaker - a tool for authoring interactive virtual characters”. In: *J. Multimodal User Interfaces* 6.1-2, pp. 3–11. DOI: 10.1007/s12193-011-0077-1. URL: https://doi.org/10.1007/s12193-011-0077-1.

Gebhard, Patrick et al. (2014). “Exploring interaction strategies for virtual characters to induce stress in simulated job interviews”. In: *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pp. 661–668.

Gebhard, Patrick et al. (2018). “MARSSI: Model of Appraisal, Regulation, and Social Signal Interpretation”. In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 497–506.

Gedam, Shruti and Sanchita Paul (2021). “A review on mental stress detection using wearable sensors and machine learning techniques”. In: *IEEE Access* 9, pp. 84045–84066.

Giannakakis, Giorgos et al. (2019). “Review on psychological stress detection using biosignals”. In: *IEEE Transactions on Affective Computing* 13.1, pp. 440–460.

Giberti, H., S. Cinquemani, and S. Ambrosetti (2013). “5R 2dof parallel kinematic manipulator – A multidisciplinary test case in mechatronics”. In: *Mechatronics* 23.8, pp. 949 –959. ISSN: 0957-4158. DOI: https://doi.org/10.1016/j.mechatronics.2012.09.006. URL: http://www.sciencedirect.com/science/article/pii/S0957415812001353.

Gillberg, Christopher (1998). “Asperger syndrome and high-functioning autism”. In: *The British journal of psychiatry* 172.3, pp. 200–209.

Giraud, Tom et al. (2021). “"Can you help me move this over there?": Training children with ASD to joint action through tangible interaction and virtual agent”. In: *TEI ’21: Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction, Online Event / Salzburg, Austria, February 14-19, 2021*. Ed. by Raphael Wimmer et al. ACM, 27:1–27:12. DOI: 10.1145/3430524.3440646. URL: https://doi.org/10.1145/3430524.3440646.

Google (2021). *Fitbit Inspire 3 — store.google.com*. https://store.google.com/it/product/fitbit_inspire_3?hl=en. [Accessed 08-11-2024].

Goris, Judith et al. (2020). “The relation between preference for predictability and autistic traits”. In: *Autism Research* 13.7, pp. 1144–1154.

Govaerts, Renée et al. (Aug. 2021). "Prevalence and incidence of work-related musculoskeletal disorders in secondary industries of 21st century Europe: a systematic review and meta-analysis". In: *BMC Musculoskeletal Disorders* 22.1, p. 751. ISSN: 1471-2474. DOI: 10.1186/s12891-021-04615-9. (Visited on 01/06/2023).

Grand View Research, . (2021). *Rehabilitation Robots Market Size & Share Report, 2022-2030 — grandviewresearch.com*. https://www.grandviewresearch.com/industry-analysis/rehabilitation-robots-market-report. [Accessed 17-09-2024].

– (2022). *Collaborative Robots Market Share & Growth Report, 2030 — grandviewresearch.com*. https://www.grandviewresearch.com/industry-analysis/collaborative-robots-market. [Accessed 15-10-2024].

Gross, James J (2013). *Handbook of emotion regulation*. Guilford publications.

Gross, Stephanie and Brigitte Krenn (2024). "A communicative perspective on human–robot collaboration in industry: Mapping communicative modes on collaborative scenarios". In: *International Journal of Social Robotics* 16.6, pp. 1315–1332.

Haas, John K (2014). "A history of the unity game engine". In.

Hadar, Uri et al. (1983). "Kinematics of head movements accompanying speech during conversation". In: *Human Movement Science* 2.1-2, pp. 35–46. URL: https://doi.org/10.1016/0167-9457(83)90004-0.

Harel, David (1987). "Statecharts: A visual formalism for complex systems". In: *Science of computer programming* 8.3, pp. 231–274.

Hart, Sandra G and Lowell E Staveland (1988). "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, 1988". In: *Advances in Human Psychology: Human Mental Workload. Elsevier Science*.

Hassan, Teena et al. (2019). "Automatic detection of pain from facial expressions: a survey". In: *IEEE transactions on pattern analysis and machine intelligence* 43.6, pp. 1815–1831.

Hayward, Susan M, Keith R McVilly, and Mark A Stokes (2019). "Autism and employment: What works". In: *Research in autism spectrum disorders* 60, pp. 48–58.

Heimbuch, Sven and Daniel Bodemer (2017). "Effects of Implicit Guidance on Contribution Quality in a Wiki-Based Learning Environment". In: *International Conference of the Learning Sciences*. Singapore: International Society of the Learning Sciences. DOI: https://doi.org/10.31234/osf.io/hm2cn.

Heimerl, Alexander et al. (2023). "ForDigitStress: A multi-modal stress dataset employing a digital job interview scenario". In: *arXiv preprint arXiv:2303.07742*.

Heinz, Melinda et al. (2013). "Perceptions of technology among older adults". In: *Journal of gerontological nursing* 39.1, pp. 42–51.

Hektner, Joel M, Jennifer A Schmidt, and Mihaly Csikszentmihalyi (2007). *Experience sampling method: Measuring the quality of everyday life*. Sage.

Helton, William S (2004). "Validation of a short stress state questionnaire". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 48. 11. Sage Publications Sage CA: Los Angeles, CA, pp. 1238–1242.

Hendricks, Dawn (2010). "Employment and adults with autism spectrum disorders: Challenges and strategies for success". In: *Journal of vocational rehabilitation* 32.2, pp. 125–134.

Hermens, Hermie J et al. (1999). "European recommendations for surface electromyography". In: *Roessingh research and development* 8.2, pp. 13–54.

Hess, Ursula and Agneta Fischer (2013). "Emotional mimicry as social regulation". In: *Personality and social psychology review* 17.2, pp. 142–157.

Hogan, Neville (Mar. 1985). "Impedance Control: An Approach to Manipulation: Part I—Theory". In: *Journal of Dynamic Systems, Measurement, and Control* 107.1, pp. 1–7. ISSN: 0022-0434. DOI: 10.1115/1.3140702. eprint: https://asmedigitalcollection.asme.org/dynamicsystems/article-pdf/107/1/1/5492345/1_1.pdf. URL: https://doi.org/10.1115/1.3140702.

Hong, Joanna H et al. (2021). "The positive influence of sense of control on physical, behavioral, and psychosocial health in older adults: An outcome-wide approach". In: *Preventive Medicine* 149, p. 106612.

Hooff, Madelon LM van and Edwin AJ van Hooft (2014). "Boredom at work: Proximal and distal consequences of affective work-related boredom." In: *Journal of occupational health psychology* 19.3, p. 348. URL: https://psycnet.apa.org/doi/10.1037/a0036821.

Hormaza, Leire Amezua et al. (2019). "On-line training and monitoring of robot tasks through virtual reality". In: *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*. Vol. 1. IEEE, pp. 841–846.

Hoshina, Atsushi et al. (2020). "Preliminary Experiment for Mentally Supporting Rehabilitation Robot based on Emotion Estimation". In: *2020 International Conference on Image Processing and Robotics (ICIP)*. IEEE, pp. 1–5.

Hu, Yue et al. (2022). "Toward Active Physical Human–Robot Interaction: Quantifying the Human State During Interactions". In: *IEEE Transactions* *on Human-Machine Systems* 52, pp. 367–378. DOI: 10.1109/thms.2021.3138684.

Huang, Chien-Ming and Bilge Mutlu (2016). "Anticipatory robot control for efficient human-robot collaboration". In: *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, pp. 83–90.

Huang, Tsung-Ren et al. (2020). "Asynchronously Embedding Psychological Test Questions into Human–Robot Conversations for User Profiling". In: *International Journal of Social Robotics* 13, pp. 1359 –1368. DOI: 10.1007/s12369-020-00716-y.

Huijnen, C. et al. (2017). "How to Implement Robots in Interventions for Children with Autism? A Co-creation Study Involving People with Autism, Parents and Professionals". In: *Journal of Autism and Developmental Disorders* 47, pp. 3079 –3096. DOI: 10.1007/s10803-017-3235-9.

Hull, Clark Leonard (1943). "Principles of behavior: an introduction to behavior theory." In.

Ielegems, Elke (2014). "Universal design, a methodological approach". In.

Ilies, Remus, Nikolaos Dimotakis, and Irene E De Pater (2010). "Psychological and physiological reactions to high workloads: Implications for well-being". In: *Personnel Psychology* 63.2, pp. 407–436. URL: https://doi.org/10.1111/j.1744-6570.2010.01175.x.

IRIARTE, Ainhoa APRAIZ, Ganix LASA ERLE, and Maitane Mazmela Etxabe (2021). "EVALUATING USER EXPERIENCE WITH PHYSIOLOGICAL MONITORING: A SYSTEMATIC LITERATURE REVIEW". In: 8. DOI: 10.6036/NT10072.

Jang, Jinkyu et al. (2016). "Application of experiential locus of control to understand users' judgments toward useful experience". In: *Computers in Human Behavior* 54, pp. 326–340.

Jatesiktat, P. et al. (2024). "Anatomical-Marker-Driven 3D Markerless Human Motion Capture." In: *IEEE journal of biomedical and health informatics* PP. DOI: 10.1109/JBHI.2024.3424869.

John, Alka Rachel et al. (2024). "Prediction of cognitive conflict during unexpected robot behavior under different mental workload conditions in a physical human–robot collaboration". In: *Journal of Neural Engineering* 21. DOI: 10.1088/1741-2552/ad2494.

Joosen, P., V. Exadaktylos, and D. Berckmans (2015). "An investigation on mental stress-profiling of race car drivers during a race". In: *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–4. DOI: 10.1109/BSN.2015.7299413.

Just, Fabian et al. (2018). "Exoskeleton transparency: feed-forward compensation vs. disturbance observer". In: *at-Automatisierungstechnik* 66.12, pp. 1014–1026.

Kadir, Bzhwen A, Ole Broberg, and Carolina Souza da Conceição (2019). "Current research and future perspectives on human factors and ergonomics in Industry 4.0". In: *Computers & Industrial Engineering* 137, p. 106004.

Kato, Ryu, Marina Fujita, and Tamio Arai (2010). "Development of advanced cellular manufacturing system with human-robot collaboration". In: *19th international symposium in robot and human interactive communication*. IEEE, pp. 355–360.

Kawamichi, H., Y. Kikuchi, and S. Ueno (2005). "Magnetoencephalographic measurement during two types of mental rotations of three-dimensional objects". In: *IEEE Transactions on Magnetics* 41.10. cited By 6, pp. 4200–4202. DOI: 10.1109/TMAG.2005.854802.

Kee, Dohyung and Waldemar Karwowski (Feb. 2007). "A Comparison of Three Observational Techniques for Assessing Postural Loads in Industry". In: *International journal of occupational safety and ergonomics : JOSE* 13, pp. 3–14. DOI: 10.1080/10803548.2007.11076704.

Ken Conley, Dirk Thomas and Jacob Perron (2012). *rospy*. Retrieved November 21, 2024. ROS. URL: https://wiki.ros.org/rospy.

Kesavayuth, Dusanee, Joanna Poyago-Theotoky, Vasileios Zikos, et al. (2020). "Locus of control, health and healthcare utilization". In: *Economic Modelling* 86, pp. 227–238.

Kiesler, S. et al. (2008). "Anthropomorphic interactions with a robot and robot-like agent". In: *Social Cognition* 26.2. cited By 190, pp. 169–181. DOI: 10.1521/soco.2008.26.2.169.

Kim, Tae-Ho, J. Vanloo, and W. Kim (2021). "3D Origami Sensing Robots for Cooperative Healthcare Monitoring". In: *Advanced Materials Technologies* 6. DOI: 10.1002/admt.202000938.

Klein, J., N. Roach, and E. Burdet (2014). "3DOM: A 3 Degree of Freedom Manipulandum to Investigate Redundant Motor Control". In: *IEEE Transactions on Haptics* 7.2, pp. 229–239. DOI: 10.1109/TOH.2013.59.

Knight, Victoria F., John C. Wright, and Andrea DeFreese (2019). "Teaching Robotics Coding to a Student with ASD and Severe Problem Behavior". In: *Journal of Autism and Developmental Disorders* 49, pp. 2632 –2636. DOI: 10.1007/s10803-019-03888-3.

Koceska, Natasa, Saso Koceski, and Biserka Simonovska (2021). “Review of stress recognition techniques and modalities”. In: *Balkan Journal of Applied Mathematics and Informatics* 4.2, pp. 21–32.

Kothig, Austin et al. (2021). “Connecting Humans and Robots Using Physiological Signals – Closing-the-Loop in HRI”. In: *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pp. 735–742. DOI: 10.1109/RO-MAN50785.2021.9515383.

Krebs, Hermano I et al. (2004). “Rehabilitation robotics: pilot trial of a spatial extension for MIT-Manus”. In: *Journal of neuroengineering and rehabilitation* 1, pp. 1–15.

Krägeloh, Christian U. et al. (2019). “Questionnaires to Measure Acceptability of Social Robots: A Critical Review”. In: *Robotics* 8, p. 88. DOI: 10.3390/robotics8040088.

KUKA (2024). *KUKA Deutschland Gmbh*. https://www.mindbot.eu/consortium/kuka/. [Accessed 30-10-2024].

Kunz, M. et al. (2017). “Problems of video-based pain detection in patients with dementia: A road map to an interdisciplinary solution”. In: *BMC Geriatrics* 17.1. DOI: 10.1186/s12877-017-0427-2.

Kwakkel, Gert, Boudewijn J Kollen, and Hermano I Krebs (2008). “Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review”. In: *Neurorehabilitation and neural repair* 22.2, pp. 111–121.

Kyrarini, Maria et al. (2021). “A survey of robots in healthcare”. In: *Technologies* 9.1, p. 8.

Lackas, Jessica (Feb. 2021). *Explicit and implicit guidance to emotion regulation to support collaborative tasks: A model based on socio-cognitive conflict and flow parameters*. Master's Thesis, Universität des Saarlands.

Lagomarsino, Marta et al. (2022). “Pick the Right Co-Worker: Online Assessment of Cognitive Ergonomics in Human–Robot Collaborative Assembly”. In: *IEEE Transactions on Cognitive and Developmental Systems* 15, pp. 1928–1937. DOI: 10.1109/TCDS.2022.3182811.

Landi, Chiara Talignani et al. (2018). “Relieving operators' workload: Towards affective robotics in industrial scenarios”. In: *Mechatronics*. DOI: 10.1016/J.MECHATRONICS.2018.07.012.

Larradet, Fanny et al. (2020). “Toward emotion recognition from physiological signals in the wild: approaching the methodological issues in real-life data collection”. In: *Frontiers in psychology* 11, p. 1111.

Larson, Michael J et al. (2011). “Feedback and reward processing in high-functioning autism”. In: *Psychiatry Research* 187.1-2, pp. 198–203.

Lasi, Heiner et al. (2014). "Industry 4.0". In: *Business & information systems engineering* 6, pp. 239–242.

Lavit Nicora, Matteo et al. (2021). "A human-driven control architecture for promoting good mental health in collaborative robot scenarios". In: *2021 30th IEEE international conference on robot & human interactive communication (RO-MAN)*. IEEE, pp. 285–291.

Lavit Nicora, Matteo et al. (2023). "Towards social embodied cobots: The integration of an industrial cobot with a social virtual agent". In: *arXiv preprint arXiv:2301.06471*.

Lavit Nicora, Matteo et al. (2024). "Gaze detection as a social cue to initiate natural human-robot collaboration in an assembly task". In: *Frontiers in Robotics and AI* 11.

Lee, Daniel D and H Sebastian Seung (1999). "Learning the parts of objects by non-negative matrix factorization". In: *nature* 401.6755, pp. 788–791.

Lee, Wonhyong et al. (2020). "Design of Effective Robotic Gaze-Based Social Cueing for Users in Task-Oriented Situations: How to Overcome In-Attentional Blindness?" In: *Applied Sciences*. DOI: 10.3390/app10165413.

Lee, Y. et al. (2023). "Developing Social Robots with Empathetic Non-Verbal Cues Using Large Language Models". In: *ArXiv* abs/2308.16529. DOI: 10.48550/arXiv.2308.16529.

Lefter, Iulia, Gertjan J Burghouts, and Léon JM Rothkrantz (2015). "Recognizing stress using semantics and modulation of speech and gestures". In: *IEEE Transactions on Affective Computing* 7.2, pp. 162–175.

Lencioni, T et al. (2021). "A randomized controlled trial on the effects induced by robot-assisted and usual-care rehabilitation on upper limb muscle synergies in post-stroke subjects". In: *Scientific reports* 11.1, p. 5323.

Leng, Jiewu et al. (2022). "Industry 5.0: Prospect and retrospect". In: *Journal of Manufacturing Systems* 65, pp. 279–295.

Leso, Veruscka, Luca Fontana, and Ivo Iavicoli (2018). "The occupational health and safety dimension of Industry 4.0". In: *La Medicina del lavoro* 109.5, p. 327.

Lewis, Michael (2008). "Self-conscious emotions: Embarrassment, pride, shame, and guilt." In.

Li, Ming ming et al. (2024). "Interactive effects of users' openness and robot reliability on trust: evidence from psychological intentions, task performance, visual behaviours, and cerebral activations." In: *Ergonomics*, pp. 1–21. DOI: 10.1080/00140139.2024.2343954.

Liang, C. et al. (2019). "A vision-based marker-less pose estimation system for articulated construction robots". In: *Automation in Construction*. DOI: 10.1016/J.AUTCON.2019.04.004.

Liu, Changchun, Pramila Rani, and N. Sarkar (2006). "Human-Robot Interaction Using Affective Cues". In: *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 285–290. DOI: 10.1109/ROMAN.2006.314431.

Liu, Changchun et al. (2007). "Affect recognition in robot assisted rehabilitation of children with autism spectrum disorder". In: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, pp. 1755–1760.

Liu, Li et al. (2024). "Application, development and future opportunities of collaborative robots (cobots) in manufacturing: A literature review". In: *International Journal of Human–Computer Interaction* 40.4, pp. 915–932.

Logitech (2021). *Webcam Logitech C920 PRO HD, video 1080p con audio stereo — logitech.com*. https://www.logitech.com/it-it/products/webcams/c920-pro-hd-webcam.960-001055.html?srsltid=AfmBOopM52webExXXttYVfbddVywu630jtE [Accessed 08-11-2024].

Loksa, Dastyni et al. (2016). "Programming, problem solving, and self-awareness: Effects of explicit guidance". In: *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 1449–1461.

Lombardi, M. et al. (2022). "iCub Being Social: Exploiting Social Cues for Interactive Object Detection Learning". In: *ArXiv* abs/2207.13552. DOI: 10.48550/arXiv.2207.13552.

Loomes, Rachel, Laura Hull, and William Polmear Locke Mandy (2017). "What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis". In: *Journal of the American Academy of Child & Adolescent Psychiatry* 56.6, pp. 466–474.

Lopes, Alexandre, Roberto Souza, and Helio Pedrini (2022). "A survey on RGB-D datasets". In: *Computer Vision and Image Understanding* 222, p. 103489.

Lorenzini, Marta et al. (2023). "Ergonomic human-robot collaboration in industry: A review". In: *Frontiers in Robotics and AI* 9, p. 813907.

Lu, Sin-Ru et al. (2024). "Implementation of Engagement Detection for Human–Robot Interaction in Complex Environments". In: *Sensors (Basel, Switzerland)* 24. DOI: 10.3390/s24113311.

Lucey, P. et al. (2011). "Painful data: The UNBC-McMaster shoulder pain expression archive database". In: pp. 57–64. DOI: 10.1109/FG.2011.5771462.

Mace, Jonathan (2012). *rosbridge*. Retrieved November 21, 2024. ROS. URL: https://wiki.ros.org/rosbridge_suite.

Maciejasz, P. et al. (2014). “A survey on robotic devices for upper limb rehabilitation”. In: *J Neuroeng Rehabil* 11, p. 3.

Majumder, Sumit et al. (2017). “Smart homes for elderly healthcare—Recent advances and research challenges”. In: *Sensors* 17.11, p. 2496.

Manjunatha, H., S. Jujjavarapu, and E. Esfahani (2020). “Classification of Motor Control Difficulty using EMG in Physical Human-Robot Interaction”. In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2708–2713. DOI: 10.1109/SMC42975.2020.9283016.

Martini, Enrico et al. (2024). “A Robust Filter for Marker-less Multi-person Tracking in Human-Robot Interaction Scenarios”. In: *ArXiv* abs/2406.01832. DOI: 10.48550/arXiv.2406.01832.

Maselli, MV et al. (2023). “Continuous Action Recognition in Manufacturing Contexts by Deep Graph Convolutional Networks”. In: *Intelligent Systems Conference*. Springer, pp. 156–173.

Matheson, Eloise et al. (2019). “Human–robot collaboration in manufacturing applications: A review”. In: *Robotics* 8.4, p. 100.

Mavridis, N. (2015). “A review of verbal and non-verbal human-robot interactive communication”. In: *Robotics and Autonomous Systems* 63.P1. cited By 226, pp. 22–35. DOI: 10.1016/j.robot.2014.09.031.

Mayetin, Umut and Serdar Kucuk (2022). “Design and experimental evaluation of a low cost, portable, 3-dof wrist rehabilitation robot with high physical human–robot interaction”. In: *Journal of Intelligent & Robotic Systems* 106.3, p. 65.

Mazefsky, C. A. et al. (July 2013). “The role of emotion regulation in autism spectrum disorder”. In: *J Am Acad Child Adolesc Psychiatry* 52.7, pp. 679–688.

MEDEA (2024). *IRCCS – Associazione La Nostra Famiglia ‘Istituto Scientifico Eugenio Medea’*. https://www.mindbot.eu/consortium/medea/. [Accessed 30-10-2024].

Mehrabian, Albert and James A. Russell (1974). *An approach to environmental psychology / Albert Mehrabian and James A. Russell*. MIT Press. ISBN: 978-0-262-13090-5.

Memar, Amirhossein H. and E. Esfahani (2018). “Physiological Measures for Human Performance Analysis in Human-Robot Teamwork: Case of Tele-Exploration”. In: *IEEE Access* 6, pp. 3694–3705. DOI: 10.1109/ACCESS.2018.2790838.

Microsoft (2021a). *Azure Kinect DK documentation*. Retrieved November 21, 2024. Microsoft. URL: https://docs.microsoft.com/en-us/azure/kinect-dk/.

– (2021b). *GitHub - microsoft/Azure-Kinect-Sensor-SDK: A cross platform (Linux and Windows) user mode SDK to read data from your Azure Kinect device. — github.com*. https://github.com/microsoft/Azure-Kinect-Sensor-SDK. [Accessed 08-11-2024].

– (2021c). *Microsoft Surface Laptop Models and Lineup | Microsoft Surface — microsoft.com*. https://www.microsoft.com/surface/devices/surface-laptop-models. [Accessed 08-11-2024].

Miller, Linda et al. (2021). "More Than a Feeling—Interrelation of Trust Layers in Human-Robot Interaction and the Role of User Dispositions and State Anxiety". In: *Frontiers in Psychology* 12. DOI: 10.3389/fpsyg.2021.592711.

Minamoto, Masahiko et al. (2018). "Tele-Operation of Robot by Image Processing of Markers Attached to Operator's Head". In: *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 2414–2419. DOI: 10.1109/ICMA.2018.8484620.

Moher, David et al. (2009). "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement". In: *Annals of internal medicine* 151.4, pp. 264–269.

Molaei, Amir et al. (2022). "A portable low-cost 3D-printed wrist rehabilitation robot: Design and development". In: *Mechanism and Machine Theory* 171, p. 104719.

Mollahosseini, Ali, Behzad Hassani, and Mohammad H. Mahoor (2019). "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild". In: *IEEE Trans. Affective Computing* 10.1, pp. 18–31.

Mondellini, Marta et al. (2023). "Behavioral patterns in robotic collaborative assembly: comparing neurotypical and Autism Spectrum Disorder participants". In: *Frontiers in Psychology* 14, p. 1245857.

Mondellini, Marta et al. (2024). "Exploring the Dynamics between Cobot's Production Rhythm, Locus of Control and Emotional State in a Collaborative Assembly Scenario". In: *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*. IEEE, pp. 1–6.

Morone, Giovanni et al. (2023). *Robot-and Technology-Boosting Neuroplasticity-Dependent Motor-Cognitive Functional Recovery: Looking towards the Future of Neurorehabilitation*.

MRMSOSP (2024). *Republic of Croatia – Ministry of Labour, Pension System, Family and Social Policy*. https://www.mindbot.eu/consortium/mrms/. [Accessed 30-10-2024].

Müller, Sarah L et al. (2017). "Subjective stress in hybrid collaboration". In: *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*. Springer, pp. 597–606.

Naddeo, Alessandro and Nicola Cappetti (2014). "New trend line of research about comfort evaluation: proposal of a framework for weighing and evaluating contributes coming from cognitive, postural and physiologic comfort perceptions". In: (visited on 01/13/2023).

Nah, Fiona Fui-Hoon et al. (2014). "Flow in gaming: literature synthesis and framework development". In: *International Journal of Information Systems and Management* 1.1-2, pp. 83–124.

Nahavandi, Saeid (2019). "Industry 5.0—A human-centric solution". In: *Sustainability* 11.16, p. 4371.

Nam, Hyung Seok et al. (2019). "Vision-assisted interactive human-in-the-loop distal upper limb rehabilitation robot and its clinical usability test". In: *Applied Sciences* 9.15, p. 3106.

NASA (2021). *NASA Antropometry and Biomechanics*. Retrieved November 21, 2024. NASA. URL: https://msis.jsc.nasa.gov/sections/section03.htm.

Nathanson, Donald L (1994). *Shame and pride: Affect, sex, and the birth of the self*. New York City: WW Norton & Company.

Nexus (2007). *Vicon Upper Limb Model*. https://docs.vicon.com/display/Nexus214?preview=/83296552/83296566/Model_UpperLimb_ProductGuide_Rev1.0_2007Jul.pdf. [Accessed 29-11-2024].

Nicholas, David et al. (2019). "Perspectives of employers about hiring individuals with autism spectrum disorder: Evaluating a cohort of employers engaged in a job-readiness initiative". In: *Journal of vocational Rehabilitation* 50.3, pp. 353–364.

Nomura, T. et al. (2008). "Prediction of Human Behavior in Human–Robot Interaction Using Psychological Scales for Anxiety and Negative Attitudes Toward Robots". In: *IEEE Transactions on Robotics* 24, pp. 442–451. DOI: 10.1109/TRO.2007.914004.

Nomura, Tatsuya et al. (2006a). "Measurement of anxiety toward robots". In: *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, pp. 372–377.

– (2006b). “Negative Attitudes toward Robots Scale”. In: *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems*.

Norton, T. et al. (2018). “Automated real-time stress monitoring of police horses using wearable technology”. In: *Applied Animal Behaviour Science* 198, pp. 67–74.

Nunnari, Fabrizio and Alexis Heloir (2019). “Yet another low-level agent handler”. In: *Computer Animation and Virtual Worlds* 30.3-4. e1891 cav.1891, e1891. DOI: `https://doi.org/10.1002/cav.1891`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/cav.1891`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/cav.1891`.

Nunnari, Fabrizio et al. (2023). “Understanding and mapping pleasure, arousal and dominance social signals to robot-avatar behavior”. In: *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, pp. 1–8.

Oosterhof, Nikolaas N and Alexander Todorov (2008). “The functional basis of face evaluation”. In: *Proceedings of the National Academy of Sciences* 105.32, pp. 11087–11092.

Operto, Stefania (2019). “Evaluating public opinion towards robots: a mixed-method approach”. In: *Paladyn, Journal of Behavioral Robotics* 10.1, pp. 286–297.

Parenti, Lorenzo, Marwen Belkaid, and A. Wykowska (2023). “Differences in Social Expectations About Robot Signals and Human Signals”. In: *Cognitive science* 47 12, e13393. DOI: `10.1111/cogs.13393`.

Patrizi, Alfredo, E. Pennestrì, and P. Valentini (2016). “Comparison between low-cost marker-less and high-end marker-based motion capture systems for the computer-aided assessment of working ergonomics”. In: *Ergonomics* 59, pp. 155 –162. DOI: `10.1080/00140139.2015.1057238`.

Pearce, Jon (2005). “Engaging the learner: how can the flow experience support e-learning?” In: *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*. Association for the Advancement of Computing in Education (AACE), pp. 2288–2295.

Pennisi, Paola et al. (2016). “Autism and social robotics: A systematic review”. In: *Autism Research* 9. DOI: `10.1002/aur.1527`.

Philip, Nada Y et al. (2021). “Internet of Things for in-home health monitoring systems: Current advances, challenges and future directions”. In: *IEEE Journal on Selected Areas in Communications* 39.2, pp. 300–310.

Polar (2021). *Polar H10 | Polar Global — polar.com*. https://www.polar.com/en/sensors/h10-heart-rate-sensor?srsltid=AfmBOorUovHkqBE63yYbdo999LFrzydWpTqQAv52bSKTYeu. [Accessed 08-11-2024].

Prajod, Pooja and Elisabeth André (2022). "On the Generalizability of ECG-based Stress Detection Models". In: *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, pp. 549–554.

Prajod, Pooja, Tobias Huber, and Elisabeth André (2022). "Using Explainable AI to Identify Differences Between Clinical and Experimental Pain Detection Models Based on Facial Expressions". In: *International Conference on Multimedia Modeling*. Springer, pp. 311–322.

Prajod, Pooja, Bhargavi Mahesh, and Elisabeth André (2024). "Stressor Type Matters!–Exploring Factors Influencing Cross-Dataset Generalizability of Physiological Stress Detection". In: *arXiv preprint arXiv:2405.09563*.

Prajod, Pooja et al. (2021). "Do Deep Neural Networks Forget Facial Action Units?—Exploring the Effects of Transfer Learning in Health Related Facial Expression Recognition". In: *International Workshop on Health Intelligence*. Springer, pp. 217–233.

Prajod, Pooja et al. (2023). "Gaze-based attention recognition for human-robot collaboration". In: *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, pp. 140–147.

Prajod, Pooja et al. (2024). "Flow in human-robot collaboration—multimodal analysis and perceived challenge detection in industrial scenarios". In: *Frontiers in Robotics and AI* 11, p. 1393795.

Proaction International, . (2022). *Industry 5.0: Revolutionizing Work by Putting People First — blog.proactioninternational.com*. https://blog.proactioninternational.com/en/industry-5.0-the-next-industrial-revolution-is-people-centric. [Accessed 08-10-2024].

Provost, E. et al. (2007). "Investigating Implicit Cues for User State Estimation in Human-Robot Interaction Using Physiological Measurements". In: *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 1125–1130. DOI: 10.1109/ROMAN.2007.4415249.

Pruss, Ethel et al. (2023). "Restoring Engagement in Human-Robot Interaction: A Brain-Computer Interface for Adaptive Learning with Robots". In: *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3247–3252. DOI: 10.1109/SMC53992.2023.10394055.

Qassim, Hassan M and WZ Wan Hasan (2020). "A review on upper limb rehabilitation robots". In: *Applied Sciences* 10.19, p. 6976.

Quigley, Morgan et al. (2009). "ROS: an open-source Robot Operating System". In: *ICRA workshop on open source software*. Vol. 3. 3.2. Kobe, Japan, p. 5.

Ragusa, Francesco et al. (2021). "The meccano dataset: Understanding human-object interactions from egocentric videos in an industrial-like domain". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1569–1578.

Rathi, P. et al. (2017). "Real-Time, Wearable, Biomechanical Movement Capture of Both Humans and Robots with Metal-Free Electrodes". In: *ACS Omega* 2, pp. 4132 –4142. DOI: 10.1021/acsomega.7b00491.

Ray, Céline, Francesco Mondada, and Roland Siegwart (2008). "What do people expect from robots?" In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vol. 41. 1, pp. 22–26. DOI: 10.1109/IROS.2008.4650714.

Redaelli, Davide Felice, Fabio Alexander Storm, and Giulia Fioretta (Nov. 2021). *MindBot Planetary Gearbox*. DOI: 10.5281/zenodo.5675810. URL: https://doi.org/10.5281/zenodo.5675810.

Reinhardt, Jakob et al. (2017). "Dominance and movement cues of robot motion: A user study on trust and predictability". In: *2017 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, pp. 1493–1498.

Reinkensmeyer, David J et al. (2012). "Comparison of three-dimensional, assist-as-needed robotic arm/hand movement training provided with Pneu-WREX to conventional tabletop therapy after chronic stroke". In: *American Journal of Physical Medicine & Rehabilitation* 91.11, S232–S241.

Research, Directorate-General for and . Innovation (2021). *Industry 5.0 - Towards a sustainable, human-centric and resilient European industry — research-and-innovation.ec.europa.eu*. https://research-and-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/all-publications/industry-50-towards-sustainable-human-centric-and-resilient-european-industry_en. [Accessed 17-09-2024].

– (2022). *Industry 5.0, a transformative vision for Europe — research-and-innovation.ec.europa.eu*. https://research-and-innovation.ec.europa.eu/knowledge-publications-tools-and-data/publications/all-publications/industry-50-transformative-vision-europe_en. [Accessed 17-09-2024].

Rethink (2021). *Pickit3D detection camera*. https://www.pickit3d.com/en/. [Accessed 08-11-2024].

Rethink (2022). *ROS interface Pickit 3.0 documentation*. https://docs.pickit3d.com/en/3.0/robot-integrations/ros/index.html. [Accessed 08-11-2024].

Riener, Robert, Tobias Nef, and Gery Colombo (2005). "Robot-aided neurorehabilitation of the upper extremities". In: *Medical and biological engineering and computing* 43, pp. 2–10.

Rihet, Mathias, Aurélie Clodic, and Raphaëlle N. Roy (2024). "Robot Noise: Impact on Electrophysiological Measurements and Recommendations". In: *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. DOI: 10.1145/3610978.3640708.

Rivas, Jesús J et al. (2015). "Detecting affective states in virtual rehabilitation". In: *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE, pp. 287–292.

Robotiq (2021). *Adaptive Grippers | Robotiq — robotiq.com*. https://robotiq.com/products/adaptive-grippers. [Accessed 08-11-2024].

Rodgers, Helen et al. (2019). "Robot assisted training for the upper limb after stroke (RATULS): a multicentre randomised controlled trial". In: *The Lancet* 394.10192, pp. 51–62.

Romat, Hugo et al. (2016). "Natural human-robot interaction using social cues". In: *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 503–504. DOI: 10.1109/HRI.2016.7451827.

Romeo, Laura et al. (2021a). "Performance analysis of body tracking with the microsoft azure kinect". In: *2021 29th Mediterranean Conference on Control and Automation (MED)*. IEEE, pp. 572–577.

Romeo, Laura et al. (2022). "Microsoft azure kinect calibration for three-dimensional dense point clouds and reliable skeletons". In: *Sensors* 22.13, p. 4986.

Romeo, Laura et al. (2024). "A Dataset on Human-Cobot Collaboration for Action Recognition in Manufacturing Assembly". In: *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, pp. 866–871.

Romeo, M. et al. (2021b). "Predicting apparent personality from body language: benchmarking deep learning architectures for adaptive social human–robot interaction". In: *Advanced Robotics* 35, pp. 1167 –1179. DOI: 10.1080/01691864.2021.1974941.

Rotter, Julian B (1966). "Generalized expectancies for internal versus external control of reinforcement." In: *Psychological monographs: General and applied* 80.1, p. 1.

Russakovsky, Olga et al. (2015). "Imagenet large scale visual recognition challenge". In: *International journal of computer vision* 115.3, pp. 211–252.

Saadatian, Aboozar, Mansour Sahebozamani, and Mohammad Taghi Karimi (2023). "Contrast of Maximum Functional Torque in the Shoulder Joint in Overhead Athletes with and without Sub-acromion Impingement during Sitting Throw". In: *Journal of Rehabilitation Sciences & Research* 10.1, pp. 44–48.

Saleh, M., F. A. Hanapiah, and H. Hashim (2020). "Robot applications for autism: a comprehensive review". In: *Disability and Rehabilitation: Assistive Technology* 16, pp. 580 –602. DOI: 10.1080/17483107.2019.1685016.

Santos, Laura et al. (2021). "Design of a Robotic Coach for Motor, Social and Cognitive Skills Training Toward Applications With ASD Children". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29, pp. 1223–1232. DOI: 10.1109/TNSRE.2021.3091320.

Santos, Wilk Oliveira dos et al. (2018). "Flow theory to promote learning in educational systems: Is it really relevant?" In: *Revista Brasileira de Informática na Educação* 26.2.

Sarabia, Miguel et al. (2018). "Assistive robotic technology to combat social isolation in acute hospital settings". In: *International Journal of Social Robotics* 10, pp. 607–620. URL: https://doi.org/10.1007/s12369-017-0421-z.

Saran, Akanksha et al. (2018). "Human gaze following for human-robot interaction". In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 8615–8621. URL: https://doi.org/10.1109/IROS.2018.8593580.

Sarri, Kyriaki and Christine K. Syriopoulou-Delli (2021). "Robotics for enhancing independent living skills in adolescents and young adults with autism spectrum disorder: a systematic review". In.

Sbargoud, Fazia et al. (2021). "HYBRID CLASSIFICATION STRATEGY OF EMG SIGNALS FOR ROBOTIC HAND CONTROL". In: *Biomedical Engineering: Applications, Basis and Communications*, p. 2150015. DOI: 10.4015/S1016237221500150.

Scano, Alessandro, Franco Molteni, and Lorenzo Molinari Tosatti (2019). "Low-cost tracking systems allow fine biomechanical evaluation of upper-limb daily-life gestures in healthy people and post-stroke patients". In: *Sensors* 19.5, p. 1224.

Scano, Alessandro et al. (2018). "Robotic assistance for upper limbs may induce slight changes in motor modules compared with free movements in stroke survivors: a cluster-based muscle synergy analysis". In: *Frontiers in human neuroscience* 12, p. 290.

Scano, Alessandro et al. (2020). "Analysis of upper-limb and trunk kinematic variability: Accuracy and reliability of an RGB-D sensor". In: *Multimodal Technologies and Interaction* 4.2, p. 14.

SceneMaker, Visual (2012). *Visual SceneMaker*. URL: http://scenemaker.dfki.de/.

Schmidt, P. et al. (2018). "Introducing WeSAD, a multimodal dataset for wearable stress and affect detection". In: pp. 400–408. DOI: 10.1145/3242969.3242985.

Sciutti, Alessandra et al. (2018). "Humanizing human-robot interaction: On the importance of mutual understanding". In: *IEEE Technology and Society Magazine* 37.1, pp. 22–29.

Sellner, J., P. Thiam, and F. Schwenker (2019). "Visualizing facial expression features of pain and emotion data". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11377 LNAI, pp. 101–115. DOI: 10.1007/978-3-030-20984-1_9.

Selye, Hans (1956). *The Stress of Life*. McGraw Hill, pp. 323–325.

Sener, Fadime et al. (2022). "Assembly101: A large-scale multi-view video dataset for understanding procedural activities". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21096–21106.

Seraji, H. (1994). "Adaptive admittance control: an approach to explicit force control in compliant motion". In: *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, 2705–2712 vol.4. DOI: 10.1109/ROBOT.1994.350927.

Shafti, A. et al. (2018). "Real-time Robot-assisted Ergonomics*". In: *2019 International Conference on Robotics and Automation (ICRA)*, pp. 1975–1981. DOI: 10.1109/ICRA.2019.8793739.

Shao, Shiliang et al. (2021). "Comparison Analysis of Different Time-Scale Heart Rate Variability Signals for Mental Workload Assessment in Human-Robot Interaction". In: *Wireless Communications and Mobile Computing*. DOI: 10.1155/2021/8371637.

Simonyan, Karen and Andrew Zisserman (2014). "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556*.

Singh, Gaganpreet et al. (2019). "Physiologically Attentive User Interface for Robot Teleoperation - Real Time Emotional State Estimation and Interface

Modification using Physiology, Facial Expressions and Eye Movements". In: pp. 294–302. DOI: 10.5220/0006733002940302.

Smets, Elena, Walter De Raedt, and Chris Van Hoof (2018). "Into the wild: the challenges of physiological stress detection in laboratory and ambulatory settings". In: *IEEE journal of biomedical and health informatics* 23.2, pp. 463–473.

Smets, Elena et al. (2013). "Monitoring the Mental Status of Football Players." In: *icSPORTS*, pp. 206–213.

Smith, Paul, Mubarak Shah, and Niels da Vitoria Lobo (2003). "Determining driver visual attention with one camera". In: *IEEE transactions on intelligent transportation systems* 4.4, pp. 205–218.

Sorensen, Glorian et al. (2019). "Improving working conditions to promote worker safety, health, and wellbeing for low-wage workers: The workplace organizational health study". In: *International journal of environmental research and public health* 16.8, p. 1449.

Souza, Rhaíra Helena Caetano e and Eduardo Lázaro Martins Naves (2021). "Attention detection in virtual environments using EEG signals: a scoping review". In: *frontiers in physiology* 12, p. 727840.

Staffa, M. and Silvia Rossi (2022). "Enhancing Affective Robotics via Human Internal State Monitoring". In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 884–890. DOI: 10.1109/RO-MAN53752.2022.9900762.

Stamatelopoulou, Foteini et al. (2018). ""Being in the zone": a systematic review on the relationship of psychological correlates and the occurrence of flow experiences in sports' performance". In: *Psychology* 9.08, p. 2011.

Stein, Joel (2012). *Robotics in rehabilitation: technology as destiny*.

Sterling-Turner, Heather E and Sara S Jordan (2007). "Interventions addressing transition difficulties for individuals with autism". In: *Psychology in the Schools* 44.7, pp. 681–690.

Su, Hang et al. (2023). "Recent advancements in multimodal human–robot interaction". In: *Frontiers in Neurorobotics* 17, p. 1084000.

Swangnetr, M. (2010). "Analysis of Patient-Robot Interaction Using Statistical and Signal Processing Methods". In.

Taelman, Joachim et al. (2016). "Stress Level Monitoring in Car Racing - Examples of Measurements during Races". In: *Proceedings of the 4th International Congress on Sport Sciences Research and Technology Support - Volume 1: icSPORTS*, INSTICC. SciTePress, pp. 59–62. ISBN: 978-989-758-205-9. DOI: 10.5220/0006084500590062.

Takayama, Leila et al. (2011). “Assisted driving of a mobile remote presence system: System design and controlled user evaluation”. In: *2011 IEEE international conference on robotics and automation*. IEEE, pp. 1883–1889.

Tayibnapis, Iman Rahmansyah, Min-Kook Choi, and Soon Kwon (2018). “Driver’s gaze zone estimation by transfer learning”. In: *2018 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, pp. 1–5.

Teasell, Robert et al. (2005). “The Role of Timing and Intensity of Rehabilitation Therapies”. In: *Topics in Stroke Rehabilitation* 12.3, pp. 46–57.

Thomas, Patricia A and Seoyoun Kim (2021). “Lost touch? Implications of physical touch for physical health”. In: *The Journals of Gerontology: Series B* 76.3, e111–e115.

Thompson, Edmund R (2007). “Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS)”. In: *Journal of cross-cultural psychology* 38.2, pp. 227–242.

Toisoul, Antoine et al. (Jan. 2021). “Estimation of continuous valence and arousal levels from faces in naturalistic conditions”. en. In: *Nature Machine Intelligence* 3.1, pp. 42–50. ISSN: 2522-5839. DOI: 10.1038/s42256-020-00280-0. URL: https://www.nature.com/articles/s42256-020-00280-0 (visited on 04/27/2023).

Tropea, Peppino et al. (2013). “Effects of early and intensive neuro-rehabilitative treatment on muscle synergies in acute post-stroke patients: a pilot study”. In: *Journal of neuroengineering and rehabilitation* 10, pp. 1–15.

Tseng, Kevin C et al. (2024). “Portable robots for upper-limb rehabilitation after stroke: a systematic review and meta-analysis”. In: *Annals of Medicine* 56.1, p. 2337735.

Turner-Stokes, L. et al. (2015). “Patient engagement and satisfaction with goal planning: Impact on outcome from rehabilitation”. In: *International Journal of Therapy and Rehabilitation* 22.5. cited By 35, pp. 210–216. DOI: 10.12968/ijtr.2015.22.5.210.

Tutam, Mahmut (2022). “Warehousing 4.0 in Logistics 4.0”. In: *Logistics 4.0 and Future of Supply Chains*, pp. 95–118.

UAU (2024). *Universitaet Augsburg*. https://www.mindbot.eu/consortium/uau/. [Accessed 30-10-2024].

UMIL (2024). *Università degli Studi di Milano*. https://www.mindbot.eu/consortium/umil/. [Accessed 30-10-2024].

Unger, Darlene D (2002). “Employers’ attitudes toward persons with disabilities in the workforce: myths or realities?” In: *Focus on autism and other developmental disabilities* 17.1, pp. 2–10.

Unity Technologies, . (2022). *GitHub - Unity-Technologies/Unity-Robotics-Hub: Central repository for tools, tutorials, resources, and documentation for robotics simulation in Unity. — github.com*. https://github.com/Unity-Technologies/Unity-Robotics-Hub. [Accessed 25-11-2024].

Upasani, Satyajit et al. (2023). "Eye-Tracking in Physical Human–Robot Interaction: Mental Workload and Performance Prediction". In: *Human Factors* 66, pp. 2104 –2119. DOI: 10.1177/00187208231204704.

Vagnetti, Roberto et al. (2023). "Instruments for Measuring Psychological Dimensions in Human-Robot Interaction: Systematic Review of Psychometric Properties". In: *Journal of Medical Internet Research* 26. DOI: 10.2196/55597.

Van Volkom, Michele, Janice C Stapley, and Johnna Malter (2013). "Use and perception of technology: Sex and generational differences in a community sample". In: *Educational Gerontology* 39.10, pp. 729–740.

Velana, Maria et al. (2017). "The senseemotion database: A multimodal database for the development and systematic validation of an automatic pain-and emotion-recognition system". In: *IAPR Workshop on Multimodal Pattern Recognition of Social Signals in Human-Computer Interaction*. Springer, pp. 127–139.

Vera-Ortega, Pablo et al. (2022). "Enabling Remote Responder Bio-Signal Monitoring in a Cooperative Human–Robot Architecture for Search and Rescue". In: *Sensors (Basel, Switzerland)* 23. DOI: 10.3390/s23010049.

Vigni, Francesco, Antonio Andriella, and Silvia Rossi (2024). "A rosbag tool to improve dataset reliability". In: *Companion of the 2024 ACM/IEEE international conference on human-robot interaction*, pp. 1085–1089.

Villani, Valeria et al. (2019). "Humans interacting with multi-robot systems: a natural affect-based approach". In: *Autonomous Robots* 44, pp. 601 –616. DOI: 10.1007/s10514-019-09889-6.

Volpe, Bruce T et al. (2000). "A novel approach to stroke rehabilitation: robot-aided sensorimotor stimulation". In: *Neurology* 54.10, pp. 1938–1944.

Wagner, Johannes et al. (2013). "The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time". In: *Proceedings of the 21st ACM international conference on Multimedia*, pp. 831–834.

Walter, Steffen et al. (2013). "The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system". In: *2013 IEEE international conference on cybernetics (CYBCO)*. IEEE, pp. 128–131.

Wang, F. et al. (2018). "Regularizing face verification nets for pain intensity regression". In: vol. 2017-September, pp. 1087–1091. DOI: 10.1109/ICIP.2017.8296449.

Wang, Ju et al. (2021). "Unobtrusive health monitoring in private spaces: The smart home". In: *Sensors* 21.3, p. 864.

Washabaugh, Edward P et al. (2018). "A portable passive rehabilitation robot for upper-extremity functional resistance training". In: *IEEE Transactions on Biomedical Engineering* 66.2, pp. 496–508.

Watson, David, Lee Anna Clark, and Auke Tellegen (1988). "Development and validation of brief measures of positive and negative affect: the PANAS scales." In: *Journal of personality and social psychology* 54.6, p. 1063.

Weiss, Astrid, Ann-Kathrin Wortmeier, and Bettina Kubicek (2021). "Cobots in industry 4.0: A roadmap for future practice studies on human–robot collaboration". In: *IEEE Transactions on Human-Machine Systems* 51.4, pp. 335–345.

Weiss, Mary Jane and Sandra L Harris (2001). "Teaching social skills to people with autism". In: *Behavior modification* 25.5, pp. 785–802.

Wick, F.A. et al. (2019). "Perception in dynamic scenes: What is your Heider capacity?" In: *Journal of Experimental Psychology: General* 148.2, pp. 252–271. DOI: 10.1037/xge0000557.

Windolf, M., N. Götzen, and M. Morlock (2008). "Systematic accuracy and precision analysis of video motion capturing systems–exemplified on the Vicon-460 system." In: *Journal of biomechanics* 41 12, pp. 2776–80. DOI: 10.1016/j.jbiomech.2008.06.024.

Wixted, Fiona and Leonard O'Sullivan (2014). "The effect of automated manufacturing environments on employee health". In: *Irish Ergonomics Society*, p. 80.

World Helath Organization, . (2024). *Rehabilitation — who.int*. https://www.who.int/news-room/fact-sheets/detail/rehabilitation. [Accessed 17-09-2024].

Wynford-Thomas, R. and N.P. Robertson (2017). "The economic burden of chronic neurological disease". In: *Journal of Neurology* 264.11. cited By 8, pp. 2345–2347. DOI: 10.1007/s00415-017-8632-7.

Xiang, Xiang et al. (2022). "Imbalanced regression for intensity series of pain expression from videos by regularizing spatio-temporal face nets". In: *Pattern Recognition Letters* 163, pp. 152–158.

Xu, Xun et al. (2021). "Industry 4.0 and Industry 5.0—Inception, conception and perception". In: *Journal of manufacturing systems* 61, pp. 530–535.

Yamine, Jawad et al. (2020). "A planar parallel device for neurorehabilitation". In: *Robotics* 9.4, p. 104.

Yan, Yuchen and Yunyi Jia (Jan. 2022). "A Review on Human Comfort Factors, Measurements, and Improvements in Human–Robot Collaboration". en. In: *Sensors* 22.19, p. 7431. ISSN: 1424-8220. DOI: 10.3390/s22197431. (Visited on 03/17/2023).

Yang, Jin et al. (2016). "Adaptive control with a fuzzy tuner for cable-based rehabilitation robot". In: *International Journal of Control, Automation and Systems* 14.3, pp. 865–875.

Yazdanirad, Saeid et al. (2018). "Comparing the Effectiveness of Three Ergonomic Risk Assessment Methods-RULA, LUBA, and NERPA-to Predict the Upper Extremity Musculoskeletal Disorders". eng. In: *Indian journal of occupational and environmental medicine* 22.1. 29743780[pmid], pp. 17–21. ISSN: 0973-2284. DOI: 10.4103/ijoem.IJOEM_23_18.

Yi, Fangqiu, Hongyu Wen, and Tingting Jiang (2021). "Asformer: Transformer for action segmentation". In: *arXiv preprint arXiv:2110.08568*.

Yoon, David J et al. (2022). "The balance between positive and negative affect in employee well-being". In: *Journal of Organizational Behavior* 43.4, pp. 763–782.

Yu, Yantao et al. (2019). "An automatic and non-invasive physical fatigue assessment method for construction workers". In: *Automation in construction* 103, pp. 1–12.

Yun, Sangseok et al. (2017). "Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system". In: *Autism Research* 10. DOI: 10.1002/aur.1778.

Zaletelj, Janez and Andrej Košir (2017). "Predicting students' attention in the classroom from Kinect facial and body features". In: *EURASIP journal on image and video processing* 2017.1, pp. 1–12.

Zhang, Xucong et al. (2020). "Eth-xgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation". In: *European Conference on Computer Vision*. Springer, pp. 365–381.

Zhang, Xue, Zan Yue, and Jing Wang (2017). "Robotics in Lower-Limb Rehabilitation after Stroke". In: *Behavioural Neurology* 2017, pp. 1–13. ISSN: 0953-4180, 1875-8584. DOI: 10.1155/2017/3731802. URL: https://www.hindawi.com/journals/bn/2017/3731802/ (visited on 11/18/2020).