1 2 9 0

UNIVERSIDADE D
COIMBRA

Eduardo Simões Araújo

Evaluating Human-Robot Collaboration through Virtual Reality Scenarios

Dissertation in the context of the Master in Multimedia Design, advised by Professor Doctor Paula Alexandra Silva and co-advised by Professor Doctor Artur Pilacinski and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

January of 2023

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE Đ
COIMBRA

DEPARTMENT OF INFORMATICS ENGINEERING

Eduardo Simões Araújo

Evaluating Human-Robot Collaboration through Virtual Reality Scenarios

Dissertation in the context of the Master in Multimedia Design, advised by Professor Doctor Paula Alexandra Silva and co-advised by Professor Doctor Artur Pilacinski and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

January of 2023

1 2 9 0

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE Ð
COIMBRA

DEPARTAMENTO DE ENGENHARIA INFORMÁTICA

Eduardo Simões Araújo

Avaliação de Colaboração Humano-Robô através de Cenários de Realidade Virtual

Dissertação no âmbito do Mestrado em Design e Multimédia, orientada pela Professora Doutora Paula Alexandra Silva e co-orientada pelo Professor Doutor Artur Pilacinski apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Janeiro de 2023

Acknowledgements

This work was only possible through several people's backing, help and support.

First, I would like to thank my advisor, Professor Paula Alexandra Silva, for the consistency of her support, the weekly meetings and the advice that moulded the work. My co-advisor, Professor Artur Pilacinski, for his guidance while working at the PROACTION Laboratory and for providing insightful feedback throughout the work.

PhD student Diogo Branco and his advisor Professor Sergi Bermúdez I Badia, from the University of Madeira, for supplying their Unity project and for their cooperation in the meetings and the initial design and development phase. PhD student Soraia Oliveira for her help in recruiting test participants and conducting the test sessions.

Although not by name, I would also like to thank everyone from DEI and PROACTION Laboratory who helped me with their suggestions or overcoming difficulties.

Lastly, I would like to thank my family for helping me to accomplish my goals and for being present.

Abstract

Human-Robot Collaboration (HRC) is a highly inter- and multidisciplinary, still emerging, field of study. As technology develops and robot presence increases in society, it is expected that the coexistence, interaction, and collaboration between humans and robots will also increase. It is therefore important to investigate what factors affect the collaboration between humans and robots. This research explores the effect of robot characteristics, such as morphology and behaviour, on human performance, while executing a collaborative assembly task in Virtual Reality (VR). Three collaborative robots with different levels of anthropomorphism are used as participating agents in the collaboration: Sawyer and Baxter, from Rethink Robotics, and LBR iiwa, from Kuka. Both sequential and simultaneous collaboration are studied as distinct collaborative approaches. An interactive VR simulation was conceptualised, designed and developed to reproduce the collaborative assembly task's conditions. Then, an experimental evaluation was carried out with 36 participants. The participants' eyes and pulse data were collected using the VR system embedded eye-tracking and a pulse sensor. The participants' task completion times were also measured. Psychological data was gathered through the Godspeed Questionnaire Series (GQS) and the VR Presence Questionnaire (PQ), with the former reporting on how the participants perceived the robots and the latter the participants' sense of presence in the VR simulation. The analysis of the results indicates that human performance is affected by the type of robot and the collaborative approach used, where collaboration is most efficient with Sawyer using a simultaneous approach. These results contribute to understanding what factors influence human performance in HRC. Furthermore, the designed and developed interactive VR simulation is shared, allowing future HRC researchers to build upon the project and use it in future studies.

Keywords

Human-Robot Collaboration, Virtual Reality

Resumo

Colaboração humano-robô (CHR) é um campo de estudo emergente, altamente inter- e multidisciplinar. À medida que a tecnologia evoluí e a presença de robôs aumenta entre a sociedade, também é esperado que a coexistência, interação e colaboração entre humanos e robôs aumente. Portanto, é importante investigar os fatores que afetam a colaboração entre humanos e robôs. Esta pesquisa explora o efeito das características do robô, como morfologia e comportamento, no desempenho humano, durante a execução de uma tarefa de montagem colaborativa em Realidade Virtual (RV). Três robôs colaborativos com diferentes níveis de antropomorfismo são usados como agentes participantes na colaboração: Sawyer e Baxter, da Rethink Robotics, e LBR iiwa, da Kuka. Tanto a colaboração sequencial como a simultânea são estudadas como abordagens colaborativas distintas. Foi realizada a conceptualização, design e desenvolvimento de uma simulação RV interativa para reproduzir as condições da tarefa de montagem colaborativa. Em seguida, foi realizada uma avaliação experimental com 36 participantes. Os dados dos olhos e do pulso dos participantes foram adquiridos usando o sensor de monitorização ocular integrado do sistema VR e um sensor de pulso. Os tempos de conclusão das tarefas também foram medidos. Os dados psicológicos foram recolhidos através da Série de Questionários Godspeed e do Questionário de Presença de RV, com o primeiro indicando a percepção dos participantes em relação aos robôs e o segundo o indicando a sensação de presença dos participantes na simulação de RV. A análise dos resultados indica que o desempenho humano é afetado pelo tipo de robô e pela abordagem colaborativa utilizada, onde a colaboração é mais eficiente com o robô Sawyer usando uma abordagem simultânea. Estes resultados contribuem para entender os fatores que influenciam o desempenho humano em CHR. Além disso, a simulação de RV interativa projetada e desenvolvida é partilhada, permitindo que futuros investigadores de CHR desenvolvam o projeto e o usem em estudos futuros.

Palavras-Chave

Colaboração Humano-Robô, Realidade Virtual

Contents

Acronyms

AI Artificial Intelligence.

AMR autonomous mobile robot.

ATM Automated Teller Machine.

BPM Beats per Minute.

CSV Comma-separated values.

DOF degrees of freedom.

FOV Field of View.

GQS Godspeed Questionnaire Series.

GUI Graphical User Interface.

HCI Human-Computer Interaction.

HG Hand Guiding.

HMD Head Mounted Display.

HRC Human-Robot Collaboration.

HRI Human-Robot Interaction.

IBI Inter Beat Interval.

IEEE Institute of Electrical and Electronics Engineers.

IFR International Federation of Robotics.

IPD interpupillary distance.

ISO International Standards Organization.

LSL Lab Streaming Layer.

LTS Long-term support.

PFL Power and Force Limiting.

PPG Photoplethysmography.

PQ Presence Questionnaire.

ROS Robot Operating System.

RPA Robotic Process Automation.

SDK Software Development Kit.

SMS Safety-Rated Monitored Stop.

SSM Speed and Separation Monitoring.

URP Universal Render Pipeline.

UX User Experience.

VR Virtual Reality.

List of Figures

List of Tables

Chapter 1

Introduction

This chapter is divided into four sections. The first section explains the motivation of this work, the second section contextualises the work, and the third section presents the objectives of this dissertation and its contributions. Finally, the last section outlines the document structure.

1.1 Context and Motivation

We live in an era where technology dominates almost every aspect of our daily lives. In the last two decades, the number of robots deployed in the most diverse applications has been increasing, from space exploration to elderly care (Goodrich and Schultz, 2007). The increased robotic presence in society will eventually lead us to coexist, interact and collaborate ever more frequently with these machines, but in different ways than how we interact with computers or smartphones, due to robots' embodied nature.

Since a long time ago, humans have been fascinated and inspired by nature, attempting to understand and simulate it in various forms. Robots are no exception, these are machines built to simulate and amplify human and animal capabilities, and there have always been attempts at anthropomorphising robots, both through their morphology and behaviour (Dautenhahn, 2007a; Mori et al., 2012). Robots are continuously evolving along several dimensions and are increasingly capable of highly complex behaviours. As the presence of robots increases in society, so does the need for novel and adequate forms of interaction and collaboration with these machines. In Human-Robot Collaboration (HRC) applications, humans are in close proximity to robots, performing collaborative activities that benefit from both human and robot qualities, thus improving productivity and ergonomics (Dianatfar et al., 2021).

Although safety measures are implemented in robots designed to interact with humans, there is a danger component that can not be ignored due to the intrinsic physicality involved in the interactions. The rapid evolution of Virtual Reality (VR) technology, its potential and applications have also expanded (LaValle, 2020), including its use in the field of HRC (Dianatfar et al., 2021). The use of VR technology in HRC offers a more time and cost-efficient alternative to robotics research and development, allowing the study and evaluation of prototype concepts without compromising user safety (Badia et al., 2022).

In this dissertation, VR will be used to simulate a human-robot collaborative assembly task with robots of different morphologies, using two collaborative approaches, sequential and simultaneous collaboration. This simulation is then used to study aspects of human performance while interacting with robots. The use of VR benefits this study by reducing costs, the development and evaluation time, and decreasing human safety risks, as opposed to using real robots. Additionally, human performance is evaluated in an HRC context by analysing humans' responses to the stimuli presented. As human performance is directly tied to productivity and ergonomics, the findings of this study may improve future HRC research and applications.

This dissertation was developed in the context of the project Neurocobots: Enhancing Human-Robot Collaboration[1], which seeks to develop a brain-machine interface to improve dynamic collaboration between humans and robots in a professional context. Besides the University of Coimbra, the project is being developed by the Institute for Systems and Robotics, the University of Madeira, Ruhr-University Bochum and other international partners in Germany and Estonia.

1.2 Objectives and Contribution

The main objective of this dissertation is to study the effect of different robot characteristics and collaborative approaches on human performance. To approach this research, the background and related work provided a useful understanding of the fields and related work, some of the challenges they pose, and novel proposed forms of dealing with those challenges.

A review of the current HRC real-world applications, the observation of a real-life human-human collaborative assembly task and the analysis of existent VR simulations supported the design and development phase of the interactive VR simulation proposed as a solution to reproduce the previously conceptualised scenario and task. In the final stages of the design and development phase, a pilot study was carried out with seven participants to assess problems or inconsistencies in the simulation and the evaluation procedures.

To evaluate human performance in HRC, the formal evaluation procedures were established, and a final study with 36 recruited subjects was accomplished. The participants' physiological and psychological data were collected, processed and analysed. The careful evaluation and interpretation of the obtained results are shown and discussed in this dissertation.

This research contributes to better understand how human performance can be affected by different types of robots and collaboration approaches while executing a collaborative assembly task through VR. The interactive VR simulation de-

[1] https://proactionlab.fpce.uc.pt/en/news-entry/neurocobots-enhancing-human-robot-collaboration

signed and developed in the context of this dissertation is also shared as a contribution to future research.

1.3 Document structure

This document is divided into seven chapters. The introduction chapter includes the motivation, context, objectives and contribution of this research and also describes the structure of this document. The second chapter presents the background and related work concerning the research's main topics, robots and human-robot collaboration. The third chapter explains the research phases, the approach followed, and the research and experimental tools used. Chapter four details the conceptualisation, design and development of the interactive VR simulation. Chapter five shows the evaluation measures and procedures, the test conditions, the study participants' descriptions, the data collection, description, processing and analysis, and the results. The sixth chapter contains the study limitations, results discussion and future work. Finally, the last chapter presents a conclusion of this work.

The present chapter provided the necessary introduction for this research. The following chapter introduces the background and related work of the two main topics studied, robots and HRC.

Chapter 2

Background & Related Work

This chapter presents the background and related work regarding the main topics of this dissertation. The first section presents the topic of robots, followed by a brief overview of relevant subjects in Human-Robot Collaboration (HRC).

2.1 Robot Types and Uses

The term *robot* originated from the ancient Czech word *robota*, meaning *corvée* or forced labour, and was coined by the Čapek's brothers in the play R.U.R., "Rossum's Universal Robots" (1920) to denote one of its fictional humanoid characters. What humans consider robots to be at present might not be the same things we will consider robots to be in the future; for now, we still call the autonomous vacuum cleaners robots, but how long until it is assumed all vacuum cleaners are autonomous and subsequently the lack of necessity to call them robot vacuum cleaners and just vacuum cleaners. The concept of a robot is constantly changing (Dautenhahn, 2013, 2018) primarily due to technological advancements, continuously expanding the boundaries for how robots can be constituted and the settings in which they can act (Alenljung et al., 2017). According to the International Federation of Robotics (IFR) and International Standards Organization (ISO), a robot is defined as "an actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment to perform intended tasks, while mechanisms that lack the number of programmable axes or that are fully teleoperated, end-effectors are called robotic devices". Software ("bots", Artificial Intelligence (AI), Robotic Process Automation (RPA)), remote-controlled drones, unmanned (aerial, ground and underwater) vehicles, voice assistants, autonomous cars, Automated Teller Machine (ATM) or cash machines, and smart washing machines are all examples that do not fit the criteria of a robot. An article written on the website Robots from Institute of Electrical and Electronics Engineers (IEEE) (Guizzo, 2020) presents 15 different types of robots, as presented in table 2.1, which are categorised according to shared features amongst robots. The IFR and ISO consider that drones, unmanned vehicles and autonomous cars are not robots, as these do not fit their robot classification.

Type of robot	Description	Examples
Aerospace	Includes all sorts flying robots but also robots that can operate in space.	SmartBird, Mars rovers
Consumer	Average consumer can buy and use, either for fun or to help with tasks.	Aibo, Roomba
Disaster response	Dangerous jobs like searching for survivors in the aftermath of an emergency.	Packbots
Drones	Unmanned aerial vehicles, exist in different sizes and levels of autonomy.	DJI Phantom, Global Hawk
Education	Aimed at teaching and learning, for use at home or in classrooms.	Sets from Lego, EMYS
Entertainment	Designed to evoke an emotional response.	RoboThespian
Exoskeletons	Physical rehabilitation and for aiding paralysed patient's to walk again.	Ironhand, EksoVest
Humanoids	Robots that resemble humans in appearance and sometimes in behaviour.	Asimo, Geminoid series
Industrial	Repetitive tasks, also includes collaborative factory robots.	Unimate, Baxter
Medical	Healthcare applications, also bionic protheses and exoskeletons.	da Vinci, bionic prostheses
Military & Security	Scout for explosive devices, carrying heavy gear, surveillance.	PackBot, BigDog, Cobalt
Research	Conduct research in universities and research laboratories.	Emys, Baxter
Self-Driving Cars	Autonomous vehicles that can drive themselves around.	Tesla, Waymo
Telepresence	Allow humans to be present somewhere without actually being there.	Double 3, Kubi, Ava
Underwater	Robots whose primary deployment environment is water.	Aquanaut, Ocean One

Table 2.1: Types of robots according to the IEEE Robots website (Guizzo, 2020).

Since forever, humans have been fascinated and curious by nature, and since then, we also have been attempting to understand it and simulate it, inspired by nature's processes and, particularly ourselves, human beings. This made us build machines that attempt to simulate nature's appearance and behaviour, from a simple robotic articulated arm, based on humans, as depicted in figure 2.1; to RoboBee, a small flapping-wing robot inspired by the biology of a fly, as depicted in figure 2.2a; to even humanoid robots which are built as similar as possible to humans' likeness, as depicted in figure 2.2b (Dautenhahn, 2007b; Fong et al., 2003; Mori et al., 2012; Terveen, 1995).

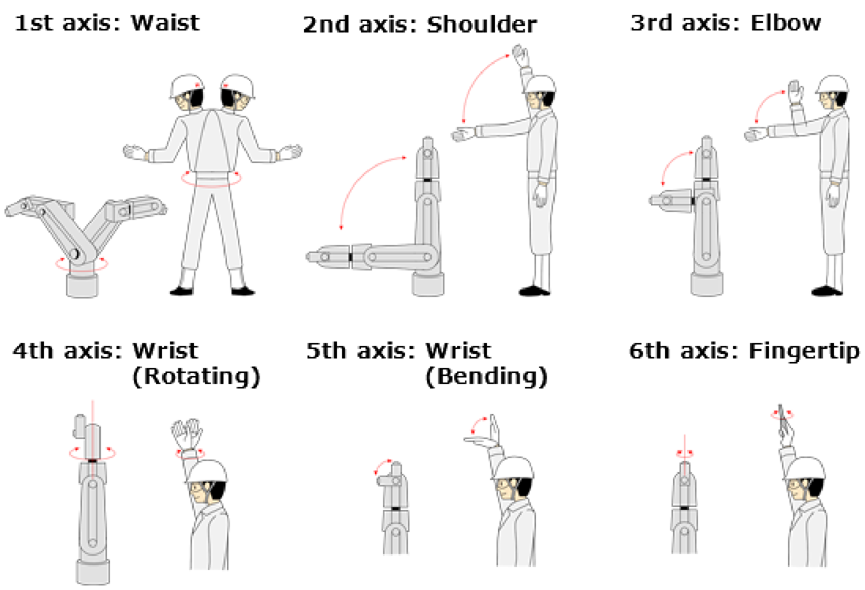
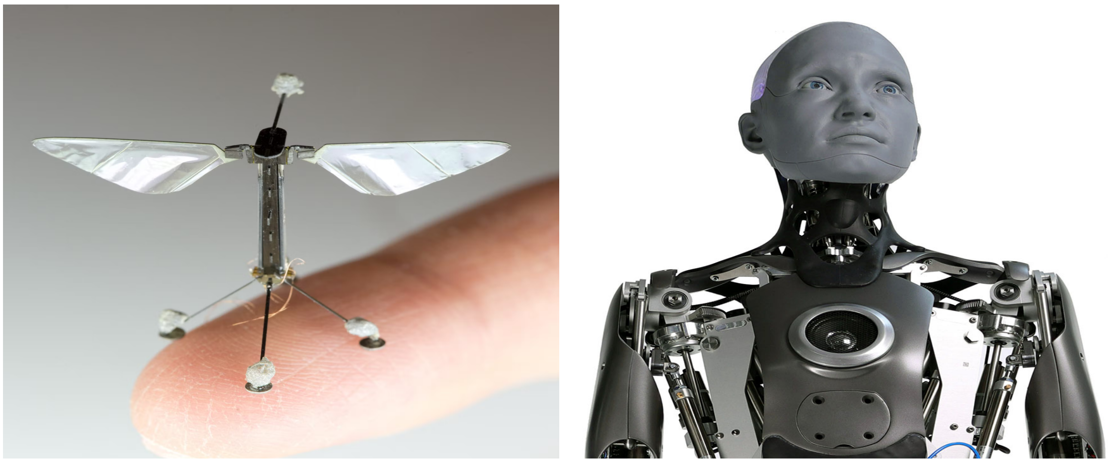


Figure 2.1: A comparison between robot and human movement.

Robots are ever more present in today's world and society (Bieller, 2021); as expressed previously, technological achievements are the primary propellant for the evolution of the robotics field, and as technology evolved, so evolved the latter (Alenljung et al., 2017; Demir et al., 2019). Although the industry is still the most prominent application area for robots, other areas are increasing in popularity, e.g. the area of service has seen an increase in the presence of both professional and personal/consumer robots (Bieller, 2021). Furthermore, recently there has



(a) RoboBee (b) Ameca

Figure 2.2: Different types of robots inspired by nature.

been an increase in the number of socially interactive robots in human environments, and their level of participation in everyday activities is becoming more sophisticated (Dautenhahn, 2007b; Thrun, 2004). Robots are very complex and can vary along multiple dimensions, e.g., appearance, kinematics, behaviour, mechanical structure, etc.; and more specifically in Human-Robot Interaction (HRI), e.g., task type and criticality, morphology, interaction roles, autonomy level, etc., (Yanco and Drury, 2004) and its the multitude of dimensions that make it cumbersome to classify robots generally, hence they are frequently categorised according to their intended application(s) and area of application(s). In the next subsection, different applications will be presented to the main groups of robot development.

2.1.1 Application Areas

Most robot deployments fit into two main categories, based on their application area: industrial and service (professional and personal) areas (Alenljung et al., 2017; Bieller, 2021; Thrun, 2004). The service area has the highest expected growth rate in society (Alenljung et al., 2017; Bieller, 2021). A robot application is described as how the robot is intended to be used and its purpose. However, this does not mean all robots from an application category are equal, it means that robots inside the same category possess similar capabilities or functionalities. Accordingly, in table 2.1, Aibo and Roomba are both consumer service robots, yet they serve completely different purposes, Aibo is a robotic companion dog, while Roomba is a robotic vacuum cleaner. These two main areas of robot development are presented in further detail below.

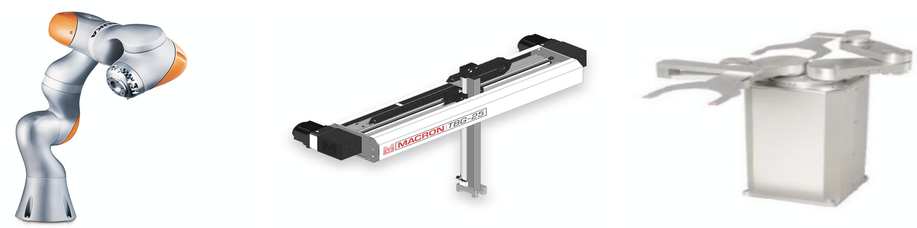
Industry Robots

The industry is one of the primary areas of robots application, where the global average of robot density in the manufacturing industry, the ratio between humans and robots, has grown by 12%, with a total of 384000 industrial robot installations in 2020 (Bieller, 2021). The IFR defines *multipurpose industrial robot* as an :

> Automatically controlled, reprogrammable, multipurpose, manipulator that is programmable in at least three axes, either fixed in place or mobile and intended for and typically used in industrial automation applications.

The IFR classifies industrial robot types according to their mechanical structure, which also dictates the robot kinematics i.e. the robot's motion; different mechanical structures allow for different types of movement, which also translates in different coverage ranges and reach; specific applications may benefit from a specific type of mechanical structure and movement, e.g. parallel robots allow for very fast part handling due to most of the mechanisms being fixed, most commonly in a ceiling. Figure 2.3 presents different types of robots based on their mechanical structure and the IFR classification:

- **Articulated**: a robot whose arm has at least three rotary joints;
- **Cartesian** (linear/gantry): a robot whose arm has three prismatic joints and whose axes are correlated with a cartesian coordinate system;
- **Cylindrical**: a robot whose axes form a cylindrical coordinate system;
- **Parallel**: a robot whose arms have concurrent prismatic or rotary joints;
- **SCARA**: a robot which has two parallel rotary joints to provide compliance in a plane;
- **Others**: robots not covered by one of the above classes;



(a) Articulated robot (b) Cartesian robot (c) Cylindrical robot

(d) Parallel robot (e) SCARA robot

Figure 2.3: Different types of robots classified by their mechanical structure.

Most often, industrial robots are deployed to completely replace the human part in dangerous, dirty, dull or dear settings or jobs: dangerous jobs are self-explanatory they impose risks for human life, and therefore are safer to conduct these jobs with robots, e.g. bomb disposal and space exploration. Dirty jobs are activities that most people don't know or don't think about but need to be done, sometimes they present health hazards, e.g. sewer reconnaissance and mine exploration. Dull jobs are tedious and repetitive tasks where human capabilities would be going to waste e.g. handling and cleaning or disinfecting. Lastly dear jobs are simply activities in which using robots would be more efficient, either economically or timely (Demir et al., 2019; Marr, 2017).

According to the IFR statistics from World Robotics 2021, the more prominent worldwide industry areas for robot deployment are electronics, automotive, metal and machinery, plastic and chemical products, food and lastly, others/unspecified. Regarding specific applications within these areas, handling, welding, assembling, clean rooms, dispensing and processing are amongst the most common, among which handling applications take the lead with over 150000 units of robots installed worldwide for three years in a row (Bieller, 2021). The table in 2.2 summarises the main differences between typical industrial robots and collaborative industrial robots. Having provided an overview of industrial robots, the next section focuses on service robots.

Characteristic	**Industrial robots**	**Collaborative industrial robots**
Role	Replace worker	Assist worker
Human interaction	Commands, and programming assigning locations movements and gripping.	Intelligent interaction: gesture recognition, speech recognition, and anticipating operator moves.
Camera and Computer Vision	External camera, and external system, when exists.	Built -in standard (as part of the cobot), coupled with artificial intelligence.
Workspace	Separate safe workspace for robots and operators. Usually fenced.	Sharing the same workspace. No fencing.
Re-programming	Rare	Frequent
Physical disruptions	Mostly hazardous. Setup required for re-initiation.	Safe, with easy re-initiation.
Agility	Rapid motions	Slow motions
Payload	May be heavy	Not heavy
Dynamic environment	No	Yes

Table 2.2: The major differences between industrial robots and collaborative robots. Adapted from (Cohen et al., 2021).

Service Robots

The service area is the second biggest area with immense opportunity for robot implementation and use (Bieller, 2021). As previously mentioned, professional and personal *service robots* have the highest expected growth rate in society. The IFR defines a service robot as:

> "A robot which operates semi- or fully autonomously to perform services useful to the well-being of humans and equipment, excluding manufacturing operations."

The main difference between service robots and industrial robots is the environment or settings in which the robots operate, and while industrial robots are deployed in industrial settings to replace humans in tasks, service robots act mainly outside of industrial environments to assist humans (Alenljung et al., 2017; Thrun, 2004).

(a) Freight (b) PackBot (c) Da Vinci

Figure 2.4: Different types of professional service robots. Freight (a) is used for transportation in warehouses; PackBot (b) is used for military dangerous missions; Da Vinci (c) is used as a remote surgical robot.

(a) Vector (b) Humman Support Robot (HSR) (c) Emys

Figure 2.5: Different types of consumer service robots. Vector (a) is used as a companion robot; HSR (b) is used to watch family members and fetch objects; Emys (c) is used in education to teach foreign languages to kids.

Accordingly to Thrun, professional service robots are defined as robots designed to assist humans in achieving professional goals, e.g. nuclear waste cleanup (Falter et al., 1995), and abandoned mines navigation and mapping (Thrun et al., 2003). Examples of professional service robots and their applications are presented in figure 2.4. Although personal service robots are also designed to provide humanitarian assistance and are capable of providing entertainment, they differ by being employed in domestic settings, e.g. vacuum cleaners, elderly care and therapy (Thrun, 2004), as opposed to the deployment in professional settings with the professional service robots. Examples of personal service robots and their applications are presented in figure 2.5. According to the IFR statistics from World Robotics 2021 (Bieller, 2021), in 2020, there were over 130000 new professional service robots and over 19 million new consumer service robots. The primary applications, in terms of unit sales, are transportation and logistics, professional cleaning, medical robots, hospitality and agriculture; also presented in these statistics were the top application trends for service robots, which include autonomous mobile robot (AMR) and delivery robots, medical and rehabilitation, partially due to Covid-19, cleaning and disinfection, social robots, i.e. telepresence, and lastly automated restaurant, i.e. staff support and reducing human contact.

2.1.2 Robot Morphology

There are many different robot morphologies, and as mentioned previously, most of them are inspired by nature. Each robot has unique aesthetic features, ranging from size, shape, axis, degrees of freedom (DOF), materials, etc. Thus, when it comes to robot morphology, most robots are categorised as anthropomorphic, i.e., human-like, zoomorphic, i.e. animal-like or mechanic, i.e. machine-like. The figures 2.6, 2.7 and 2.8 display different robot morphologies, from human-like to animal-like and machine-like.

(a) Asimo (b) Geminoid DK (c) Kismet

Figure 2.6: Robots whose appearances resemble humans or human-like.

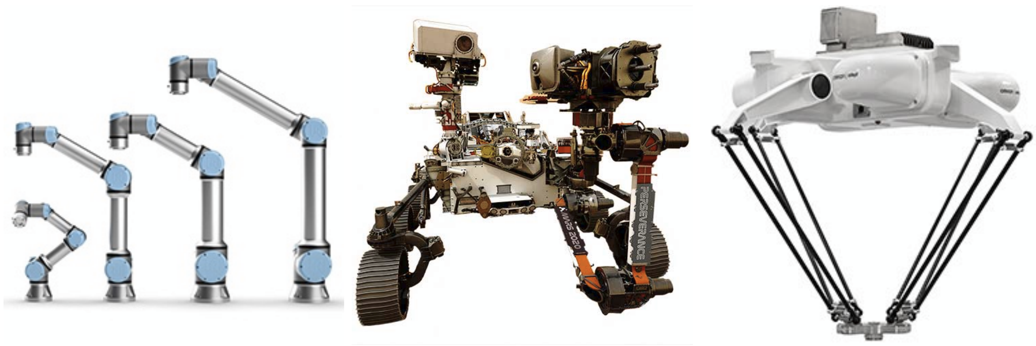
However, robots are frequently classified on a scale that relates their appearance with anthropomorphism, i.e. attribution of human characteristics, in this scale, robots can be classified from machine-like, robotness or productness to human-like or humanness (Kwak, 2014). In this classification scale, a traditional industrial robot whose body consists of an articulated arm is categorised as a product-oriented robot, and a humanoid social robot, whose appearance resembles that of a human, is categorised as a human-oriented robot. According to Di Salvo, a machine or product-oriented robot appearance is designed to comply with the robot's functionalities, while on the other end, a human-oriented robot is designed to resemble a human's appearance (DiSalvo et al., 2002).

Research on robot morphology has shown that robots' appearance influence how

(a) ACM-R5H (b) Aibo (c) Paro

Figure 2.7: Robots whose appearances resemble animals or animal-like.

humans perceive the robots and their capabilities (Fong et al., 2003; Goetz et al., 2003; Gray and Wegner, 2012). Fong et al. state that appearance biases interaction and that a robot's morphology should match its intended function. The uncanny valley theory, proposed by Mori et al., discusses a feeling of uncanniness that human beings, and other living beings, exhibit when interacting with entities whose appearance resembles their own (Mori et al., 2012). The uncanny valley theory is presented in further detail in subsection 2.1.4.



(a) UR cobots (b) Perseverance (c) Quattro

Figure 2.8: Robots whose appearance is mechanical or machine-like.

DiSalvo et al. (2002), surveyed a robot's heads appearance to measure people's perception of humanness and the importance of facial features presence such as a nose, a mouth, eyes and eyelids and in (Goetz et al., 2003) where the authors studied the effect of matching appearance and behaviour to improve human-robot cooperation. Statistical analysis showed a systemic relation between a robot's appearance and people's perception of the robot, as well as people's willingness to comply with the robot's instructions. The authors also state the importance of a robot's morphology and behaviour towards developing better personal service robots.

Also worth mentioning is the research by Gray and Wegner (2012), where robots with higher anthropomorphic levels, or more human-like, are perceived as having a mind. Lastly, (Kwak, 2014) studied the effect of different appearances on the perceived social presence, sociability and service evaluation and concluded that human-like robots performed better in human-robot social interaction.

2.1.3 Robot Behaviour

Industrial robots are mostly categorised as fully autonomous robots, requiring very minimal or no human intervention, usually even having their operational space separated from humans. In these scenarios, the majority of interfaces used are programming languages and graphical simulation tools, with some robots learning by demonstration (Thrun, 2004). On the other end of the spectrum, the area with the largest expected growth, there is the service robots area and more specifically social robots, which are defined as robots capable of social interaction, not only human-robot interaction but also robot-robot interaction in individualised societies (Fong et al., 2003); as social robots are deployed in the same social and spatial settings as humans, the requirements regarding the quality of the interaction between humans and robots increases (Lindblom and Cort, 2016).

A robot's behaviour can be described as the set of actions the robot is able to perform; from low-level actions, like turning an actuator on and off, to high-level actions, such as contouring an object, inverse kinematics or speech output. As previously mentioned, for a long time, humans have been trying to simulate nature's processes, particularly human beings, not only morphologically but also behaviourally. Weistroffer et al. investigated the acceptability of HRC regarding, amongst others, the robot's movement profile. Two profiles were used, machine-like and human-like. The authors concluded that motion perception is influenced by the robot's morphology, where human-like motion was preferred in industrial robots but not on humanoids (Weistroffer et al., 2013).

Most of the actions mentioned have to be pre-programmed in the robot's software, but machine learning and AI have allowed for big developments in regard to what is considered the robot's 'mind'. Dautenhahn presents two important phases regarding the evolution of socially intelligent robots as well as its contributing factors: the first phase of research was focused on the simulation of specific isolated human activities such as playing chess or proving theorems; although in the second phase, with the emergence of AI, a similar paradigm was followed, research was more focused on the 'mind' instead of only one or few human activities, which proved to be a bigger challenge than expected (Dautenhahn, 2007b). Goodrich and Schultz state that every robot application requires some form of interaction, either remote or proximate, from teleoperation, where the interaction might be more evident, i.e. the operator controlling the robot remotely through dedicated interfaces to a fully autonomous robot, where the interaction can be high-level supervision and direction of the robot, i.e. monitoring the robot behaviour through observation or even programming its behaviour through software (Goodrich and Schultz, 2007).

2.1.4 Uncanny Valley

The uncanny valley is a theory presented by Masahiro Mori in 1970 and depicted in figure 2.9. "I have noticed that, in climbing toward the goal of making robots appear human, our affinity for them increases until we come to a valley, which I call the uncanny valley" (Mori et al., 2012).

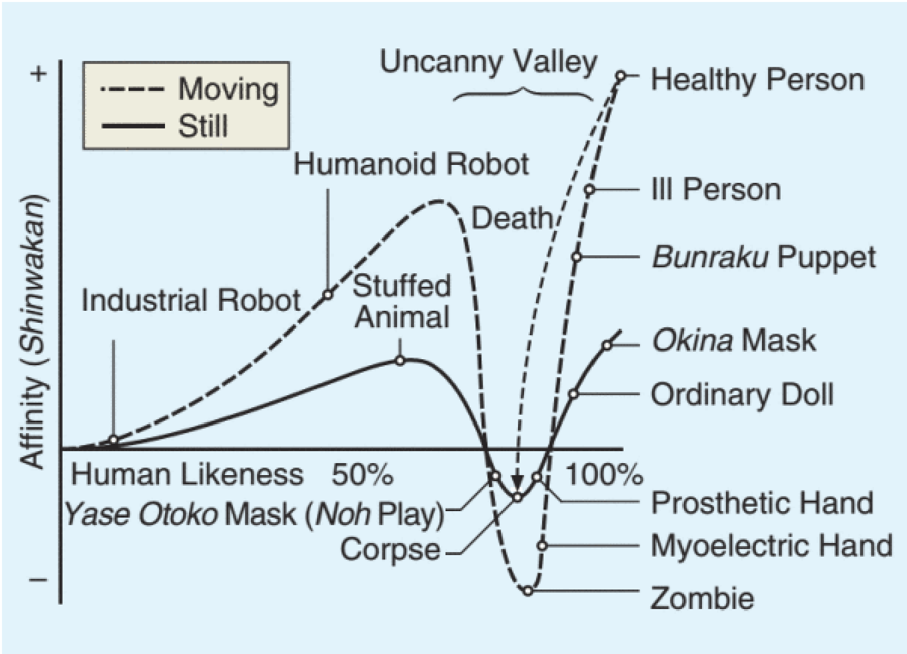


Figure 2.9: The graph depicts the uncanny valley, the proposed relation between the human likeness of an entity and the perceiver's affinity for it (Mori et al., 2012).

The term "uncanny valley" is frequently used in the field of robotics, but it also has its place amongst computer-generated imagery, e.g., in video games, films and animations (Gee et al., 2005; MacDorman et al., 2009). As previously mentioned, in (DiSalvo et al., 2002) research was done on robots' faces, the more facial features a robot has, meaning eyes, nose, mouth, *et cetera*, the more human-like it was perceived. In their research, the authors concluded that facial features, such as the nose, eyelids and mouth were deemed of great importance when it came to human likeness ratings. The head shape is also an important factor, for the more square the head is, the less human-like it was perceived. Varied chin length and forehead height were also studied, and the authors found that a shorter chin made the eyes appear larger, causing *kindchenschema* (cuteness) (Powers and Kiesler, 2006). (Rosenthal-von der Pütten and Krämer, 2014) claim that in the same sense that attractive people are also rated more positively in other aspects holds true for "virtual" agents (robots) as well.

The expected cognitive capabilities of a robot and its expected behaviour are linked to its appearance, and human-like robots prompt attributions of mind to the robots (Gray and Wegner, 2012). People systematically preferred robots for jobs when the robot's human–likeness matches the sociability required in those jobs (Goetz et al., 2003). With the development of technology and the recent advancements of AI, some people defend the existence of an uncanny response concerning the "mind" (Stein and Ohler, 2017). However, Mori's theory has also been criticised for being simple and because its dimensions are not well defined since they are themselves complex dimensions (Bartneck et al., 2009a). The uncanny valley theory is also criticised for neglecting relevant factors besides the characteristics of the robots, such as participants' age, culture, religion or previous experiences (Gee et al., 2005). Lastly, most studies around this theory only focus on the facial features instead of the agent's full body.

Although the human feeling of *uncanniness* can be related to appearance, researchers cannot fully pinpoint which appearance traits are responsible for it (Rosenthal-von der Pütten and Krämer, 2014). Furthermore, Gee et al. defend that this theory serves as an indicator that it is easier to detect discrepancies in movement rather than in appearance, which means that humans feel more *uncanniness* due to the movement rather than the appearance (Gee et al., 2005). In the following section, the background and related work concerning HRC is shown, presenting definitions for interaction and collaboration, different types of collaborative approaches, and lastly, the safety and user experience components in HRC.

2.2 Human-Robot Collaboration (HRC)

HRI and HRC are both challenging, inter- and multidisciplinary fields that comprise knowledge from study fields such as psychology, cognitive science, social sciences, artificial intelligence, computer science, robotics, engineering, human-computer interaction, etc. (Bauer et al., 2008; Dautenhahn, 2007a). Having started to emerge in the mid-1990s both fields of HRI and HRC are considered to be in their infancy. The main contributions for the fields have originated from scientific meetings frequented by people from different scientific communities, and also incentives in the form of competitions, where a goal is defined and the scientific community attempts to achieve it (Goodrich and Schultz, 2007).

2.2.1 Interaction and Collaboration

The terms *interaction* and *collaboration,* are very frequent in the fields of HRI and HRC, still it is important to understand what they entail to avoid semantics errors. According to the Oxford Advanced Learner's Dictionary, *interaction* is defined as "two things having an effect of each other", while *collaboration* is defined as "the action of working with someone to produce something" (Oxford Advanced Learner's Dictionary, 2022a). However, in the Cambridge Dictionary *interaction* is defined as "an occasion when two or more people or things communicate with or react to each other", while *collaboration* is defined as "the situation of two or more people working together to create or achieve the same thing" (Cambridge Dictionary, 2022a). Considering these definitions, it can be interpreted that *interaction* requires two or more entities to occur, as well as mutual action, influence or communication; and that *collaboration* requires the existence of a shared or common goal between two or more working entities.

According to Goodrich and Schultz, interaction can be split into two broad categories: remote interaction and proximate interaction. In remote interaction, humans and robots are not co-located, they are separated spatially or even temporally (Goodrich and Schultz, 2007). For e.g., Perseverance, one of NASA's rovers, is separated both in space and time, the figure 2.8b depicts the rover. In proximate interaction, humans and robots are co-located, as it happens, for example, when social service robots are deployed in elderly care centres, where the robots will be close to humans.

Bauer et al. define *interaction* as any action that involves humans and robots, not implying profit for one of the parts, while *collaboration* is defined as working with someone or something to achieve a common goal; interaction is then considered to be a more general term that includes collaboration (Bauer et al., 2008). In a review paper by Matheson et al., four different interaction types are presented: (i) coexistence, when the human and the robot are in the same environment but don't interact; (ii) synchronised, when the human and the robot work in the same environment at different times; (iii) cooperation, when the human and the robot work on separate tasks but are in the same environment at the same time; and lastly (iv) collaboration, when the human and the robot work on the same task, in the same environment, and simultaneously (Matheson et al., 2019). These definitions present a good base for differentiating both concepts and to be applied to the fields of HRI and HRC. The relationships between the previously presented terms becomes clearer in an illustration, as seen in figure 2.10 that synthesises the framework by Castro et al., and which will guide the work of this dissertation.

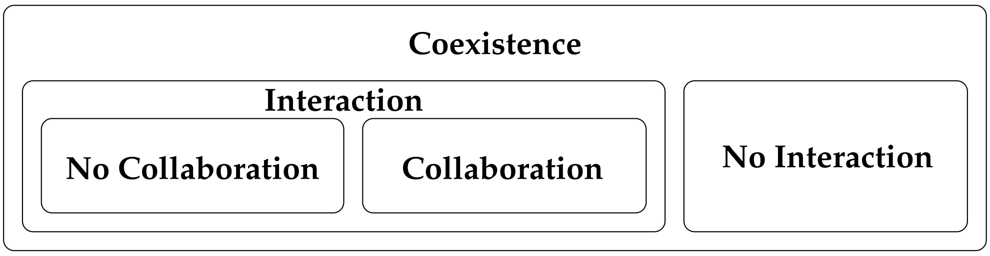


Figure 2.10: Relations between different concepts regarding humans and robots. Adapted from (Castro et al., 2021).

Communication

Communication is another crucial concept related to the above-mentioned definitions, where its absence can significantly impact the effectiveness of human-robot interaction and collaboration, leading to low confidence or trust, which negatively impacts the interaction and the collaboration (Freedy et al., 2007; Kolbeinsson et al., 2019; Weiss et al., 2021).

When applied to a human-human context, communication can take different forms related to our five sensory organs, i.e. eyes, ears, skin, nose, and mouth, allowing humans to exchange information bi-directionally through various interfaces, mainly speech, vision and touch. According to Yanco and Drury, and Goodrich and Schultz, in HRI and HRC, communication is dependent on the communication medium and the communication format, with the media being transmitted through:

- visual displays, as Graphical User Interface (GUI), AR/VR interfaces;

- gestures, hand and facial movements, and communication of intention based on movement, intent signalling;
- speech and natural language, non-speech audio, frequently used in alerting;

Additionally, the format of the communication varies across the previously presented mediums e.g. sound can be exchanged in the form of auditory cues and in the form of speech (Goodrich and Schultz, 2007; Yanco and Drury, 2004). The authors Castro et al. also present different types of communication interfaces used in HRC and the respective sensors responsible that serve as interfaces or communication channels: (i) directly encoded orders, in the form of graphical user interfaces or terminal instructions; (ii) gestures and motion, in the form of visual and depth cameras, LIDAR and inertial sensors, accelerometers and gyroscopes; (iii) voice, through a microphone; (iv) haptics and contact, in the form of tactile arrays and load cells that convert force into electrical signals; and lastly (v) muscular and brain activity, through electromyography, and electroencephalography, analysing muscle's and brain's electrical activity (Castro et al., 2021).

In the work of Bauer et al., communication is approached as communication of intentions and is divided into two types: explicit and implicit. *Explicit communication* of intention, which is deliberate and used to ensure that the receiving end gets information about the intentions, e.g. speech, gesture and haptics. *Implicit communication* of intention is communicated implicitly and unconsciously, mostly through actions, and requires a lot of interpretation by the receiving end, e.g. emotion, and physiological signals. The authors also state that intentions communicated and interpreted correctly lead to joint intentions, which in turn leads to the joint actions required for efficient collaboration (Bauer et al., 2008). The main ways of communicating intention are summarised in table 2.3.

<table>
<tr><th colspan="14">Communication of Intention</th></tr>
<tr><td colspan="2">Speech</td><td colspan="6">Gesture</td><td>Action</td><td colspan="2">Haptic Signal</td><td colspan="3">Physiological Signal</td></tr>
<tr><td>Explicit Information</td><td>Emotion</td><td colspan="2">Head/ Eyes</td><td colspan="2">Communicative Gesture</td><td colspan="2">Manipulative Gesture</td><td>Proactive Task Execution</td><td>Force/ Torque</td><td>Angles/ Orientation</td><td colspan="3">Approval/ Arousal</td></tr>
<tr><td></td><td></td><td>Facial Express.</td><td>Eye Gaze</td><td>Pointing</td><td>Signs</td><td>Motion</td><td>Object Usage</td><td></td><td></td><td></td><td>Heart Rate</td><td>Brain Activity</td><td>Muscle Activity</td></tr>
</table>

Table 2.3: Main ways of communicating intentions, explicit communication unmarked and implicit communication marked in grey. Adapted from (Bauer et al., 2008).

Interaction Roles

Building upon Norman's model for Human-Computer Interaction (HCI), Scholtz introduced a taxonomy of different interaction roles that can be attributed to HRI, thereafter presented by several authors (Goodrich and Schultz, 2007; Scholtz, 2003; Yanco and Drury, 2004). According to Scholtz, and Yanco and Drury, there are five different roles: supervisor, operator, teammate, mechanic/programmer, and bystander. Goodrich and Schultz then added the roles of mentor and information consumer. Table 2.4 provides an overview of the different applications, and their possible corresponding interaction roles, a brief description of each role is presented afterwards.

Application area	Interaction	Role	Example
Military	Remote	Human is information consumer	Dangerous reconnaissance information
Education	Proximate	Robot is mentor	Classroom assistant
Entertainment	Proximate	Robot is teammate/peer	Social companion
Domestic	Proximate	Human is supervisor	Robotic vacuum

Table 2.4: Examples of different roles humans and robots can assume in different application areas. Adapted from (Goodrich and Schultz, 2007).

Supervisor, characterised as monitoring or observation, there is no need for direct control;

Operator, characterised by requiring some interaction, e.g. teleoperation or changing the robot's behaviour;

Teammate/peer, characterised by team/peer work with a robot to accomplish a task;

Mechanic/programmer, characterised by physically altering the robot's software or hardware;

Bystander, characterised by the human requiring understanding of the situation but not requiring interaction;

Mentor, characterised by the robot being in a teaching or leadership role;

Information consumer, characterised by the user receiving information without controlling the robot.

2.2.2 Levels of Collaboration

Collaboration means to work together, to achieve a goal or produce something; automation and collaboration are fundamentally different in that automation actively seeks to remove the human from the task being performed, whereas the latter tries to maximise both human and robot capabilities through collaboration. To address collaboration in HRC, it is necessary to keep in mind the context, scenario, goal of the activity, tasks, etc., and specify which humans are involved, what kind of robots, and what kind of collaboration (Kolbeinsson et al., 2019). Types or levels of collaboration are a form of classification meant to distinguish different levels of collaboration that can occur between humans and robots, just like there are classifications for different robot automation levels.

The IFR states that in the industry area, there are five admissible collaboration levels, described in table 2.5 and illustrated in figure 2.11, and that out of those five proposed levels, most collaborative applications in industrial environments, as of today, are of the coexistence and sequential or synchronised collaboration type.

Level of collaboration	Cell	Coexistence	Sequential collaboration	Cooperation	Responsive Collaboration
Requirement for intrinsic safety features vs. external sensors	Fenced Robot	No fence and no shared workspace	Robot and worker both active in the workspace but movements are sequential	Robot and worker work on the same part at the same time - both in motion	Robots respond in real-time to movement of workers.

Table 2.5: Levels of collaboration between industrial robots and humans. As the level of collaboration increases (left to right), so does the Requirement for intrinsic safety measures vs external sensors. Source: IFR Position Paper (IFR, 2020) adapted from (Wilhelm et al., 2016).

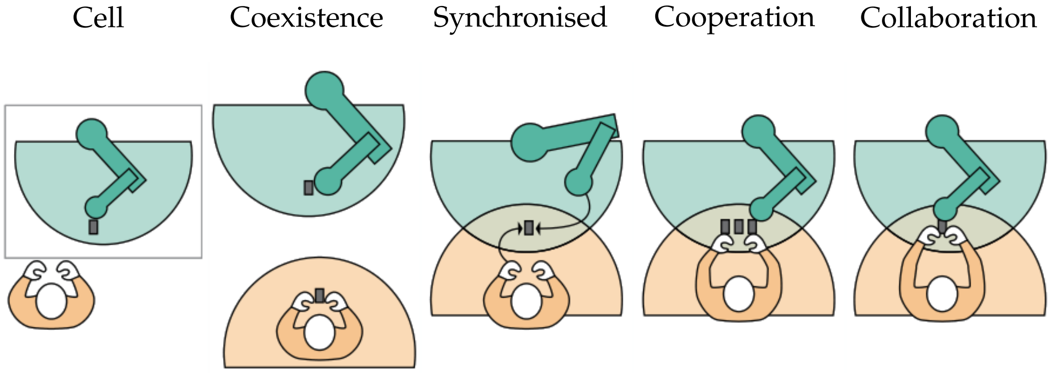


Figure 2.11: Different levels of collaboration as proposed by Wilhelm et al. and IFR. Adapted from (Wilhelm et al., 2016).

2.2.3 Safety in Human-Robot Collaboration

In HRI and HRC, particularly in industrial contexts, there is a concern regarding human safety. Industrial robots are usually separated physically from workers with fences due to the danger they pose to humans. However, collaborative robots are meant to be deployed in scenarios where close proximity (without fences or a cage) between robots and workers should not be a problem, meaning that robots should be equipped with various sensors that allow them to perceive their surroundings (Badia et al., 2022; Matheson et al., 2019).

As mentioned in table 2.5, the higher the collaboration level, the higher the need for safety measures. As an effort to protect operators in industrial settings, the technical committee, ISO/TC 299 Robotics (2016), has presented techniques developed with safety regulations regarding the operation of collaborative robots, where these techniques can be used in conjunction or independently (Shea, 2016):

Safety-Rated Monitored Stop (SMS), the robot stops moving when the operator enters the workspace and resumes motion after the operator leaves the workspace, e.g. loading and unloading parts to the end-effector and inspections;

Hand Guiding (HG), the robot is controlled manually via a hand-operated device that transmits motion, e.g. lift assist and diverse applications where the robot acts like a manual tool;

Speed and Separation Monitoring (SSM), the robot adjusts its motion velocity accordingly to its distance from the operator, the further the operator is, the faster the robot is going to move, also includes a minimum separation distance that when surpassed makes the robot stop, e.g. simultaneous tasks;

Power and Force Limiting (PFL), the robot operation is restricted in both power and force, this allows for intentional and unintentional physical contact between the robot and the human, e.g. applications where physical contact is necessary or tolerable.

Although these techniques provide safer forms of collaboration between humans and robots, the safe operation of collaborative robots must always be assessed properly and take into account multiple factors such as the space, task, robot, end-effector, payload, *et cetera*. Matheson et al. state that early studies were focused on SSM and PFL approaches and that in more recent studies SMS and HG approaches started to emerge, with particular emphasis on HG, due to enabling unskilled users to interact with the robot, by moving its end-effector towards desired interest points and having the robot replicate the transmitted movement (Matheson et al., 2019). However, safety is not only a major concern in industrial collaborative applications where the risk may be higher due to the physicality involved, but also in interaction with service robots, humanoids can also pose safety problems. Dautenhahn further explains that even in applications involving education, therapy or elderly support safety, safety is of critical importance due to the nature of the applications and the vulnerability of the users involved (Dautenhahn, 2007a).

2.2.4 User Experience in Human-Robot Collaboration

Social robots can be approached from three different perspectives: robot-centred view, robot cognition-centred view and human-centred view; however, the human-centred view approach has not been given the same attention as the other two approaches (Dautenhahn, 2007b). Although human-robot interaction is different from human-human interaction, human-animal interaction, and human-computer interaction (Dautenhahn, 2018), various researchers suggest that there is a necessity for methods and techniques for systematically designing and evaluating various aspects of the human-robot interaction, suggesting the use of methods and techniques from the fields of HCI and User Experience (UX) (Alenljung et al., 2017; Dautenhahn, 2007a; Lindblom and Cort, 2016). According to Oxford Advanced Learner's Dictionary definition, UX is described as the overall experience of a person when using a particular product, such as a website, e.g. how easy or pleasant it is to use. According to Cambridge Dictionary definition, UX refers to the experience of someone using a product, system, or service, e.g. whether they find it enjoyable.

HCI and UX are both concerned with the pragmatic and the hedonic aspects of interactive products (Hassenzahl and Roto, 2007; Hassenzahl and Tractinsky, 2006): the *pragmatic* quality is related to fitting user's behavioural goals or "do-goals", i.e. the usability component of UX, including effectiveness, efficiency, satisfaction, ease-of-use, and learnability; and the *hedonic* quality, also named "be-goals", is related to the user's emotional response when interacting with a product, i.e. personal growth or autonomy (Alenljung et al., 2017; Hassenzahl and Tractinsky, 2006; Lindblom and Cort, 2016).

Positive UX is not a part of the product *per-se*, but a part of the result of the interaction between the user and the product. In this way, UX is dependent on the user, on the product and also on the specific circumstances in which the interaction occurs (Hassenzahl and Tractinsky, 2006); thus it is not feasible to have equal user experiences, since these vary highly from individual to individual, as well as with case-specific circumstances (Lindblom and Cort, 2016). Lindblom and Cort also presented relevant challenges regarding the evaluation of user experience in HRI such as the need for an iterative UX design process in HRI; the need to use UX goals to achieve positive UX; and finally, the need of knowledge regarding UX evaluation amongst robot developers (Alenljung et al., 2017; Lindblom and Cort, 2016).

This chapter presented a review of the background and related work, concerning the two main topics of this dissertation, robots and HRC. It allows to better understand the subjects underlying this work, some of the challenges they pose, and ways of overcoming those challenges. The following chapter introduces the approach followed by this research, the research phases, and the hardware and software tools used.

Chapter 3

Research Methodology

This chapter presents and describes the main research phases and methods that guided and were used in this dissertation, as well as the research and experimental hardware and software tools utilised.

3.1 Research Phases and Approach

This research aims to understand how human performance is affected when executing a collaborative assembly task with robots of different anthropomorphic levels and two distinct collaboration approaches, sequential and simultaneous. Related to human performance, the uncanny valley effect response during the collaborative assembly task is also studied. The research approach followed in this research is divided into four distinct but related phases: background and related work, conceptualisation and design, iterative and incremental development and, lastly, formal evaluation, as detailed in the diagram in figure 3.1.

The **background and related work** is the phase where the necessary literature for the research was studied. The two significant issues examined were (i) robots, with an emphasis on application areas, morphologies, behaviours and the uncanny valley effect; and (ii) human-robot collaboration, with an emphasis on the differences between interaction and collaboration, different levels of collaboration, human safety and user experience in human-robot collaboration. This phase had a duration of approximately three months and was crucial for the overall research since it allowed important insight regarding the primary research issues and the knowledge needed for the context to be defined in the next phase.

The **conceptualisation and design** phase defines potential scenarios and tasks to integrate into the interactive Virtual Reality (VR) simulation to be developed. A review of the existent collaborative robots and their use cases (a summary table of this review can be seen in Appendix A) is conducted alongside what was apprehended in the first phase through drawings, schematics and storyboards, allowing to establish a concept and a preliminary design. This phase lasted approximately five months.

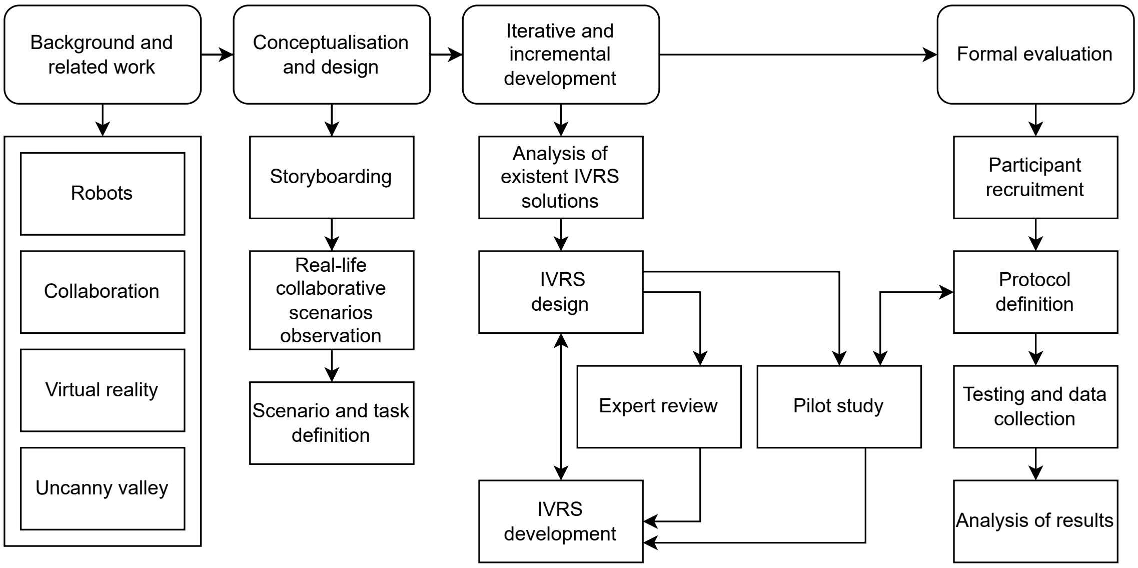


Figure 3.1: A diagram with the research phases and their relations represented by the arrows. IVRS stands for Interactive Virtual Reality Simulation.

The observation and analysis of a real-life human-human collaborative assembly task are carried out to further understand how humans collaborate when performing a similar task to the one later implemented, which consists of a scenario with interlocking building blocks. Lastly, the final scenario and task are defined based on what was gathered from the previous stages. In short, this phase served to identify, conceptualise, contextualise and design a solution for the presented research problem to later proceed with developing the solution itself. In addition, the different test conditions to which participants will be exposed are determined in conjunction with the dependent variables and preliminary tests are done to ease implementation in the development phase of the synchronised data collection from the multiple sensors used.

The physiological measures to consider are the participants' pulse Beats per Minute (BPM) and Inter Beat Interval (IBI); pupil size, blink rate, and gaze fixation duration on the robots, these are measured with the OpenBCI Cyton pulse sensor and the HTC Vive Pro Eye embedded eye-tracking. The psychological measures are the participants' perception of robots' anthropomorphism, animacy, likeability, intelligence and safety, and participants' sense of presence in the interactive virtual reality simulation, these are measured with the Godspeed Questionnaire Series (GQS) and the VR Presence Questionnaire (PQ). Additionally, the task completion time is also measured. The independent variables are robot anthropomorphism, with three robots of low to moderate anthropomorphic levels and the presence of gaze through a monitor screen for two of the robots. The participants perform the same collaborative assembly task with every robot, Kuka LBR iiwa, Baxter and Sawyer, and both types of collaboration: sequential and simultaneous. Thus, the most appropriate experimental design to implement in this study is a within-subjects design, where every participant is subjected to every condition.

The **iterative and incremental development** phase is where the interactive VR simulation is developed in the Unity Real-Time Development Platform as a way to present visual, auditory and haptic stimuli to the participants. In the early stages of this phase, existent and related interactive virtual reality simulation solutions are analysed to identify potential challenges and better understand the required functionalities. The development phase had a duration of approximately six months. An iterative and incremental approach is followed to develop the interactive simulation, meaning there is a repeating cycle of design, development and evaluation by experts through regular meetings, where new functionalities are implemented and later reviewed. In the last stage of the development phase and as preparation for the next phase of the work, a pilot study is done with previously recruited subjects to evaluate the interactive VR simulation by receiving feedback and identifying and correcting the inconsistencies detected.

The final phase, the **formal evaluation**, had a duration of approximately five months and is the phase where the definitive test protocol is determined after analysing the pilot study's feedback, followed by the recruitment of the subjects, the participants are submitted to the tests, and the tests data is collected. To process the participants' data, the Python programming language is used with the software libraries Matplotlib, NumPy, Pandas and SciPy Signal. After processing the data, the Jamovi software is used to analyse it and to obtain the final results through statistics measures, such as repeated measures analysis of variance (ANOVA) for the physiological data and arithmetic averages of the scores of the questionnaires for the psychological data. Additionally, to compare the physiological data with the GQS data, scatterplot diagrams are made to find patterns in the data. The obtained results are evaluated and interpreted and then presented and discussed.

This work was subjected to two instances of formal evaluation, for which was necessary to write and prepare a formal document and a presentation. The first instance, in February of 2022, was an intermediate milestone and the second, in January of 2023, the final milestone. The Gantt chart in figure 3.2 displays the work schedule of the above-mentioned phases, including the documentation phase.

Participants and Ethical Considerations

The study obtained ethical approval from the Ethics and Deontology Committee for Research (CEDI) of the Faculty of Psychology and Educational Sciences of the University of Coimbra: CEDI/FPCEUC:64/1, 22/06/2022. The recruitment of 39 participants was made through email invitations, the test sessions began by providing a summary of the test and asking for the participant's informed consent, ensuring the data collected were anonymous and confidential. The final tests and data collection with the recruited participants were accomplished, excluding three participants for not being able to finish the test session. In section 5.3, the subjects who participated in this research are further detailed.

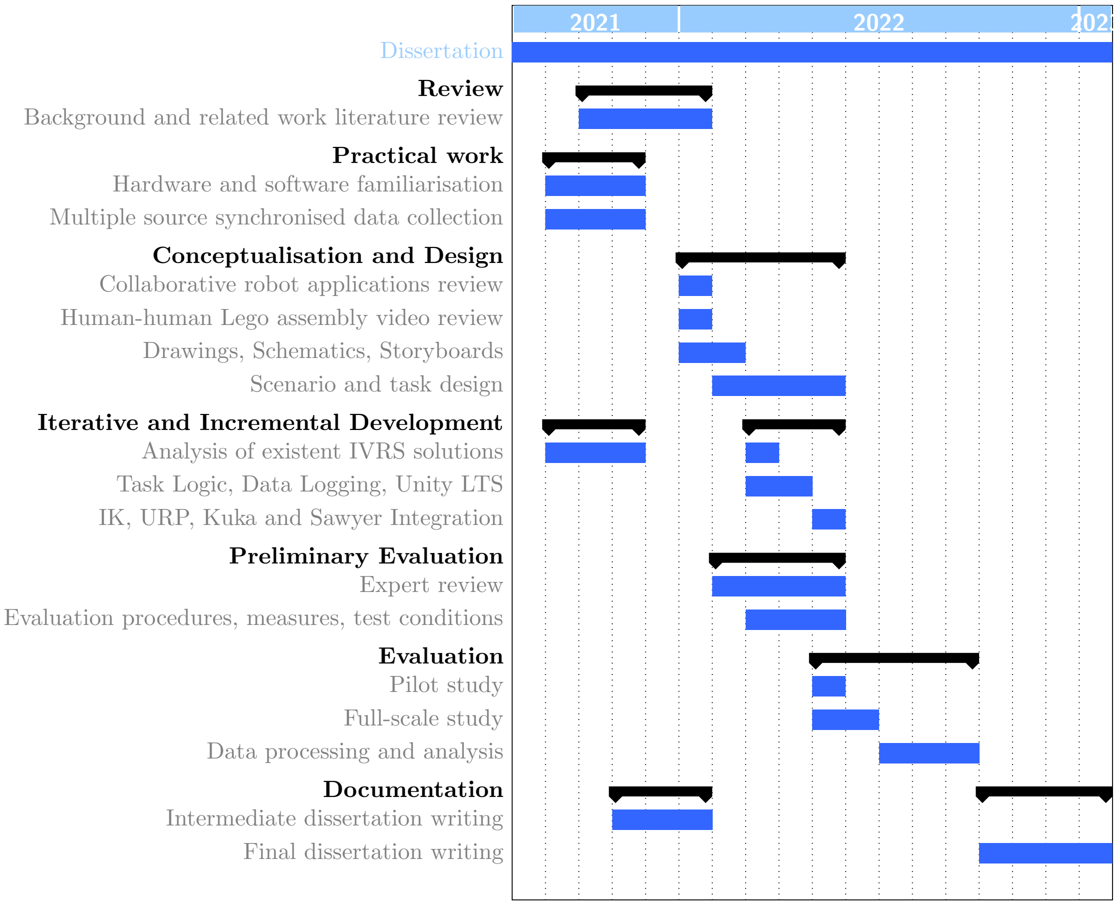


Figure 3.2: Work schedule.

3.2 Research and Experimental Tools

3.2.1 Hardware Tools

Virtual Reality (VR) system

The presented study was conducted in a VR environment, hence the need for the users to use VR hardware equipment to realise the tests. The VR system used is the HTC Vive Pro Eye[1], along with the lighthouses and the two controllers, as depicted in figure 3.3. This VR solution has a built-in eye-tracker capable of analysing gaze origin and direction, pupil position and size, and eye openness. The system is also capable of sound output through built-in adjustable headphones. In this VR system, the users' mobility is limited due to the need for a physical connection to a computer and the system requiring a direct, unobstructed line of sight to the lighthouses, which themselves only allow for five squared meters room-scale. The participants were positioned in the centre of the room to avoid system tracking failures negatively impacting the experiment and the user experience. The use of a VR system in this study allows for safe interaction and collaboration with the robots since there is no physicality involved, as well as faster development and evaluation of the interactive VR simulation.

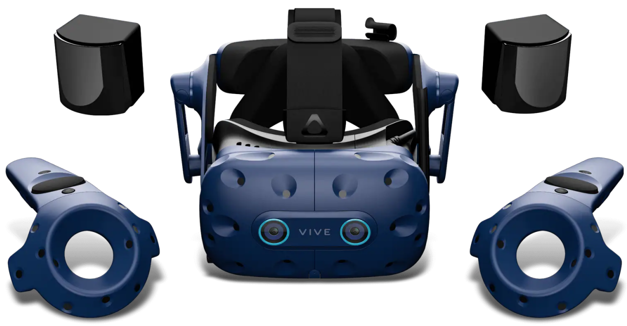


Figure 3.3: HTC Vive Pro Eye VR system, the Head Mounted Display (HMD) in the centre, the two lighthouses in the back and the two controllers in front.

Eye-tracking

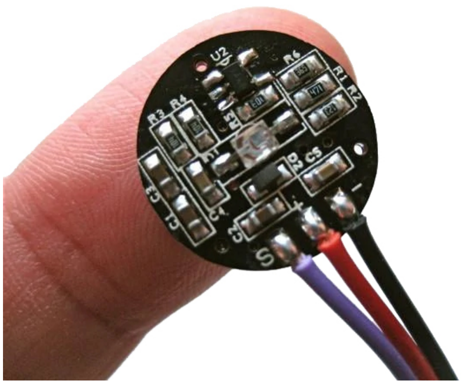
The eye-tracking sensor used is embedded into the VR system[1], more precisely in the HMD and is powered by Tobii. This eye-tracking system is capable of a maximum binocular data output frequency of 100 Hz, it has an accuracy of 0.5°-1.1° degrees within a Field of View (FOV) of 20° and also a trackable FOV of 110°. The interface utilised to integrate the eye-tracking system is the HTC Vive SRanipal SDK and, more specifically, in the case of Unity, the SRanipal Unity SDK. The eye-tracking system's purpose is to track and record different types of eye data, such as gaze origin and direction, pupil position and size, and eye openness. In this study, the analysis of the acquired eye data served to obtain

[1] https://www.vive.com/us/product/vive-pro-eye/specs/

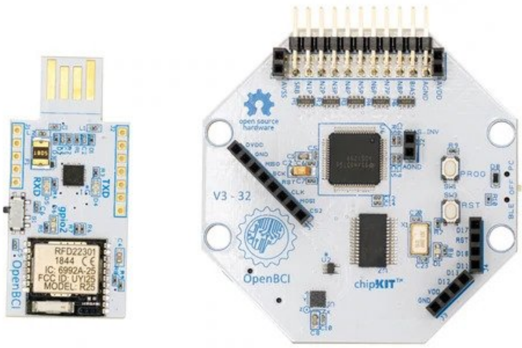
information about the pupils' response through the average pupil diameter, the blink rate and the time spent looking at different scenario locations.

Pulse sensor and Cyton biosensing board

The pulse sensor[2] is a Photoplethysmography (PPG) sensor that uses light and a photosensor to measure variations in the blood circulation volume, as displayed in figure 3.4a. The pulse sensor can be used with micro-controllers such as an Arduino or, in the case of this study, with a biosensing board, such as the OpenBCI Cyton[3], which is also Arduino-compatible, as shown in figure 3.4b. The pulse sensor can be placed either on a fingertip or the earlobe, and it should not be strapped either too tightly or too loose, as this can cause noise that will interfere with the measurement. A three-wired cable is used to connect the pulse sensor to the Cyton board, the red wire takes power from the board, the black wire is the ground, and the purple wire transmits an analogue input in the form of an electrical signal in millivolts to the board. The Cyton board has a power supply of four AA batteries, and the data transmission to the computer is done via the Bluetooth USB dongle. Initially, it was thought that Bluetooth would confer a degree of portability to the whole apparatus, thus allowing users to carry the board themselves. Still, after some tests, it was concluded that the participants' movements also caused the board to move and introduce noise to the pulse data. This issue was solved by fixing the board on a surface and extending the pulse sensor cable so the participants did not need to carry the board.



(a) Pulse Sensor



(b) Cyton board and USB dongle

Figure 3.4: In (a) the backside of the pulse sensor; (b) on the left, its USB programmable dongle for Bluetooth communication and on the right, the Cyton board.

[2] https://docs.openbci.com/ThirdParty/Pulse_Sensor/Pulse_Sensor_Landing/
[3] https://docs.openbci.com/Cyton/CytonLanding/

3.2.2 Software Tools

Unity Real-Time Development Platform

Unity[4] is a cross-platform game engine used not only for game development but also as a tool for research activity. The software has a large support community, especially regarding compatibility with third-party software. The present study uses the Unity platform to create and develop a three-dimensional interactive VR simulation that fundamentally presents different types of visual, auditory and tactile stimuli to the human in junction with the HTC Vive VR system. The initial Unity version used for the development was Unity 2019.2.8f1. Still, to assure and ease future compatibility and the development of the project, it was later updated to the Long-term support (LTS) release Unity 2020.3.30f1 version.

SRanipal Runtime and Software Development Kit (SDK)

The SRanipal[5] Runtime is required to enable eye-tracking capability and is mainly used to verify the eye-tracking system status. The SRanipal Unity SDK is used to integrate eye-tracking features directly in the Unity project, namely eye-tracking data recording and collection.

Lab streaming layer (LSL)

The Lab Streaming Layer (LSL)[6] is a system used for multiple source data collection; the system handles networking, time-synchronisation, and collection of the data itself with disk storage. This study uses LSL to stream the pulse sensor readings from OpenBCI to Unity and record and store the pulse data at the same frequency as the remainder of the acquired physiological data.

OpenBCI

The OpenBCI Graphical User Interface (GUI)[7] is a software tool for visualising, recording, and streaming data from the OpenBCI Boards, such as the one used in this study and previously mentioned in 3.2.1. The software is used to monitor pulse data in real-time to verify if there was noise in the received signals and to stream the data from the pulse sensor, pulse BPM and IBI to Unity using LSL, to keep the collection of data synchronised with the remaining sources.

[4] https://unity.com/

[5] https://developer-express.vive.com/resources/vive-sense/eye-and-facial-tracking-sdk/

[6] https://labstreaminglayer.readthedocs.io/info/intro.html

[7] https://docs.openbci.com/Software/OpenBCISoftware/GUIDocs/

Open Broadcaster Software (OBS)

The OBS[8] is an open-source, cross-platform software for video recording and live streaming. OBS was used to record the computer screen and sound of both the pilot study tests and the final study tests conducted. In addition, the test sessions were recorded to keep a registry and, in case of necessity, to allow future examinations.

Python software libraries: Matplotlib, NumPy, Pandas & SciPy Signal

Matplotlib[9], NumPy[10], Pandas[11] and SciPy [12] Signal are software libraries for the Python programming language used in data processing and analysis. Matplotlib was used to visualise the data results through graphs; NumPy was used for its implementation of high-level mathematical functions; SciPy Signal was used to process the biological signals data, namely, noise reduction through signal filtering, and signal peak detection; Pandas was used to organise, structure and manipulate all the data.

Jamovi

The Jamovi[13] software is an open statistical platform that provides a range of statistical analyses. Jamovi was utilised to conduct the repeated measures analysis of variance (ANOVA) of the physiological data and, additionally, the analysis of the psychological data obtained from the two questionnaires used in this research, the Godspeed Questionnaire Series and the Virtual Reality Presence Questionnaire. (Gallucci, 2019; R Core Team, 2021; The jamovi project, 2022)

This chapter presented and described the phases, approach, hardware, and software tools employed in this research. The following chapter introduces the conceptualisation, design and development of the interactive virtual reality simulation as the proposed solution to answer the research questions.

[8] https://obsproject.com/
[9] https://matplotlib.org/
[10] https://numpy.org/
[11] https://pandas.pydata.org/
[12] https://scipy.org/
[13] https://www.jamovi.org/

Chapter 4

Interactive Virtual Reality Simulation

The fourth chapter is divided into three sections that cover the conceptualisation, design and development of the interactive Virtual Reality (VR) simulation. The first section is dedicated to presenting the conceptualisation and design of the interactive simulation, namely the scenario, the environment, the collaborative assembly task, and the interlocking building blocks. The following section covers the development stages of the interactive VR simulation using the research tools described in section 3.2.

4.1 Conceptualisation and Design of the Interactive VR Simulation

4.1.1 Learning from real-life collaborative scenarios observation

A collaborative assembly task presumes collaboration between the participating agents, working together to achieve a common goal and assembly, i.e. putting together parts of a whole. In preparation for the conceptualisation of the interactive simulation scenario, a review of Human-Robot Collaboration (HRC) real-world applications was made, with close attention to the collaborative robot used, the type of collaboration and task-specific details, to attempt to identify possible solutions to implement as a collaborative assembly task. A summary of the review can be seen in Appendix A.

The review of possible scenarios determined that an assembly application fits this study the most because of its wide variety of potential implementations and due to being one of the most common industrial applications. Discussion regarding the assembly component also concluded that the assembly task should be simple enough so that it did not require special instructions and background knowledge or contextualisation regarding the assemblage as required in the assembly of mechanical or electric parts.

The assembly task needed to be simple not only to be feasible to design, develop and integrate with the virtual robots but also to be robust enough to allow for further complexity and scalability for this research to be future-proof. After brainstorming possible solutions and pondering on their strengths and caveats, the research team decided to use Lego as the basis for the scenario; i.e. the scene would feature Lego-like interlocking building blocks to perform the collaborative assembly task.

Lego-like interlocking building blocks are used to perform the collaborative assembly task due to the advantages the latter present for this research study. Interlocking building blocks are universal, and their affordances are clear and easily understood by the users. Task scalability is also an option due to the large number of different types of blocks that exist, allowing for an immense number of building block combinations and multiple variations of task complexity. Another benefit to using interlocking building blocks in the interactive VR simulation is the non-necessity to overwhelm the users with having to dwell and learn about specific tools or assembly tasks. Furthermore, the learning curve of the building blocks is almost non-existential when compared to the industrial assembly of electronic components or intricate mechanical components, thus assisting in time management of the planned test session saving time by not having to teach users about task-specific details. Furthermore, the building blocks can also serve as a simplified representation of various industrial assembly tasks. The motives presented above help improve the overall assembly task, namely the user performance, by minimising task errors and the user experience using objects with well-known and straightforward affordances.

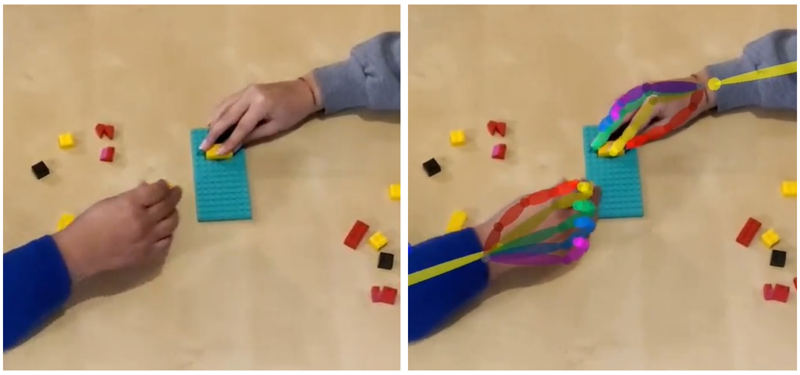


Figure 4.1: Two agents using interlocking building blocks collaboratively in an assembly task.

To build an understanding of the use of interlocking building blocks for the collaborative assembly task, two persons were asked to assemble a building block structure collaboratively, as seen in figure 4.1. Both participating agents were told to assemble a building block structure collaboratively, shown the final assembled structure, and each was given half of the necessary building blocks to assemble the structure. The human-human building block assembly collaborative task was video recorded and later analysed by team members involved in the research, one expert in Human-Computer Interaction (HCI), another expert in VR and the other expert in Human Movement Science and Collaborative Robotics. In addition to naked-eye video observation, a more detailed analysis of the gestures was done with the software OpenPose[1], to detect the agents' hand key points. The analysis of the video recordings showed that one of the agents tends to wait for the other to finish inserting the current building block before inserting the next building block due to not being able to insert it simultaneously and requiring the first levels of the structure to be assembled before being able to proceed to the next levels. While the agents took turns inserting the building blocks, it was possible to observe the communication of intention occurring mostly through gestures and eye gaze.

After having determined a starting point for the scenario and the task, sketches helped to conceptualise the scenario and the structure to be assembled, depicted in figure 4.2. A storyboard of a collaborative assembly task was also created to represent the task step by step, depicted in figure 4.3. These allowed for exploring possibilities using low-fidelity prototypes.

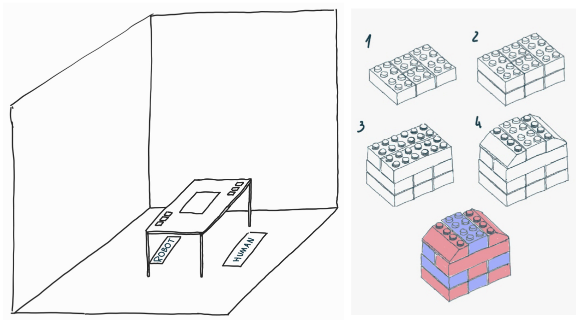


Figure 4.2: Concept sketches of the scenario on the left and of a building block structure on the right.

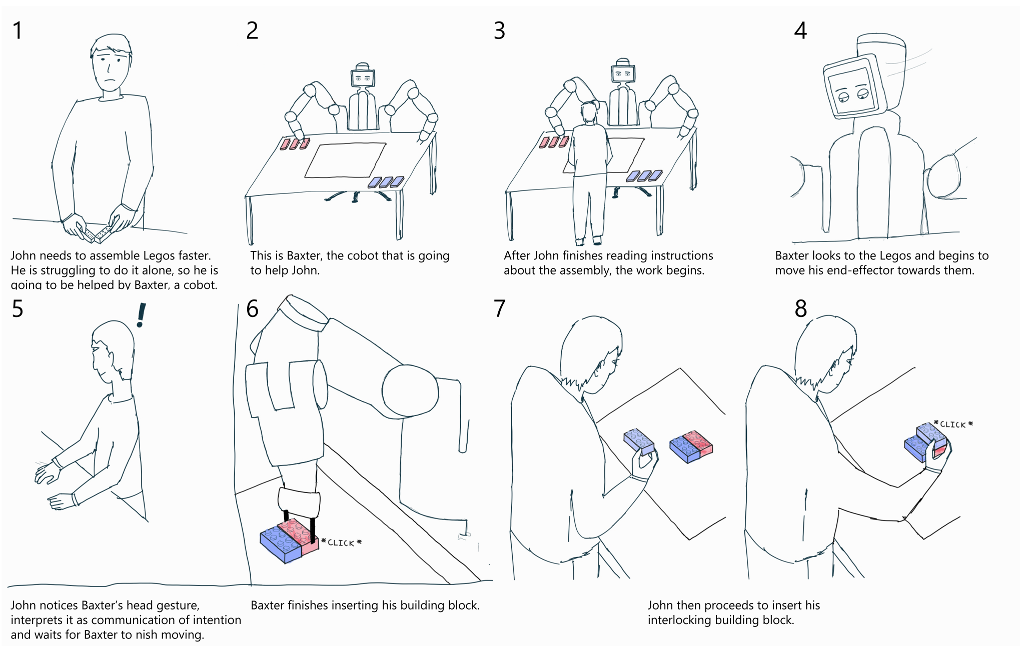


Figure 4.3: Storyboard of a collaborative assembly task.

[1]https://github.com/CMU-Perceptual-Computing-Lab/openpose

4.1.2 Scenario and Task Design

After determining the central scenario and task concepts for the interactive VR simulation, decisions on the types of collaborative robots and types of collaboration also had to be made, as explained in the following sections.

Types of Collaborative Robots

Considering the types of collaborative robots, the task is performed with three distinct robots: Kuka LBR iiwa, as shown in figure 4.4, is a one-armed collaborative robot; Rethink Robotics Baxter, as shown in figure 4.5, is a two-armed collaborative robot with an animated face and Rethink Robotics Sawyer, as shown in figure 4.6, is a one-armed collaborative robot with an animated face. Despite the robots' differences, they all have been designed and projected for collaborative industrial applications, which makes them suitable candidates to perform the previously introduced collaborative assembly task.

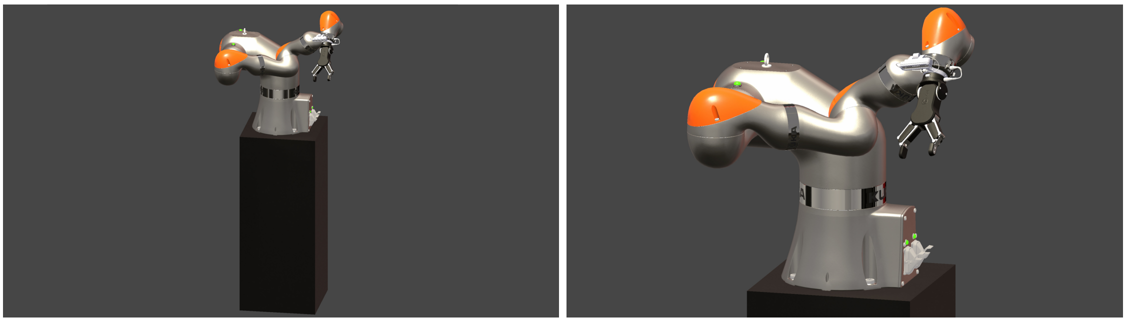


Figure 4.4: Kuka LBR iiwa robot as seen in the interactive virtual reality simulation

Figure 4.5: Baxter robot as seen in the interactive virtual reality simulation

The three robots used in this research purposely have different levels of anthropomorphism. The less anthropomorphic robot is Kuka LBR iiwa due to being a single-armed industrial-looking robot. The most anthropomorphic robot is Baxter due to its humanoid body characteristics such as its torso, two arms and monitor display that behaves like a head with animated facial expressions. On the anthropomorphic scale, Sawyer is in the middle of the two robots previously presented, Kuka LBR iiwa and Baxter, because he is a single-armed robot, like Kuka

Figure 4.6: Sawyer robot as seen in the interactive virtual reality simulation

LBR iiwa, but has a monitor display that serves as a head with animated facial expressions, like Baxter. Both Baxter's and Sawyer's monitor display act as a head with eyes and expressions that can be either turned on, when the robots' head follows the end-effector position, displaying the robots' animated eyes, or turned off, where the robots' head does not move and shows an empty black screen. All three robots communicate motion intention with sound by producing two high-pitched beeps before moving. In the case of Baxter and Sawyer, when their head is active, the robots' head and eyes will move to follow their end-effector position, which fits in the head gestures category of the intention of communication, as previously shown in subsection 2.2.1. The movement of the robots' arms is similar across the three robots because all use the same inverse kinematics solution with equal parametrisation.

Sequential and Simultaneous Collaboration

Two distinct collaborative approaches are employed in the task performed, sequential collaboration and simultaneous collaboration. In the sequential approach, as detailed on the activity diagram 4.7, the human is responsible for beginning the task, i.e. picking and assembling the first interlocking building block. The robots only move to pick and assemble their building block after the human inserts his building block, leading to a sequential assembly of the structure where each agent takes turns performing their own part of the task. Although the human participant is free to move and has the capability to perform the task at his own pace, he is still dependent on his peer, the robot, to proceed to the following structure layer, just as the robot is dependent on the human to progress in the assembly task. In the simultaneous approach, as documented on the activity diagram displayed in figure 4.8, both parties participating in the task can start picking and assembling building blocks when the test begins. The participating agents should perform their actions simultaneously since each can work on his part of the structure to assemble. However, similarly to the sequential collaborative approach, there is still the necessity to wait for each participating agent to finish inserting their building block in the current structure layer being assembled to advance to the next structure layer.

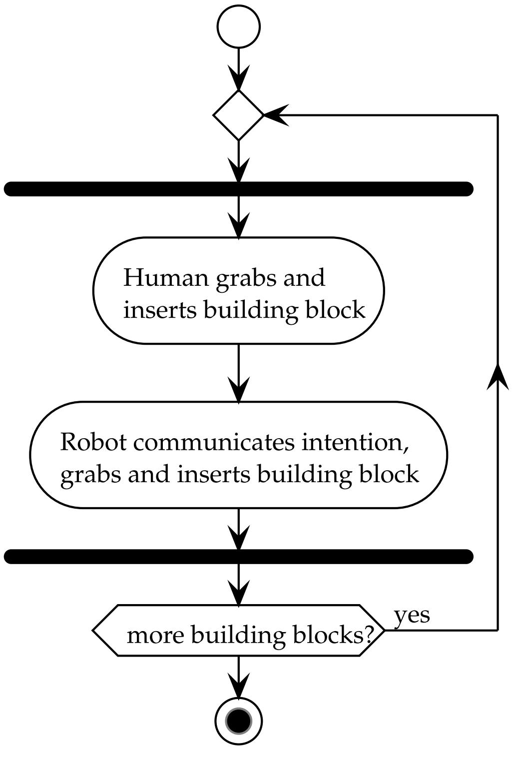


Figure 4.7: Sequential collaborative approach

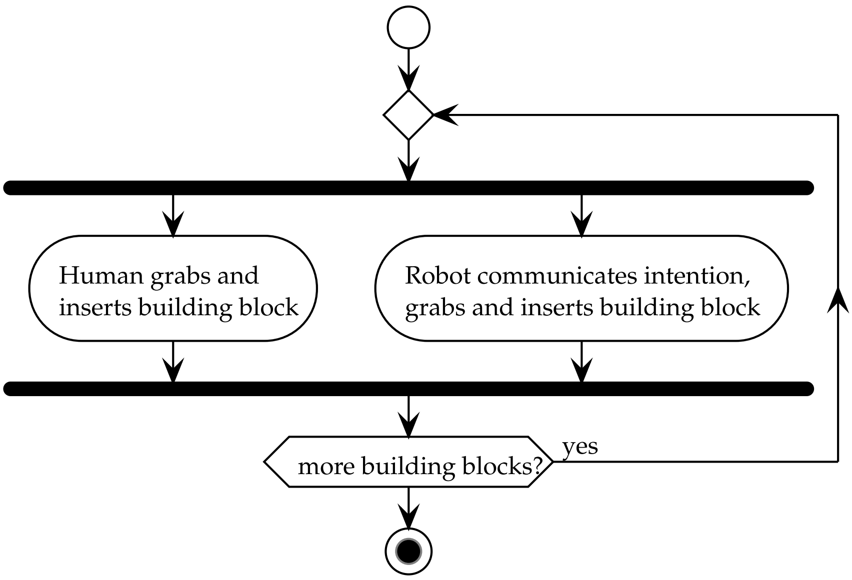
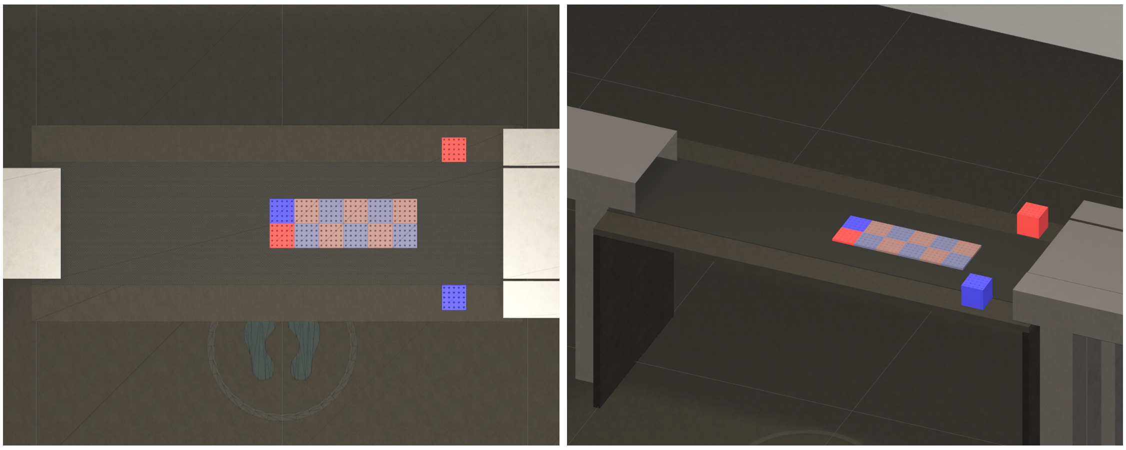


Figure 4.8: Simultaneous collaborative approach

Collaborative Assembly Task

The collaborative assembly task occurs in the virtual workplace shown in figure 4.9. The main idea for the VR environment was that it should be kept simple and free of distractions, only including the essential objects for the agents to perform the task. This design option primarily prevents the human participants' gaze from getting lost in superfluous artefacts and interfering with the data obtained through eye-tracking. Besides the room walls and door, the ceiling light and a screen display placed on the front wall to indicate the name of the current collaborative robot at the beginning of the task, the virtual environment comprises the following objects:

- **Robot**, Kuka LBR iiwa, Rethink Robotics Baxter and Sawyer
- **Workspace**, the grid to insert the interlocking building blocks, the table, the conveyor belt, the building blocks spawners and assembled structure receiver, and a display with a stopwatch timer.
- **Interlocking building blocks**, one for each of the two participating agents, the human and the robot



(a) Workspace top-view (b) Workspace perspective-view

Figure 4.9: The workspace in the VR simulation where task participating agents perform the collaborative assembly task.

The table where the shared workspace is located has a length of 200 cm by a width of 80 cm. The table's width is large enough to guarantee that the robot is kept at a minimum safety distance from the human but also sufficiently narrow to allow both task participants to reach the shared workspace where the collaboration takes place. The interlocking building block spawners, the final assembled structure receiver, and the conveyor belt were not initially present in the scenario. Still, as mentioned in subsection 4.1.2, to minimise task error and improve the scenario's realism, the spawners were introduced to deliver the interlocking building blocks one by one to the agents and to deliver the building base where both agents insert the building blocks. The receiver was introduced as an end-point to the conveyor belt so that the final assembled structure could be removed from the table to give place for the new building base. Before the collaborative assembly task test begins, instructions are given verbally to the users and through demonstration by allowing participants to practice the assembly task inside a specific VR training scene that serves the purpose of an interactive tutorial. This scene is introduced before the tests with data collection, so as not to distract the participants during the collaborative assembly task. The participants perform several trials in the practice stage and proceed to the collaborative assembly task with a given robot only after pressing a green button placed by their side. Figure 4.10 displays an overall view of the practice scene.

Figure 4.10: Scene where participants can practice for the assembly task.

The task consists of assembling an interlocking building block structure in conjunction with three separate collaborative robots and two distinct collaborative approaches. The assembling with building blocks is accomplished by interlocking the building blocks on the building base grid, side by side, in an alternated pattern. There are interlocking building blocks of two colours, blue and red, each associated with a unique task participant. The red building blocks are manipulated by the robots, and the blue building blocks are manipulated by the humans. The structure is assembled layer by layer, and both parties must insert the same building blocks per the structure's layer.

To achieve the joint goal of the task, which is to assemble the ordered structure of building blocks, both task parties, human and robot, must finish their subtasks accordingly. To inform the users after a successful insertion and interlock of a block, they receive haptic and auditory feedback, a light vibration paired with the characteristic click sound that occurs when Lego-like or interlocking building blocks snap together. In each assembled structure layer, the insertion order of the building block is alternated. The collaborative assembly task goal is to build an ordered structure of interlocking building blocks.

To increase the complexity and the dynamics of the collaborative task, the insertion of the building blocks on the grid is done alternately between the agents as depicted in figure 4.11, the position in which the agents need to insert their building block changes throughout the layers of the structure.

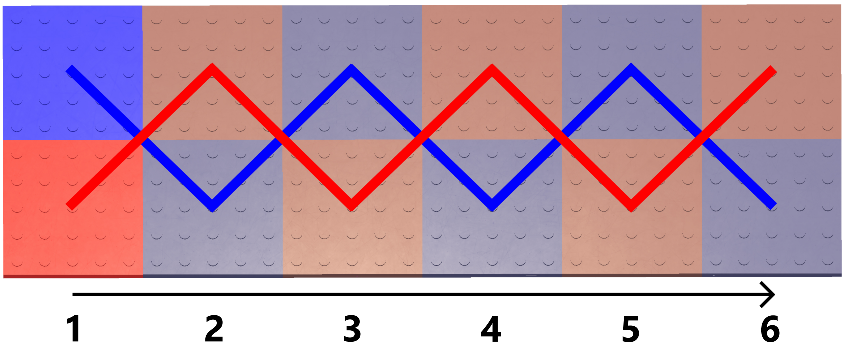


Figure 4.11: Task alternated assembly pattern, the numbers indicate the layer number and the arrow the assembly order.

The task participants cannot insert their building block in any slot other than the one highlighted in their respective colour. The reasoning behind this decision is that if humans were allowed to insert their building block in whatever insertion slot they wanted, participants would perform the task unorderly and differently from each other, thus influencing the results to be acquired. Also, the robots would need to adapt to the different participants' building block placement, which would need to be implemented in the task logic. To further improve task ease and minimise unnecessary errors, instead of providing the agents with all the necessary building blocks to assemble the structure from the beginning of the assembly task, these are instead delivered one by one after each agent finishes inserting his building block, the next one is provided; therefore the agents do not need to think about where each building block goes. The figure in 4.12 displays a participant inserting an interlocking building block in the interactive VR simulation.

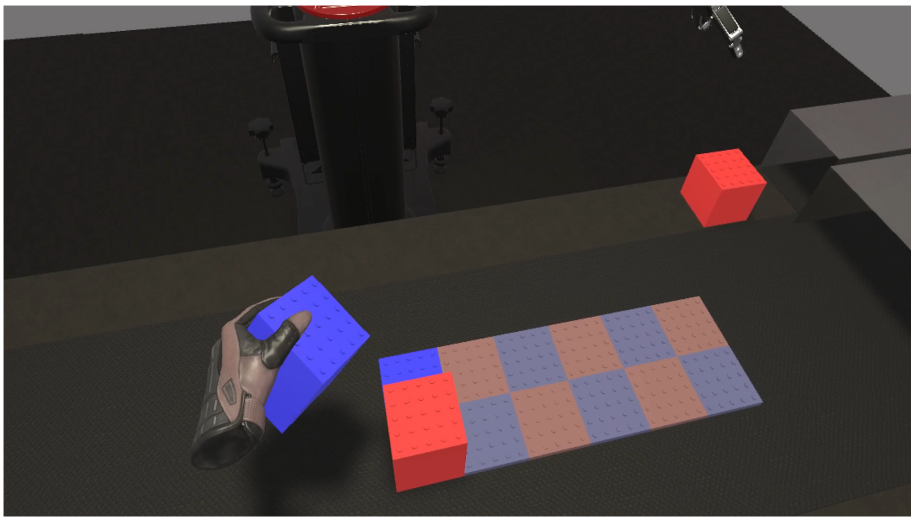


Figure 4.12: A participant inserting an interlocking building block.

4.2 Development of the Interactive VR Simulation

The development phase of the interactive VR simulation began in the middle of February and ended in the middle of May, having lasted approximately three months. The interactive VR simulation was developed using the Unity Real-Time Development Platform as presented in section 3.2, using an iterative and incremental approach. Throughout the development phase, the project was consistently discussed and evaluated on a weekly basis by members of the Neurocobots team, one whose main area of expertise is HCI, another whose main area of expertise is VR, and another whose main area of expertise is Collaborative Robotics. In total, 129 versions[23] of the Unity project were built and tested throughout the development period. The following sections introduce and explain the main development stages and the main changes between stages.

4.2.1 Analysis of Previous Work

The first stage was dedicated to analysing and understanding the previous Unity project developed by Branco and Bermúdez I Badia [4], in Unity version 2019.2.8f1. This project included an implementation of the robots Baxter and Kuka LBR iiwa, with an inverse kinematics solver that allowed control of the robot's movement but lacked functionalities, such as collision avoidance and realistic movement profiles. The three-dimensional model of the Kuka LBR iiwa robot was simple, its colours were different from the original robot, and it was missing one joint and a proper end-effector, i.e. a robotic device grip tool, but it sufficed to test the crucial components necessary to progress the development. Furthermore, the initial project lacked the logic required to perform the interlocking building block collaborative assembly task with the robots, as well as the different types of collaboration defined for the task, sequential and simultaneous collaboration, which are directly tied to how the assembly task is performed. The environment and the workspace were also not implemented, i.e. the table, timer, interlocking building blocks and spawners and receiver for the grid and the building blocks. The following subsections introduce the main development phases and the primary advances at each phase.

4.2.2 Task Logic, Data Logging, Unity LTS

The first version of the project was focused on implementing the basic elements required for the task to be performed, such as a workspace that consisted of a table, a grid to insert the interlocking building blocks and an interlocking building block. The initial versions set the foundation for the development of the task logic, which included task progression, i.e. advancing experimental trials and blocks, and task validity to ensure the task is performed accordingly to the task

[2](not publicly accessible) `https://github.com/eduardosaraujo/EHRCtVRS`
[3]Key:F85Hd2%qmNPH `https://github.com/eduardosaraujo/EHRCtVRS-BUILD/releases`
[4](not publicly accessible) `https://github.com/DiogoAABranco/DarkEnv`

requirements. Lastly, the implementation of different types of collaboration to perform the collaborative task took place. In sequential collaboration, the robot only moves after the user inserts his building block. In simultaneous collaboration, the robot moves independently from the participant. Attention is given to allow parameterisation of the task. The parameters required for the assembly task, such as task settings, i.e. the number of building blocks needed, test trials and blocks, type of robot, type of collaboration and robot movement settings, can be modified and loaded through an external configuration file in Comma-separated values (CSV) format.

After having a solid implementation of the task logic, other features were added, such as animations, interlocking building block snapping, and objects like the spawners for the building blocks, the spawner and receiver for the building block base grid, a conveyor belt, ceiling lights and lastly a floor signal indicating where the user should stand during the assembly task. These changes contribute to minimising task error, immersion, and overall realism of the simulation. Another critical feature implemented was physiological and task data logging. Recording and storing physiological data from various channels, such as the eyes and the pulse, as well as some pertinent task data, such as the current trial, the current type of robot, the current type of collaboration, and timestamp markers to indicate the occurrence of stimuli and facilitate data analysis. Lastly, the initial development phase is also important due to the migration of the whole project from Unity version 2019.2.8f1 to Unity LTS version 2020.3.30f1, thus allowing for the project to be more future-proof with higher compatibility and support. This last change also enabled the project to use Universal Render Pipeline (URP) ready materials, as explained in the next section.

4.2.3 Inverse Kinematics, Universal Render Pipeline Materials, Kuka LBR iiwa and Rethink Robotics Sawyer Integration

The biggest improvement in the third development phase, and most likely in this entire project, was the introduction of the new inverse kinematics solver, BioIK, developed by Sebastian Starke at the University of Hamburg as a part of his Master of Science thesis (Starke et al., 2017a, 2016, 2017b). The usage of BioIK enabled more complex and realistic robot movements, with collision detection, joint angle restrictions and other valuable features such as different weight inputs for different goals or collision elements. The addition of URP Unity materials increased the realism and fidelity of the overall visual dimension of the VR simulation, leading to an increase in the sense of presence and immersion. The new three-dimensional model for the Kuka LBR iiwa robot was also a notable improvement for the project since the old three-dimensional model of the Kuka LBR iiwa robot had a low fidelity and lacked an end-effector, i.e. a gripper robotic device that serves as a hand for the robot. Alongside the new Kuka LBR iiwa robot model, the addition of the new robot, Sawyer, from Rethink Robotics, was also a good complement to the project since it introduced a new type of robot with a monitor screen, like Baxter, but single-armed, like Kuka LBR iiwa. The model used for the Rethink Robotics Baxter was provided in the first version of the project, explained in section 4.2.1. The Kuka LBR iiwa model published by Bornemann and the Rethink Robotics Sawyer model published by Malawey were downloaded from the GrabCAD Community website. Besides the above-mentioned changes, other notable mentions for the third development stage are the integration of the Godspeed Questionnaire Series (GQS) in the interactive VR simulation and of the task practice scene to allow participants to practice for the task without the presence of the robots.

Answering the questionnaire in VR allowed for the procedure to be more easily and pleasantly conducted and experienced since it became unnecessary to repeatedly remove and equip the Head Mounted Display (HMD) to answer the questionnaires after going through each different test condition. This further contributed significantly to reducing the overall duration of a test session due to the time spent equipping, adjusting and removing the HMD, as well as the required eye-tracker re-calibration every time the HMD was equipped or adjusted. Adding a task practice scene allowed the participants to rehearse the task without a robot, improving their familiarity with the movements required to perform the task, the surrounding space and the necessary objects before committing to the collaborative assembly task with the robots.

More changes were made after a pilot study with seven participants, which were of critical importance:

- **Controlling the execution of the eye-tracker sensor calibration**, since the users do not need to remove the HMD during the simulation, there was no need to execute the eye calibration protocol for each test condition. The eye-tracking sensor was only required to be calibrated at the beginning of the test since the participants did not remove or adjust the HMD throughout the test.

- **Participants' handedness**, initially, the simulation only included a building block spawner on the right side of the workspace, but later an option was added to change the building blocks spawner position, left or right side, thus enabling both left-handed and right-handed participants to perform the task with ease.

- **Logic and variable resetting** after each test condition, thus avoiding restarting the simulation for each independent condition. This caused ruptures in the participants' immersion and was also time-consuming, contributing to reducing the duration of the test session.

- **Detection of gaze focus** directly within Unity with the SRanipal Runtime and Software Development Kit (SDK), and subsequently recording and storing this data, allowing to easily and quickly infer where the participants were looking on the task scene.

- **Graphical glitch** when transitioning from the collaborative assembly task scene to the GQS scene, causing discomfort and a bad experience for the test participants.

- **Screen with a timer** was added to the environment to incentivise the participants to perform the collaborative assembly task as quickly as possible.

The work presented in this chapter contributed to developing a higher fidelity interactive VR simulation, thus providing a solid solution for evaluating the HRC through the latter. The next chapter presents the evaluation measures and procedures, the study participants' descriptions, the data collection and analysis, and its results.

Chapter 5

Evaluation of Human-Robot Collaboration

This chapter describes the evaluation of human-robot collaboration, with a focus on human performance, through the previously developed interactive virtual reality simulation. The first two sections present the evaluation measures, the procedures and the test conditions. The next section describes the study participants, followed by the data collection and analysis section and finally, the results.

5.1 Evaluation Measures and Test Conditions

The main research question of this study is how human performance is impacted when executing a collaborative assembly task in the interactive VR simulation with robots of different anthropomorphic designs. This includes three robot models, Baxter and Sawyer having their eye and face animations on and off, and Kuka LBR iiwa. Two distinct collaboration approaches, sequential and simultaneous, are also evaluated. Secondarily, the uncanny valley effect on emotional response is studied to understand its possible impact on human performance. The primary data types collected and analysed to answer the research questions are physiological and behavioural data. Following this approach allows for this research to offer a multi-faceted insight into human factors related to HRC.

Physiological Measures Collected

To measure and analyse the effect of the interaction on human performance, participants' physiological data was acquired, namely from the eyes and the pulse. From the eye data, it was possible to obtain, amongst others: eye data validity (a proxy for blink rate), pupil dilation diameter in millimetres, and gaze fixation location, i.e. the location at which the user is looking during the simulation. Pupil dilation and blink rate allow for inferring whether the participants are under a lower or higher cognitive load and awareness. Additionally, it is possible to identify what objects participants spend more time looking at while performing the collaborative assembly task. From the pulse data, two different measures are obtained and analysed, the pulse rate, in beats per minute (Beats per Minute (BPM)), and the pulse inter-beat interval (Inter Beat Interval (IBI)), in milliseconds. These pulse measures make it possible to infer whether the participants were more or less stressed in different test conditions.

Psychological Measures Collected

Psychological measures were collected through two standard questionnaires, as described next.

The Godspeed Questionnaire (GQS) (Bartneck et al., 2009b), which comprises five question groups corresponding to the dimensions that reflect users' perception of the robot, such as anthropomorphism, animacy, likeability, intelligence and safety. The questionnaire implements semantic differential scales, and it has been translated into seventeen languages, a semantic differential scale uses dichotomous pairs of adjectives to allow the participants to answer each question. In this study, the questionnaire was employed in two languages, English and European Portuguese. After each test condition was presented, the GQS was answered inside the interactive VR simulation.

The VR Presence Questionnaire (PQ) (Witmer et al., 2005) was used as a tool to measure the degree of the participants' sense of presence in the interactive VR simulation. This questionnaire measures different domains of the VR experience, allowing for an understanding of what factors contributed to the overall sense of presence. This questionnaire is comprised of 24 questions answered on a seven-point Likert scale to assess VR experience factors such as realism, possibility to act and to examine, sounds and haptics. The PQ was answered on a digital form using a computer immediately after the participants finished all the test conditions and unequipped the HMD and the pulse sensor.

Test Conditions

The experimental procedure comprised five different test conditions, as summarised in table 5.1. All participants were subjected to all five test conditions, and the order in which the test conditions were presented to the participants was randomised to not interfere with the results. By order of importance, the independent variables manipulated in this research are: (i) the type of robot, (ii) the type of collaboration and (iii) the face/eyes display state, which only applies to the robots Baxter and Sawyer.

Table 5.1 also presents the control variables. The robot's distance represents the distance at which the robot is placed from the participant, which, in this study, is equal for all the robots. The maximum velocity of the robot and its target correspond to the maximum velocity parameter set in Unity's BioIK. Although the value is set to $2\,\mathrm{m\,s^{-1}}$ due to BioIK's realistic movement profile being enabled, the robots never reached this velocity. To add randomness to the movement delays and reduce the subject tendency to act automatically, an option to enable/disable the delay of robot movements as a variation of more or less 0.5 s. The timer limit represents the time it takes for the timer screen to change from the standard black colour to the colour red, thus indicating to the participants that they have surpassed the time limit to perform the task. Even though the timer screen colour changes, the task does not end until the participating agents insert the last interlocking building blocks, and the stopwatch timer keeps counting the time passed. The number of grids represents the number of final assembled structures the task participants need to finish to complete a test condition. The number of grid columns and the number of grid rows represents the number of columns and rows of the structure, respectively. Each structure column corresponds to a test trial, thus having a total of 32 trials per type of robot and 16 trials per type of collaboration per robot.

	Test #1	Test #2	Test #3	Test #4	Test #5
Type of robot	Kuka LBR iiwa	Baxter		Sawyer	
Type of collaboration	50% sequential, 50% simultaneous (grids) per test				
Robot monitor state	-	Off	On	Off	On
Distance of robot	1 m				
Maximum velocity of robot and robot's target	$2\,\mathrm{m\,s^{-1}}$				
Enable/disable robot's movement delay	$1.0 \pm 0.5\,\mathrm{s}$				
Timer limit	50.0s/grid				
Number of grids	4	2			
Number of grid columns	8				
Number of grid rows	2				

Table 5.1: The five test conditions with the independent variables highlighted in bold and the control variables not highlighted.

5.2 Evaluation Procedures

The study received ethical approval from the Ethics and Deontology Committee for Research (CEDI) of the Faculty of Psychology and Educational Sciences of the University of Coimbra: CEDI/FPCEUC:64/1, 22/06/2022. The two researchers followed a procedure script to conduct the tests, conveying the same exact instructions to all the participants. The full procedure script, translated to English, is presented in appendix B. An activity diagram presenting the main test session phases and their flow can be seen in figure 5.1. The test sessions were always conducted with two researchers present at all times, where one was responsible for handling the interactive VR simulation and the data collection, and the other was in charge of receiving the participants and providing assistance in equipping and removing the necessary hardware.

After the participants arrived at the laboratory, they were asked to read and agree to an informed consent form and answer a sociodemographic questionnaire on a computer, ensuring the data collected were anonymous and confidential. The sociodemographic questionnaire included questions about the participant's age, gender, height, education level, professional occupation, dominant hand and previous experiences with interlocking building blocks, virtual reality and robots. All participants were asked whether they were prone to feeling nauseous, dizzy or sick and were asked to tell the experimenters in case of feeling bad during the test session. Before the users were equipped with the necessary hardware, a researcher presented a summary of the overall procedure while immersed in the interactive VR simulation, presenting and explaining the following four stages:

- **Eye-tracking calibration**, the first step is to calibrate the eye-tracking sensor and verify if the HMD is fitting correctly for the sensor to have an unobstructed reading of the participant's eyes.
- **Training stage**, the second step is to practice for the assembly task without the presence of robots.
- **Collaborative task**, the third step is to perform the collaborative assembly task with the different test conditions.
- **Questionnaire**, the last step is to answer a questionnaire regarding their perception of the robot through VR at the end of each test condition.

After the summary of the test session, one of the researchers assisted the participants with equipping the HMD and the controllers, as well as attaching the pulse sensor to the fourth digit of the non-dominant hand, while performing the task. The users were instructed on how to adjust the HMD by themselves, so they were able to set it in a proper position where they could see the image focused and clear. After correctly placing the HMD, participants were given the handheld controllers and told to move around and explore the SteamVR Home to get used to VR and its controls. In the meantime, the pulse sensor was attached to the participant's non-dominant hand's fourth digit.

The researcher responsible for handling the data collection verified if the data acquired by the pulse sensor was proper and ensured that the eye-tracking system calibration was performed successfully. The primary reason for verifying the sensor signal's quality was the sensor being attached too tightly or loosely to the participant's finger. After the necessary verifications, the test session began, with the participants practising for the assembly task in the training stage. When the participants were done rehearsing, they were asked to press a button to advance to the next stage.

In this part, the participants performed the collaborative assembly task once per test condition, and after each test condition, they were prompted to answer the GQS in VR. After going through all the test conditions, the participants were assisted in removing the equipped hardware and asked to answer the VR PQ. Upon concluding, the researchers thank the participants for their participation.

Although the researchers abstained from intervening during the actual collaborative assembly task, there were instances where participants required guidance. During the calibration process of the eye-tracking system, a researcher assisted

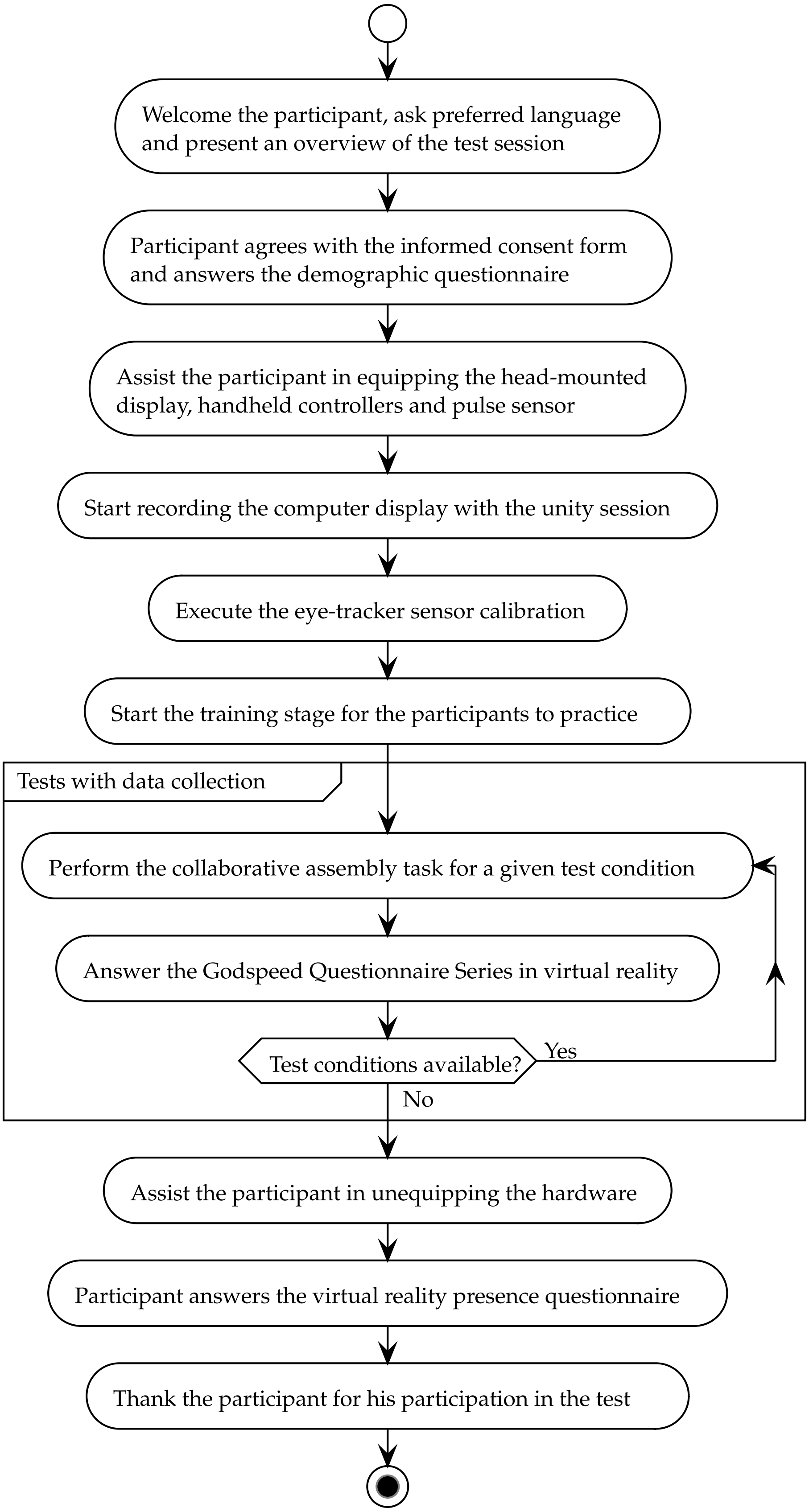


Figure 5.1: Activity diagram presenting the main test session phases and their flow.

the participants in adjusting the HMD and the interpupillary distance (IPD), when necessary, to certify that the pupil sensor had a full view of the participant's eyes and to provide an overall better user experience, a researcher verbally communicated what was happening to the participants.

In the training stage, the researchers intervene freely to assist participants with specific details of the task, such as explaining the controls to grab the cube and insert it in the corresponding slot and how to proceed to the collaborative assembly task after completing the training. While performing the collaborative task with the robots, the researchers tried not to intervene and only did so if required because of the participants reported feeling sick from nausea or dizziness or if the participant interacted with the researchers.

In between test conditions, the participants were asked if they were feeling well and not feeling nauseous or dizzy. Only after obtaining acknowledgement did the researcher advance to the following test condition. Lastly, when answering the GQS inside the interactive VR simulation, the researchers intervene to explain the controls and user interface to the participants and to inform them that if their opinion is neutral, they should leave the scale untouched. The researchers also answered if there was any misunderstanding about the content of the questionnaire.

5.3 Study Participants

A total of 39 participants were recruited to participate in the study. The participants' recruitment was done through email invitations directed mainly at the University of Coimbra student population. The recruitment of participants attempted to be gender-balanced to have similar numbers of female and male participants. Data from three participants, out of the 39, was excluded from the study because the participants were unable to finish the tests due to motion sickness or not feeling well.

From the total of 36 participants whose data were analysed, 20 participants self-identified as males and 16 participants as females. The participants' ages varied from 19 to 26 years old, with the average age being around 22 years old ($M = 22, SD = 2$), 22 years old ($M = 22, SD = 3$) for male participants and 21 years old ($M = 21, SD = 2$) for female participants. The participants' education levels varied where 19 had completed Secondary Education, 15 had Bachelor's or equivalent, and 2 had a Master's or equivalent. A total of 29 participants described their current professional occupation as being a student, one as a public servant, one as a web designer and the remaining five as not having a current professional occupation.

The sociodemographic questionnaire also comprised questions about prior contact and experience with interlocking building blocks (like Lego), robots and VR. Although most participants had contact with all three, knowing these helped characterise the sample. Out of the 36 participants, 34 answered that they had had previous contact or experience with interlocking building blocks, while the remaining 2 answered negatively. Most participants who responded positively described their contact or experience as playing with interlocking building blocks when they were children. When asked about previous contact or experience with robots, 34 participants answered positively, and the remaining 2 answered negatively. Most participants who responded positively described their contact or experience as watching movies or series where robots took part and also through video games. Regarding previous contact or experience with virtual reality, 31 participants answered positively, and the remaining 5 answered negatively. Most participants who responded positively described their contact or experience as having used it in simulators and other previous virtual reality research experiments. Overall, most participants had had some previous contact and experience with all three components presented above, which was optimal for this research.

5.4 Data Collection and Analysis

5.4.1 Data Description

Physiological Data

The physiological data file was recorded and stored in a CSV formatted text file at a frequency of 100 Hz or 10 ms period. Each test condition produced a physiological data file, totalling five physiological data files per participant. Each physiological data file contains the following 63 columns of data variables, as described in table 5.2.

Data category	Data description	Number of data columns
Time		**2**
	Unity time	
	Expressed in milliseconds	1
	Expressed in nanoseconds	1
Eyes		**39**
	Eye-tracker	
	Milliseconds timestamps	1
	Total frames count	1
	Gaze focus	
	Current object name	1
	Gaze focus point position coordinates (X, Y, Z)	3
	Gaze focus point distance (X, Y, Z)	1
	Left and right eye data validity	2
	Left and right eye openness level	2
	Pupil	
	Left and right eye pupil diameter	2
	Left and right pupil position coordinates in sensor area (X, Y)	4
	Eye gaze	
	Left and right eye gaze origin position coordinates (X, Y, Z)	6
	Left and right eye gaze direction vector normalised (X, Y, Z)	6
	Accuracy factor of the gaze ray	1
	Eye facial expression levels for character animation	6
	Convergence	
	Convergence distance validity	1
	Convergence distance from central point of right and left eye	1
	Tracking improvements count	1
Pulse		**3**
	Beats per minute (BPM)	1
	Inter-beat interval (IBI)	1
	Signal	1
Movement		**15**
	Left and right controller position coordinates (X, Y, Z)	6
	Head-mounted display (HMD)	
	HMD position coordinates (X, Y, Z)	3
	HMD rotation degrees (X, Y, Z)	3
	Robot end-effector position coordinates (X, Y, Z)	3
Task		**4**
	Current block	1
	Current trial	1
	Details	1
	Current type of collaboration	1

Table 5.2: Description and column count of the physiological and task-specific variables.

From all the physiological data acquired during the tests with participants, the data that is crucial to this study is, from the eyes data, the pupil dilation, the blink rate and the gaze fixation; from the pulse data, the pulse BPM and the pulse IBI.

Psychological Data

The psychological data file is recorded and stored in a CSV formatted text file, once, at the end of each test condition in the interactive VR simulation after the participants answer the questionnaire, totalling five psychological data files per participant. Each psychological data file contains the following four columns of data variables, as described in table 5.3.

Data category	Data description	Number of data columns
Question identifier	Question unique identifier number	1
Question answer	The answer given by the participant	1
Question semantics	The two semantics used for the question	2

Table 5.3: Description and column count of the psychological variables.

5.4.2 Data Processing

The data processing pipeline algorithm was done with the Python 3.8 programming language in a Jupyter Notebook, using Microsoft's Visual Studio Code as a code editor. In addition, the Python software libraries Matplotlib, NumPy, Pandas and SciPy Signal were used for data visualisation, high-level mathematical operations, signal filtering and signal peak detection, and data structure and organisation, respectively.

Data Pre-processing

The plots in the following section pertain to the same exemplary participant for data processing demonstration purposes. Firstly, the data is fetched from the files containing the raw data obtained from the Unity simulation. For each participant, ten data files were generated, five for the physiological and five for the psychological data; two files per test condition times the five different conditions. Each of these data files is stored in a separate Pandas DataFrame data structure, which is subsequently stored in Python lists. The DataFrames stored in the lists are then split into more DataFrames that are better categorised and parsed into epochs of interest, namely different types of robots, types of collaboration and also different monitor states, in the case of Baxter and Sawyer. The biological signals data required processing, such as signal filtering, peak detection and calculation of arithmetic averages. The parameters used to process the acquired data were constant across all the participants.

Eye Pupil Dilation Processing

The raw eye pupil diameter data comprised data for each left and right eye separately. Each eye's data was independently filtered with a Butterworth low-pass filter using Scipy Signal. Figure 5.2 shows the difference between the raw signal and the filtered signal using the function:

```
scipy.signal.butter(n, wn, fs, btype='lowpass', analog=False)
```

Where `n` is the order of the filter, the order value used was 2, the `wn` is the critical or cutoff frequency and the value used was 2 Hz. Lastly, the `fs` is the sampling frequency at which the digital signal was acquired, in this research, this value corresponds to 100 Hz. The `scipy.signal.butter()` function returns two arrays `b, a` that correspond to the numerator (`b`) and denominator (`a`) polynomials of the filter, that are then used as parameters in the Scipy Signal function:

```
scipy.signal.lfilter(b, a, x)
```

Where `b` is the numerator array, `a` is the denominator array, and `x` is the array with the signal data, this function returns `y`, which is the output of the digital signal after filtering.

After filtering the left and right eye raw signals with the low-pass filter, a combined signal of both eyes was calculated with an arithmetic average between the two signals.

Finally, a Scipy Signal Find Peaks function was used to detect the local maxima, i.e. the signal peaks that correspond to the instances when the participants blinked or a loss of signal occurred due to the eye-tracker sensor not being able to detect the eyes properly.

```
scipy.signal.find_peaks(x, height, width)
```

Where `x` is the signal data, `height` is the required height of the peaks, in this study, the minimum required peak height is −2 mm, and the maximum required peak height is 3.5 mm. Lastly, the `width` is the required width of the peaks in samples, the value used for the minimum necessary peak width is 5 samples, equalling a minimum of 50 ms duration for the detected peaks. This function returns an array with the indices of the detected peaks that satisfy all the given conditions in `x`.

After detecting the signal peaks indices, a `for` loop was used to iterate through the signal and remove the data points around the detected peaks, a total of 130 samples are removed for each peak. Additionally, 100 samples were deleted from the beginning of the signal and 150 samples from the ending. The final result is represented in figure 5.4.

The final value for the combined pupil diameter is calculated from this data using an arithmetic average between all the remaining data points with the function Numpy Nanmean function `numpy.nanmean(a)`, where `a` is the array of data points from which to infer the average, and finally its output is rounded to 2 decimal places with the function `numpy.round_(a, d)` where `a` is the output value of the

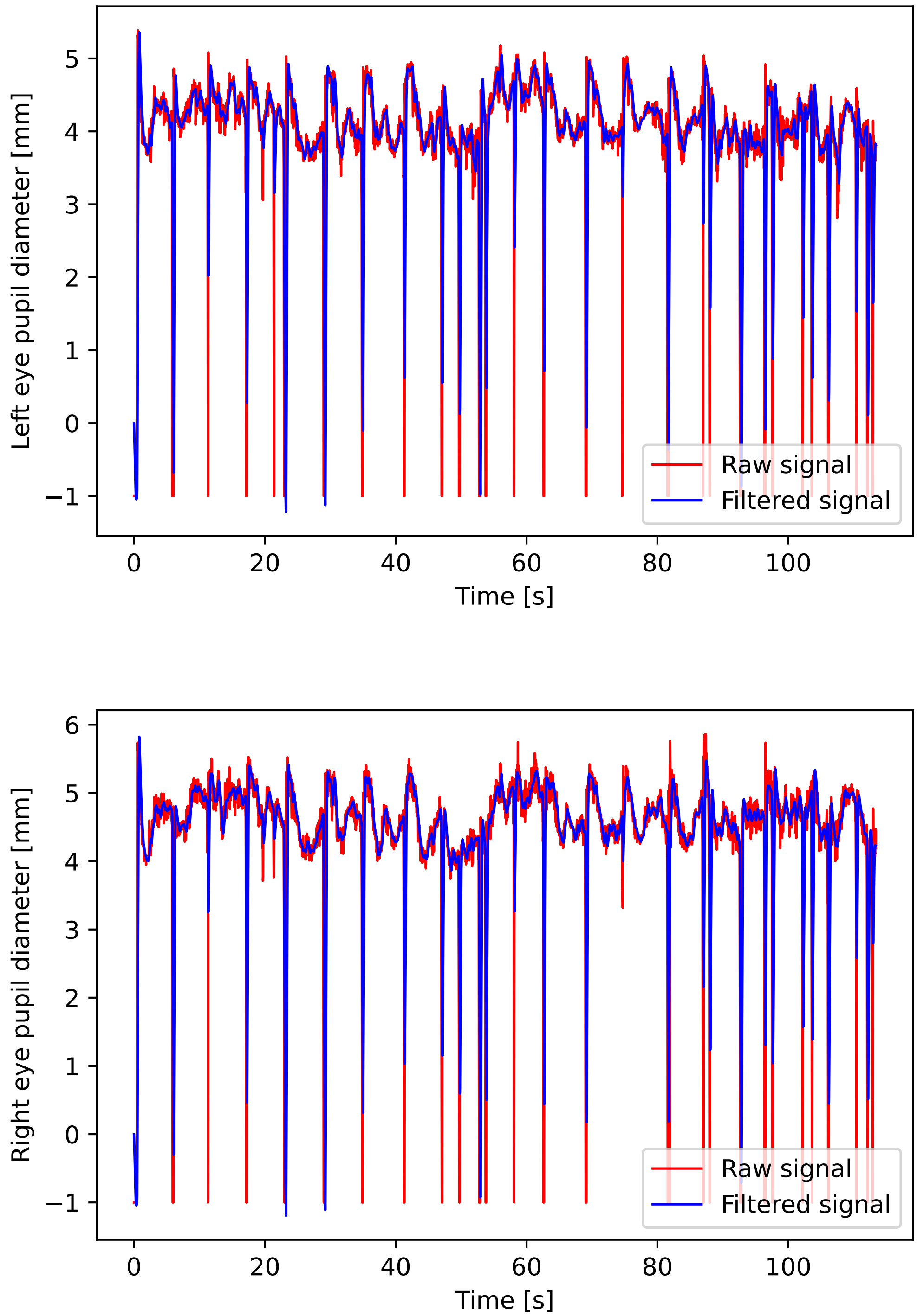


Figure 5.2: Raw and filtered left and right eye pupil diameter

arithmetic average calculation and d is the number of decimals places to preserve, in this case, the parameter is set to the value 2.

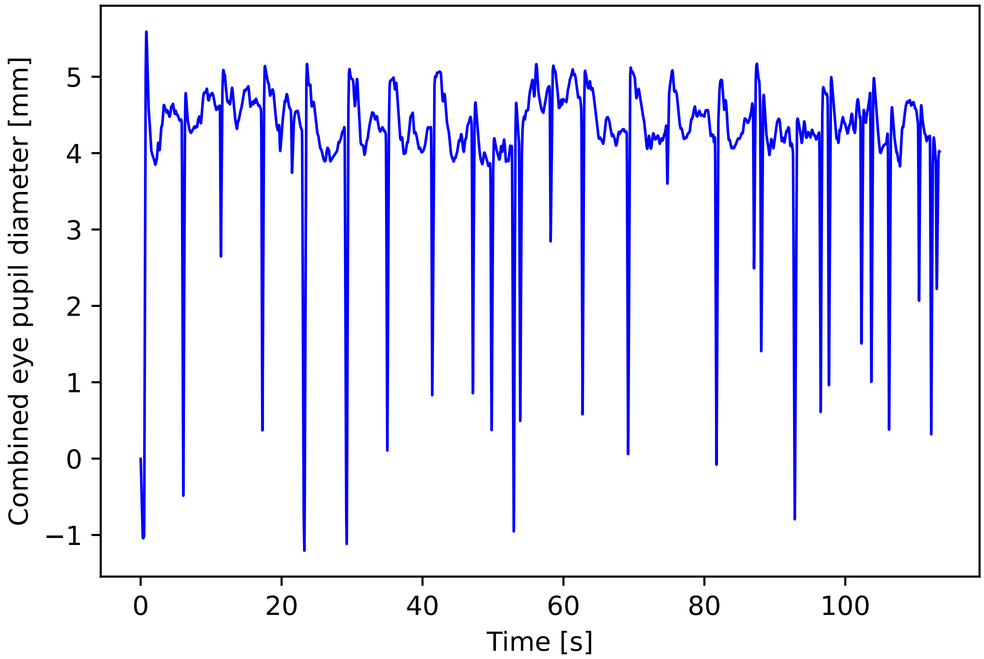


Figure 5.3: Average combined left and right eye pupil diameter

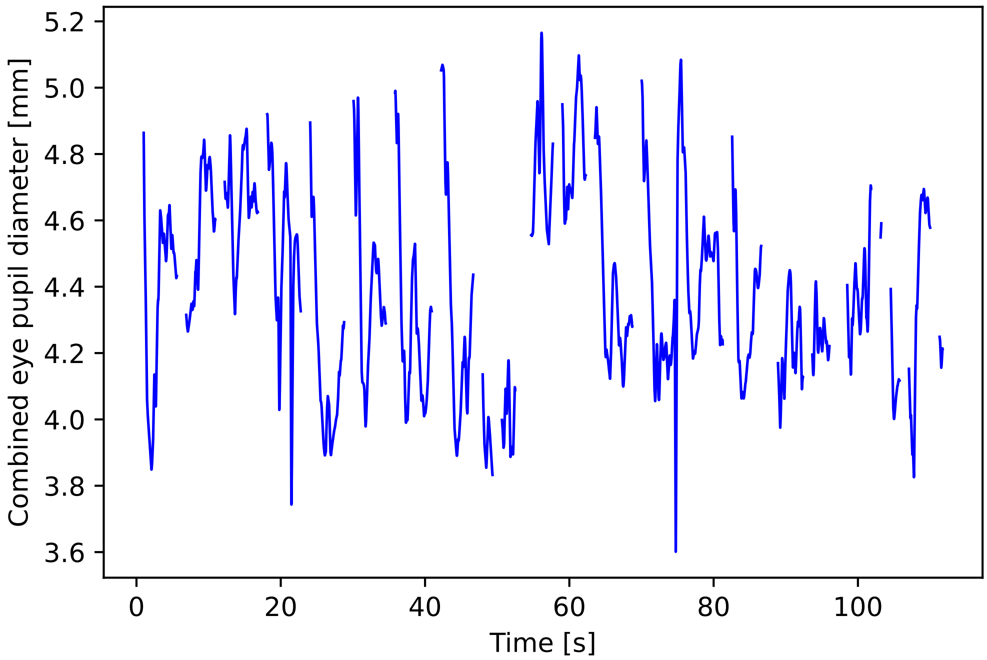


Figure 5.4: Filtered combined eye pupil diameter with peaks removed

Eye Blink Detection & Processing

The eye blinks were inferred from two different eye data variables, the eye openness level and the eye validity, with peak detection. Each of these variables was comprised of raw data pertaining to the left and right eye and was processed separately using the Scipy Signal Find Peaks function, where the following parameters were used:

$scipy.signal.find\_peaks(x,\ height,\ distance)$

Concerning the eye openness level, the parameter x is the signal data, $height$ is the required height for the peaks in the signal, with the minimum height being 0 and the maximum height being 0.35. Finally, the parameter distance is related to the horizontal distance between peak detections. This parameter was set to 10, meaning that, after detecting a peak, the next peak may only be detected after a horizontal distance of 10 samples, 10 samples equates to 100 ms intervals.

Regarding the eye validity, all the parameters have the same values and configuration, except for the $height$, where the minimum required peak height is 0, and the maximum is 10.

As mentioned before, the Find Peaks function returns an array with the indices of the detected peaks that satisfy all the given criteria in the signal x. To obtain the average blink rate, after inferring the peaks for the left and right eye, which represent the total number of blinks for each eye and variable, eye openness and eye validity, an arithmetic average was calculated between each left and right eye to obtain the combined number of blinks for the left and right eye per variable. A second arithmetic average was used to calculate the average of the two variables, eye openness and validity, conveying the average number of total eye blinks in a given session. Finally, the number of average eye blinks was divided by the total duration of the task in seconds (s) and rounded to 4 decimal places, thus calculating the average number of blinks per second. A visual representation of the peaks detected in both eyes across the two different variables, eye openness and eye validity, can be seen in the graphs in figure 5.5, and 5.6, respectively.

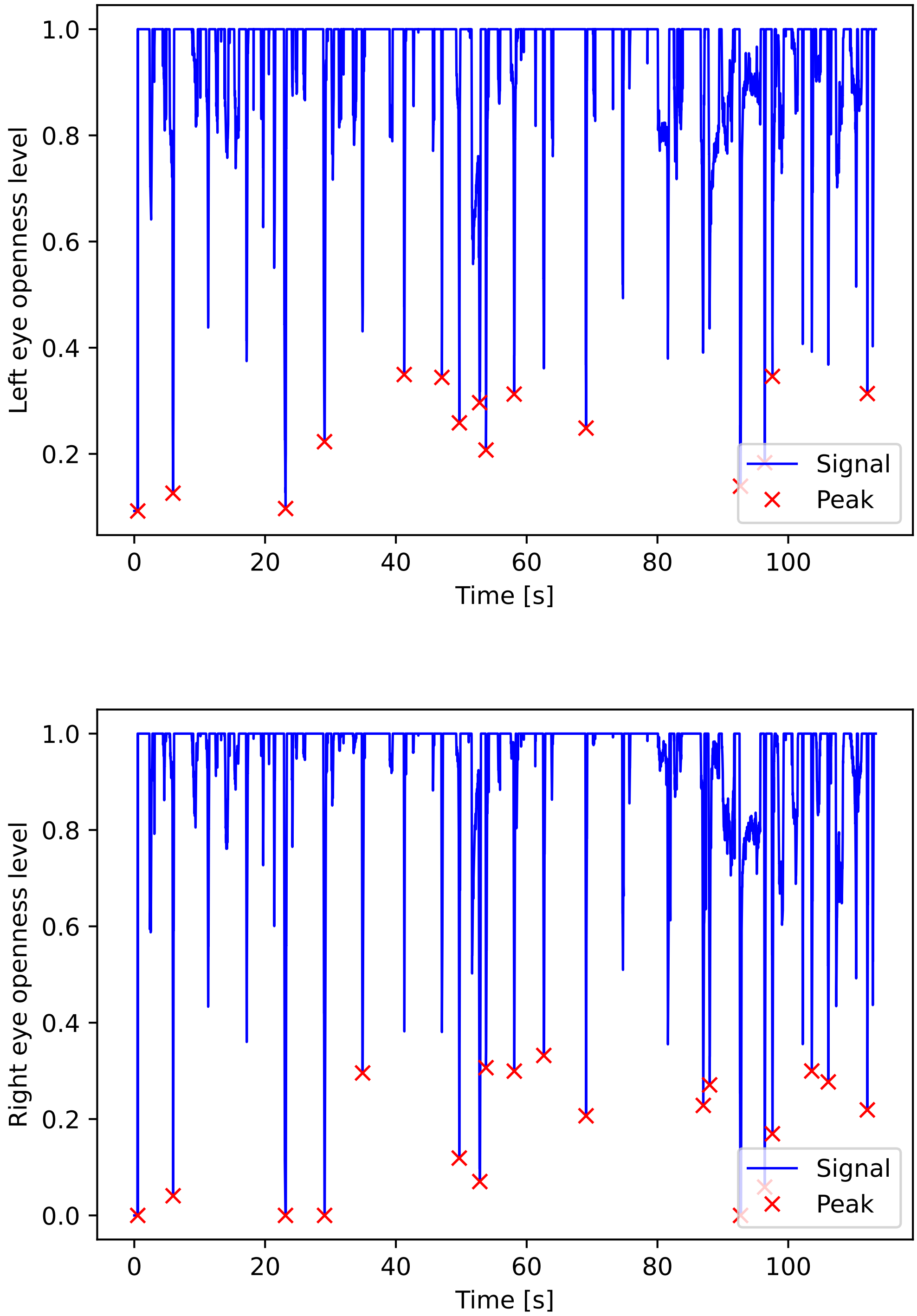


Figure 5.5: Peaks detected in the eye openness level signal for both eyes.

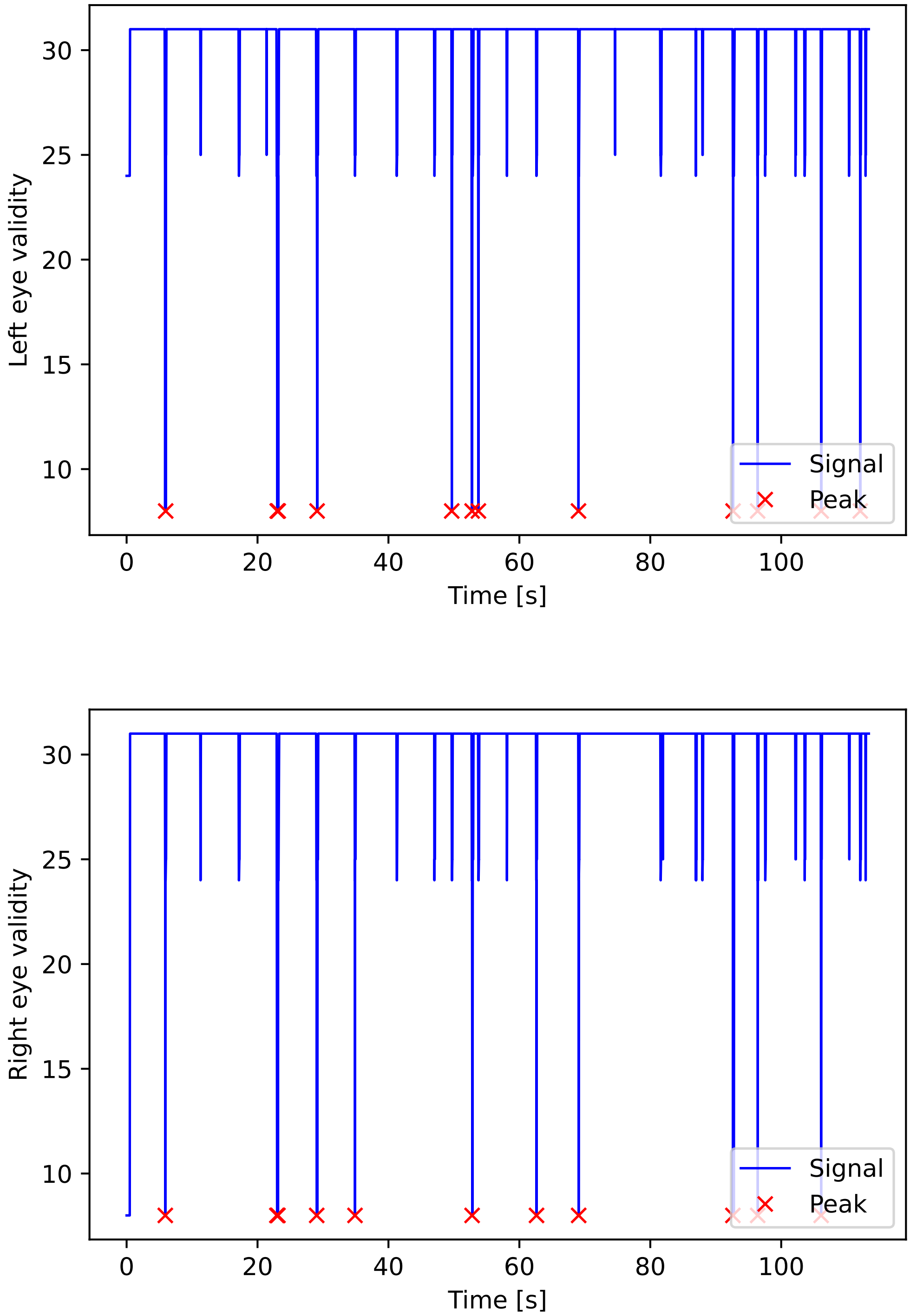


Figure 5.6: Peaks detected in the eye validity level signal for both eyes.

Pulse Rate & Pulse IBI Processing

For the pulse data, both the pulse rate and IBI were processed using a Scipy Signal Butterworth low-pass filter, similar to the one implemented in the eye pupil dilation signal processing. Figure 5.7 shows the difference between the raw signals and the filtered signals using the function:

```
scipy.signal.butter(n, wn, fs, btype='lowpass', analog=False)
```

Where `n` is the order of the filter, the order value used was 2, the `wn` is the critical or cutoff frequency and the value used was 1 Hz. Lastly, the `fs` is the sampling frequency at which the digital signal was acquired; this value corresponds to 100 Hz. As mentioned previously, the `scipy.signal.butter()` function returns two arrays `b, a` that correspond to the numerator (`b`) and denominator (`a`) polynomials of the filter, that are then used as parameters in the Scipy Signal Lfilter function:

```
scipy.signal.lfilter(b, a, x)
```

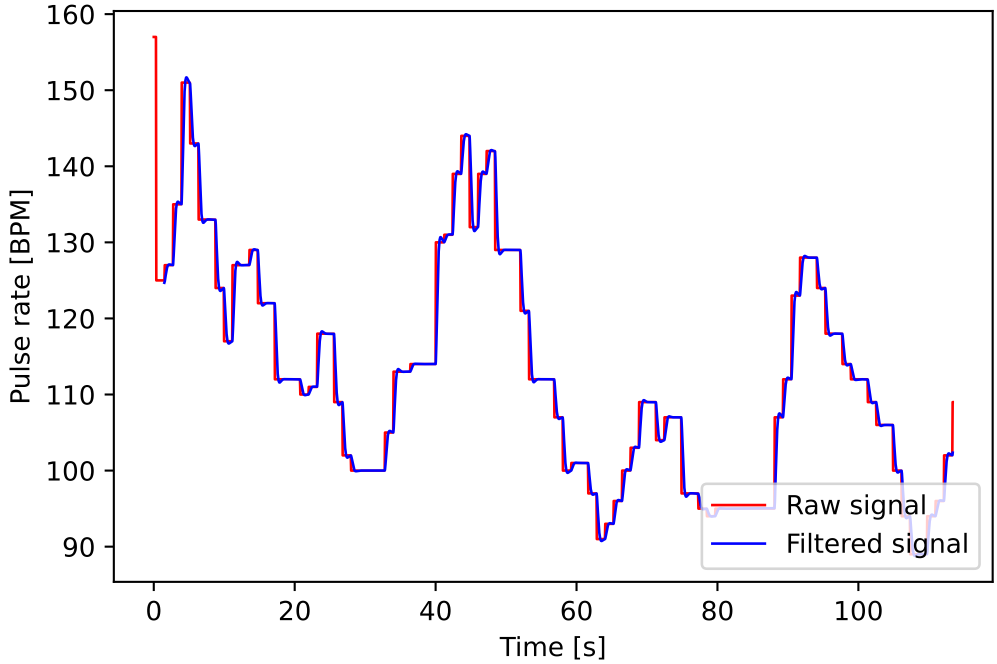
Where `b` is the numerator array, `a` is the denominator array, and `x` is the array with the signal data, this function returns `y`, which is the output of the digital signal after being filtered.

After filtering both signals with the low-pass filter the arithmetic average of the remaining data points was calculated with the Numpy Nanmean function:

```
numpy.nanmean(a)
```

Where `a` is the array of data points from which to infer the average, and finally its output is rounded to 2 decimal places with the function `numpy.round_(a, d)` where `a` is the output value of the arithmetic average calculation and `d` is the number of decimals places to preserve, in this case, the parameter is set to the value 2.

Having presented the data processing procedures, the next section focus on the results of the analysis of the collected results.



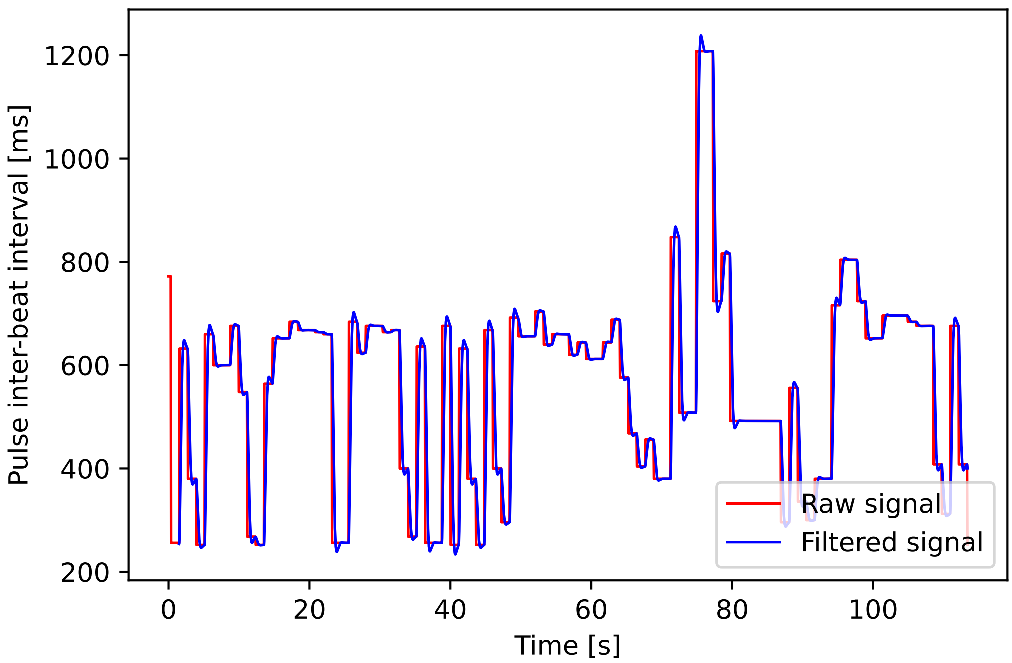


Figure 5.7: Raw and filtered pulse rate and pulse inter-beat interval signals

5.5 Results

In this subsection, the quantitative analysis performed with repeated measures analysis of variance (ANOVA) results are presented and analysed, both individually and collectively. In ANOVA analysis, a lower p-value for a particular factor signifies a higher statistical difference between the averages of the factor.

5.5.1 Repeated Measures Analysis of Variance

Collaboration with Kuka LBR iiwa versus Baxter versus Sawyer

The main factors being analysed in the following subsection are the type of robot and the type of collaboration used for the collaborative assembly task.

Eye data

When analysing the results of the within-subjects effects of the eye **pupil diameter**, displayed in figure 5.8, we can observe a statistically significant difference in the combined pupil diameters for the main effect of the type of robot ($p < 0.001$), meaning that different types of robots yielded different pupil sizes. The mean combined pupil diameter is higher for the robot Baxter and lower for Kuka LBR iiwa and Sawyer. This observation can indicate that humans are under a higher cognitive load (for example, more self-aware or observant) when performing the collaborative task with a robot whose anthropomorphism level is higher, i.e. robots' whose appearance resembles humans the most. In this case, Baxter is the robot with the highest anthropomorphism level amongst the three due to having a torso, two arms and a head, as opposed to Kuka LBR iiwa and Sawyer, which have a single arm without a head, and a single arm with a head, respectively.

Figure 5.9 displays the participants' **blink rate** for the different types of robots and types of collaboration. The p-value for the type of robot factor ($p = 0.280$) shows there is no significant statistical difference for the latter. However, there is a significant statistical difference between the two types of collaboration ($p = 0.003$). Overall the participants' blink rate is lower when performing the collaborative task sequentially as opposed to simultaneously. This can indicate that the participants are under higher cognitive load (or, alternatively, more alert) during the sequential collaboration, thus not blinking as much as in the simultaneous collaboration approach.

The **gaze fixation duration** of the participants on the robots, presented in figure 5.10, measures the time in seconds, the participants' gaze fixated on the robots during the collaborative assembly task. The gaze fixation duration was acquired for 19 participants. There was a statistically significant difference in visual fixation duration for the factor type of robot ($p = 0.007$). The robot with the highest gaze fixation duration is Kuka LBR iiwa for both types of collaboration, followed by Baxter. In contrast, the robot with the lowest overall gaze fixation du-

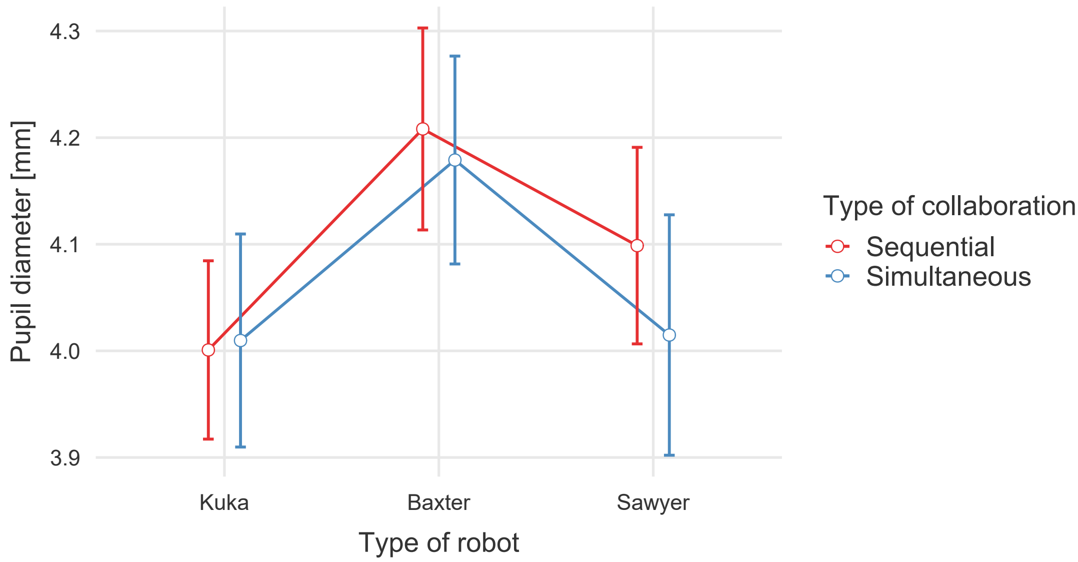
Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	1.3257	2	0.6628	14.82	< .001
Residual	3.0423	68	0.0447		
Type of collaboration	0.0633	1	0.0633	1.46	0.236
Residual	1.4771	34	0.0434		
Type of robot ✻ Type of collaboration	0.0760	2	0.0380	1.66	0.198
Residual	1.5590	68	0.0229		

Note. Type 3 Sums of Squares

Post Hoc Comparisons - Type of robot

Comparison							
Type of robot		Type of robot	Mean Difference	SE	df	t	p_{tukey}
Kuka	-	Baxter	-0.1883	0.0337	34.0	-5.59	< .001
	-	Sawyer	-0.0515	0.0326	34.0	-1.58	0.268
Baxter	-	Sawyer	0.1368	0.0405	34.0	3.38	0.005



Estimated Marginal Means - Type of robot ✻ Type of collaboration

				95% Confidence Interval	
Type of collaboration	Type of robot	Mean	SE	Lower	Upper
Sequential	Kuka	4.00	0.0836	3.83	4.17
	Baxter	4.21	0.0947	4.02	4.40
	Sawyer	4.10	0.0922	3.91	4.29
Simultaneous	Kuka	4.01	0.0999	3.81	4.21
	Baxter	4.18	0.0975	3.98	4.38
	Sawyer	4.01	0.1128	3.79	4.24

Figure 5.8: Combined (left and right) eye pupil diameter in millimetres (mm) data analysis for type of robot and collaboration.

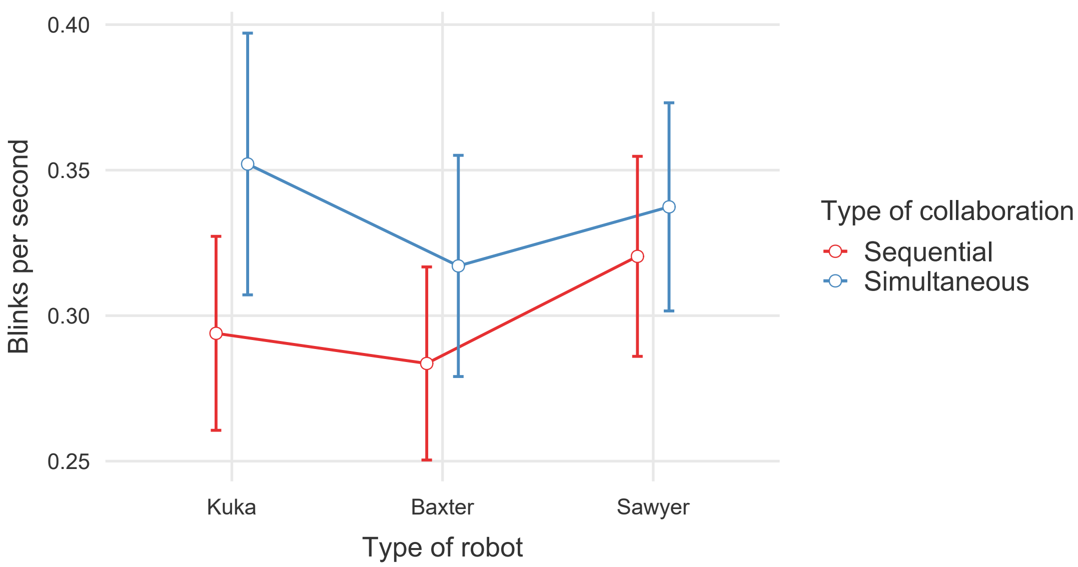
Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	0.0300	2	0.01501	1.30	0.280
Residual	0.7400	64	0.01156		
Type of collaboration	0.0650	1	0.06498	10.63	0.003
Residual	0.1957	32	0.00612		
Type of robot ✻ Type of collaboration	0.0142	2	0.00709	1.40	0.255
Residual	0.3249	64	0.00508		

Note. Type 3 Sums of Squares

Post Hoc Comparisons - Type of collaboration

Comparison							
Type of collaboration		**Type of collaboration**	**Mean Difference**	**SE**	**df**	**t**	**p_{tukey}**
Sequential	-	Simultaneous	-0.0362	0.0111	32.0	-3.26	0.003



Estimated Marginal Means - Type of robot ✻ Type of collaboration

				95% Confidence Interval	
Type of collaboration	**Type of robot**	**Mean**	**SE**	**Lower**	**Upper**
Sequential	Kuka	0.294	0.0333	0.226	0.362
	Baxter	0.284	0.0332	0.216	0.351
	Sawyer	0.320	0.0344	0.250	0.390
Simultaneous	Kuka	0.352	0.0450	0.261	0.444
	Baxter	0.317	0.0380	0.240	0.394
	Sawyer	0.337	0.0358	0.265	0.410

Figure 5.9: Blink rate in blinks-per-second data analysis for type of robot and collaboration.

ration is Sawyer, also keeping the trend in both types of collaboration. There was a statistically significant difference between the two types of collaboration ($p < 0.001$), with the participants focusing on the robots almost twice as much in the sequential collaboration approach versus the simultaneous approach. This can be explained by the fact that the participants need to wait for the robot in the sequential approach, thus having more time to look at it while waiting. In the simultaneous collaboration approach, the participants and the robot move independently, hence the lower visual focus duration, since the participants only look at the robots when necessary to confirm their position and avoid collisions.

Pulse data

In the analysis of the participants' **pulse rate**, as displayed in figure 5.11, the type of robot factor is close to being statistically significant ($p = 0.069$). The participants' pulse rate is higher for the robot Kuka LBR iiwa and lower for the robot Baxter. The higher BPM observed in this context can be explained by Kuka LBR iiwa being the less anthropomorphic robot, which may lead the participants to experience more stress while performing the collaborative task. The lower heart rate observed for both Baxter and Sawyer means that the participants are more at rest or relaxed when performing the collaborative task with these robots, with Baxter, the most anthropomorphic robot, being the one with the lowest average pulse rate. The factor type of collaboration for the pulse rate is not statistically significant ($p = 0.207$). The hearts' BPM and the IBI variables are inversely correlated, meaning that when the heart's BPM is lower, the IBI is higher and vice versa. However, the **pulse IBI**, displayed in figure 5.12, does not show any statistical significance for the factors type of robot ($p = 0.288$) and type of collaboration ($p = 0.455$).

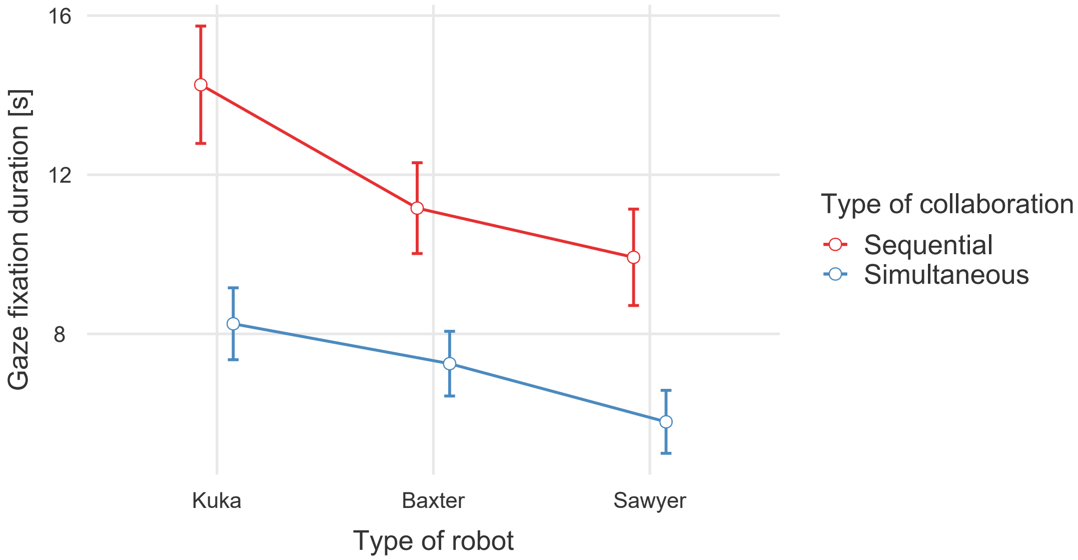
Task duration data

The **task duration**, displayed in figure 5.13, measures the average elapsed time in seconds that the participants take to finish eight trials, i.e. a fully assembled structure of interlocking building blocks. There is a statistically significant difference in both factors, the type of robot ($p < 0.001$) and collaboration ($p < 0.001$). On average, the participants take less time to complete the collaborative task with the robot Sawyer, followed by Kuka LBR iiwa and Baxter. The sequential collaborative approach appears to have a higher duration difference for both Kuka LBR iiwa and Baxter but is less noticeable in the robot Sawyer. These results are particularly interesting because they show that the participants were more effective when collaborating with Sawyer, a hybrid of an industrial robot like Kuka LBR iiwa and a more anthropomorphic robot like Baxter. This means that while anthropomorphism may be more desired for some collaborative tasks, it might not be so for all tasks. In this particular context, the task at hand was assembling interlocking building blocks, so Baxter, due to its human-like appearance, may not be the preferred robot for this specific task, unlike robots like Kuka LBR iiwa or Sawyer with lower anthropomorphism levels.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	222.9	2	111.43	5.72	0.007
Residual	701.9	36	19.50		
Type of collaboration	625.3	1	625.27	52.58	< .001
Residual	214.1	18	11.89		
Type of robot ✻ Type of collaboration	25.2	2	12.61	1.99	0.151
Residual	227.9	36	6.33		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Type of robot ✻ Type of collaboration

				95% Confidence Interval	
Type of collaboration	**Type of robot**	**Mean**	**SE**	**Lower**	**Upper**
Sequential	Kuka	14.26	1.475	11.16	17.36
	Baxter	11.16	1.141	8.76	13.56
	Sawyer	9.92	1.211	7.38	12.47
Simultaneous	Kuka	8.25	0.904	6.36	10.15
	Baxter	7.25	0.813	5.54	8.96
	Sawyer	5.79	0.791	4.13	7.45

Figure 5.10: Average gaze fixation duration on robots in seconds (s) data analysis for type of robot and collaboration.

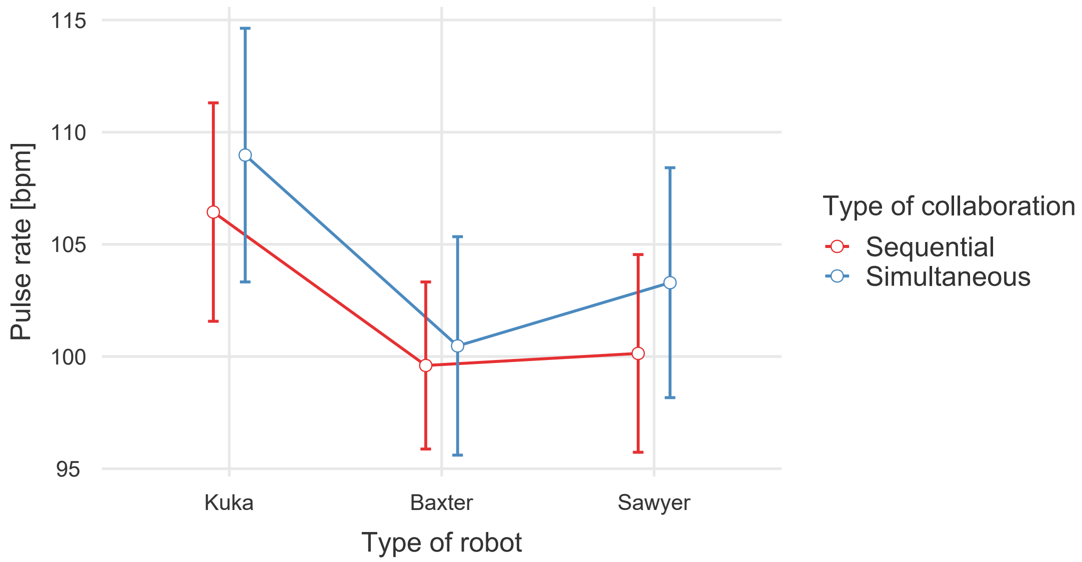
Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	2016.8	2	1008.4	2.801	0.069
Residual	21604.3	60	360.1		
Type of collaboration	222.6	1	222.6	1.664	0.207
Residual	4012.8	30	133.8		
Type of robot ✻ Type of collaboration	43.2	2	21.6	0.131	0.877
Residual	9854.6	60	164.2		

Note. Type 3 Sums of Squares

Post Hoc Comparisons - Type of robot

Comparison							
Type of robot		**Type of robot**	**Mean Difference**	**SE**	**df**	**t**	**p_{tukey}**
Kuka	-	Baxter	7.67	2.57	30.0	2.980	0.015
	-	Sawyer	5.99	3.58	30.0	1.672	0.232
Baxter	-	Sawyer	-1.68	3.92	30.0	-0.428	0.904



Estimated Marginal Means - Type of robot ✻ Type of collaboration

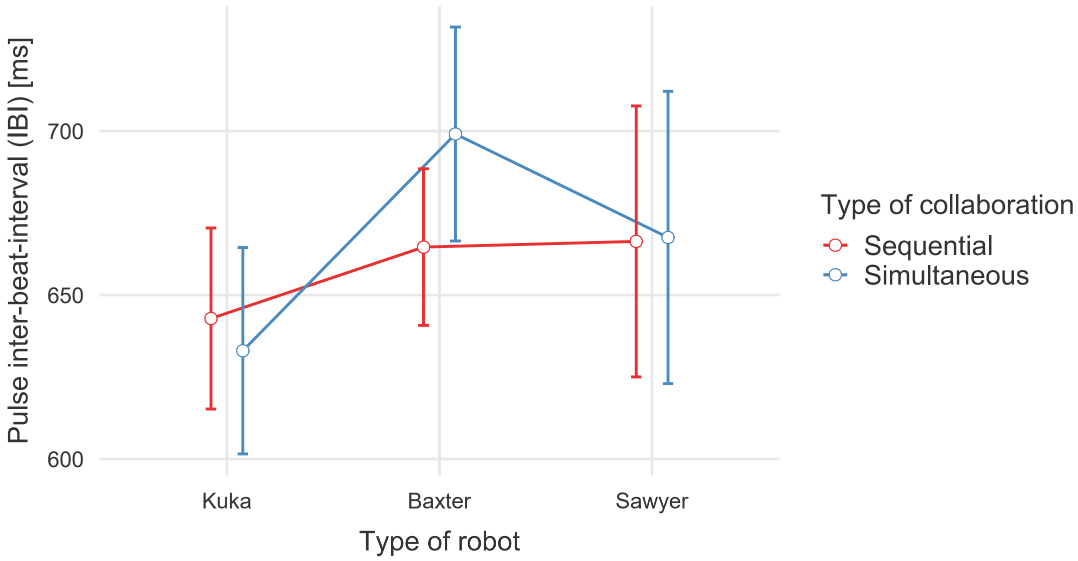
				95% Confidence Interval	
Type of collaboration	**Type of robot**	**Mean**	**SE**	**Lower**	**Upper**
Sequential	Kuka	106.4	4.87	96.5	116
	Baxter	99.6	3.72	92.0	107
	Sawyer	100.1	4.41	91.1	109
Simultaneous	Kuka	109.0	5.65	97.4	121
	Baxter	100.5	4.87	90.5	110
	Sawyer	103.3	5.12	92.8	114

Figure 5.11: Pulse rate in beats-per-minute (BPM) data analysis for the type of robot and collaboration.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	61924	2	30962	1.272	0.288
Residual	1.46e+6	60	24341		
Type of collaboration	3457	1	3457	0.574	0.455
Residual	180804	30	6027		
Type of robot ✻ Type of collaboration	16498	2	8249	0.868	0.425
Residual	570031	60	9501		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Type of robot ✻ Type of collaboration

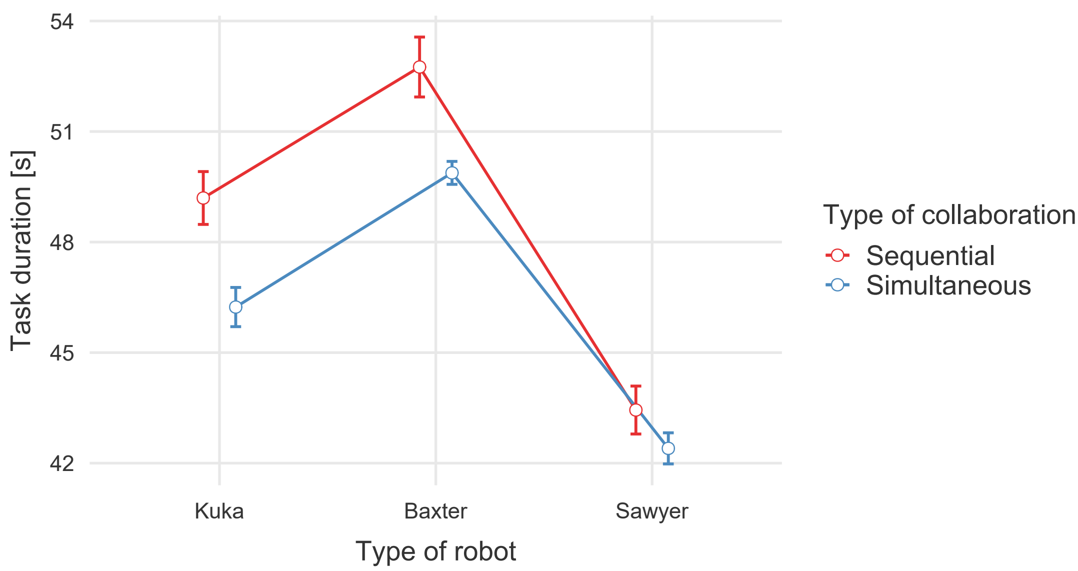
				95% Confidence Interval	
Type of collaboration	Type of robot	Mean	SE	Lower	Upper
Sequential	Kuka	643	27.6	586	699
	Baxter	665	23.9	616	713
	Sawyer	666	41.3	582	751
Simultaneous	Kuka	633	31.5	569	697
	Baxter	699	32.6	632	766
	Sawyer	668	44.6	577	759

Figure 5.12: Pulse IBI in milliseconds (ms) data analysis for the type of robot and collaboration.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	2552.9	2	1276.47	128.00	< .001
Residual	698.1	70	9.97		
Type of collaboration	283.4	1	283.41	16.55	< .001
Residual	599.5	35	17.13		
Type of robot ✻ Type of collaboration	42.3	2	21.15	3.45	0.037
Residual	429.0	70	6.13		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Type of robot ✻ Type of collaboration

				95% Confidence Interval	
Type of collaboration	**Type of robot**	**Mean**	**SE**	**Lower**	**Upper**
Sequential	Kuka	49.2	0.717	47.7	50.7
	Baxter	52.8	0.813	51.1	54.4
	Sawyer	43.4	0.650	42.1	44.8
Simultaneous	Kuka	46.2	0.532	45.2	47.3
	Baxter	49.9	0.311	49.2	50.5
	Sawyer	42.4	0.422	41.5	43.3

Figure 5.13: Task duration in seconds (s) data analysis for the type of robot and collaboration.

Baxter and Sawyer monitor display On and Off

A comparison of the data acquired between the robots Baxter and Sawyer with their monitors, i.e. animated faces, turned off and turned on, was performed to find if this human-like characteristic impacted the participants' performance. The acquired data from the eyes, pulse and task duration was analysed using repeated measures ANOVA. Table 5.4 summarises the data analysis.

The **pupil diameter** data, depicted in figure 5.14, shows a statistically significant difference for the factor type of robot ($p = 0.002$) and the factor type of collaboration ($p = 0.025$), however not as significant for the factor monitor state ($p = 0.164$), indicating that there is no difference in pupil diameters for the animated faces being turned on or off. The **blink rate** data, represented in figure 5.15, shows an almost statistically significant difference for the factor type of collaboration ($p = 0.02$), yet not significant for the factor monitor state ($p = 0.787$), indicating that there is no difference in rates for the animated faces being turned on or off. The data for **gaze fixation duration** on robots' animated faces, represented in figure 5.16, shows a statistically significant difference for the factor type of collaboration ($p < 0.001$) and for the monitor state ($p = 0.007$). This indicates a difference in gaze fixation duration on the robots' faces, showing that the participants spent more time looking at the robots' faces when turned on, with the difference being more pronounced on Baxter, and less time when turned off. The pulse data, the **pulse rate** depicted in figure 5.17 and the **pulse IBI** depicted in figure 5.18, does not show any statistically significant differences. The **task duration** data, represented in figure 5.19, shows there is a statistically significant difference for the factors type of robot ($p < 0.001$) and the factor type of collaboration ($p < 0.001$), however, no significant difference is reported for the factor monitor state ($p = 0.460$).

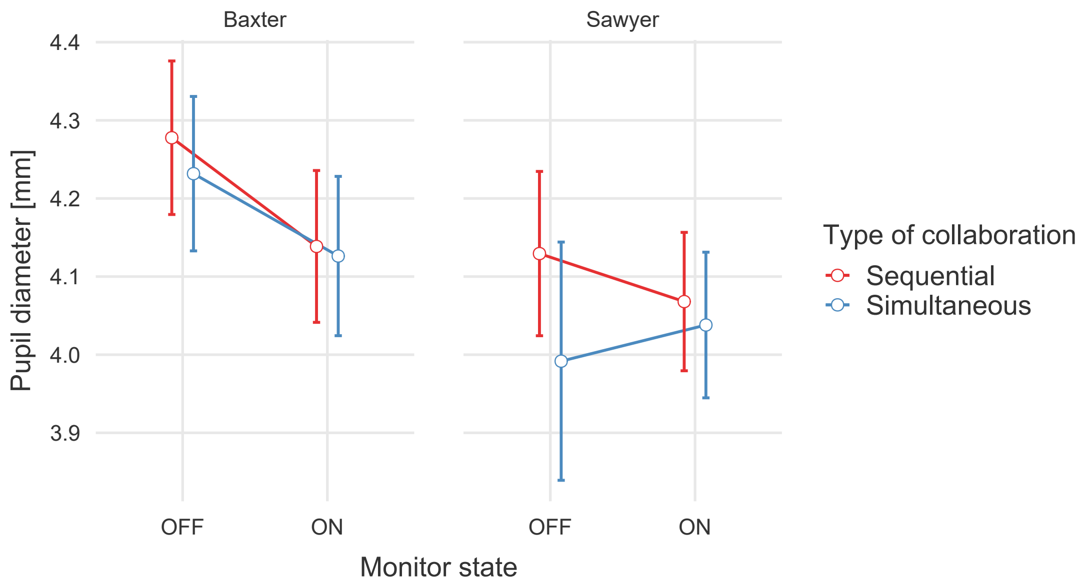
		Mean					
		Pupil diameter [mm]	**Blink rate [blinks/s]**	**Visual focus duration [s]**	**Pulse rate [bpm]**	**Pulse IBI [ms]**	**Task duration [s]**
Sequential	Baxter Off	4.28	0.370	9.49	100.2	676	56.1
	Baxter On	4.14	0.328	12.83	99.0	653	55.4
	Sawyer Off	4.13	0.389	9.81	97.3	648	46.6
	Sawyer On	4.07	0.361	10.04	103.0	684	46.3
Simultaneous	Baxter Off	4.23	0.373	5.87	102.2	678	49.9
	Baxter On	4.13	0.394	8.63	98.7	721	49.8
	Sawyer Off	3.99	0.418	5.80	98.6	651	42.6
	Sawyer On	4.04	0.390	5.78	108.0	685	42.2

Table 5.4: A summary table of the acquired data averages for the robots Baxter and Sawyer with their monitor displays turned off and turned on.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	1.310	1	1.3097	11.42	0.002
Residual	3.898	34	0.1146		
Type of collaboration	0.223	1	0.2235	5.50	0.025
Residual	1.381	34	0.0406		
Monitor state	0.295	1	0.2951	2.02	0.164
Residual	4.962	34	0.1459		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Monitor state ✻ Type of collaboration ✻ Type of robot

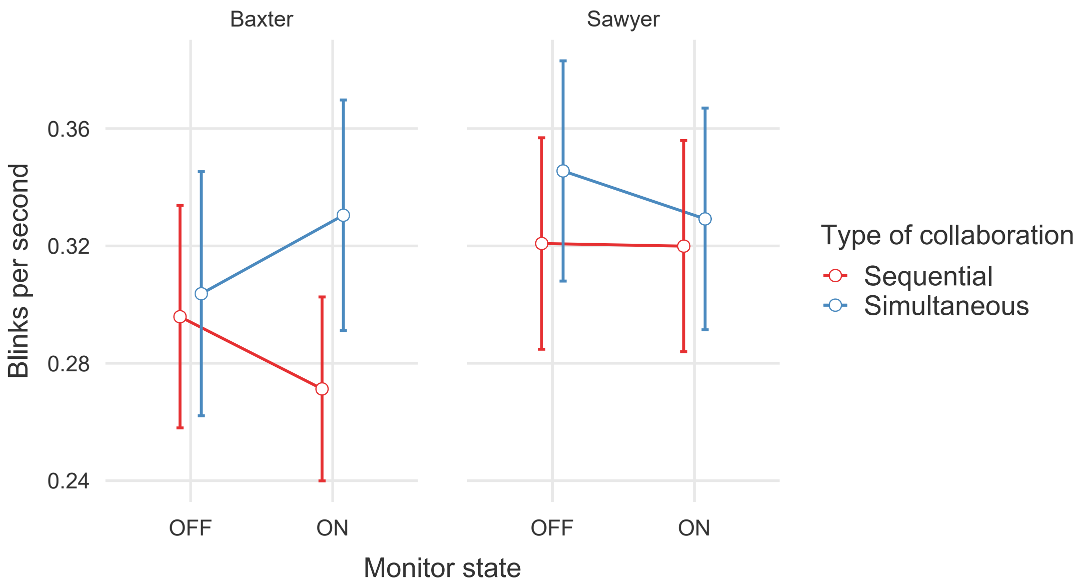
					95% Confidence Interval	
Type of robot	Type of collaboration	Monitor state	Mean	SE	Lower	Upper
Baxter	Sequential	OFF	4.28	0.0983	4.08	4.48
		ON	4.14	0.0972	3.94	4.34
	Simultaneous	OFF	4.23	0.0988	4.03	4.43
		ON	4.13	0.1020	3.92	4.33
Sawyer	Sequential	OFF	4.13	0.1052	3.92	4.34
		ON	4.07	0.0886	3.89	4.25
	Simultaneous	OFF	3.99	0.1525	3.68	4.30
		ON	4.04	0.0932	3.85	4.23

Figure 5.14: Combined (left and right) eye pupil diameter in millimetres (mm) data analysis for monitor state.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	0.0538	1	0.05380	2.9027	0.098
Residual	0.5931	32	0.01853		
Type of collaboration	0.0421	1	0.04210	6.0262	0.020
Residual	0.2236	32	0.00699		
Monitor state	9.42e-4	1	9.42e-4	0.0744	0.787
Residual	0.4052	32	0.01266		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Monitor state ✻ Type of collaboration ✻ Type of robot

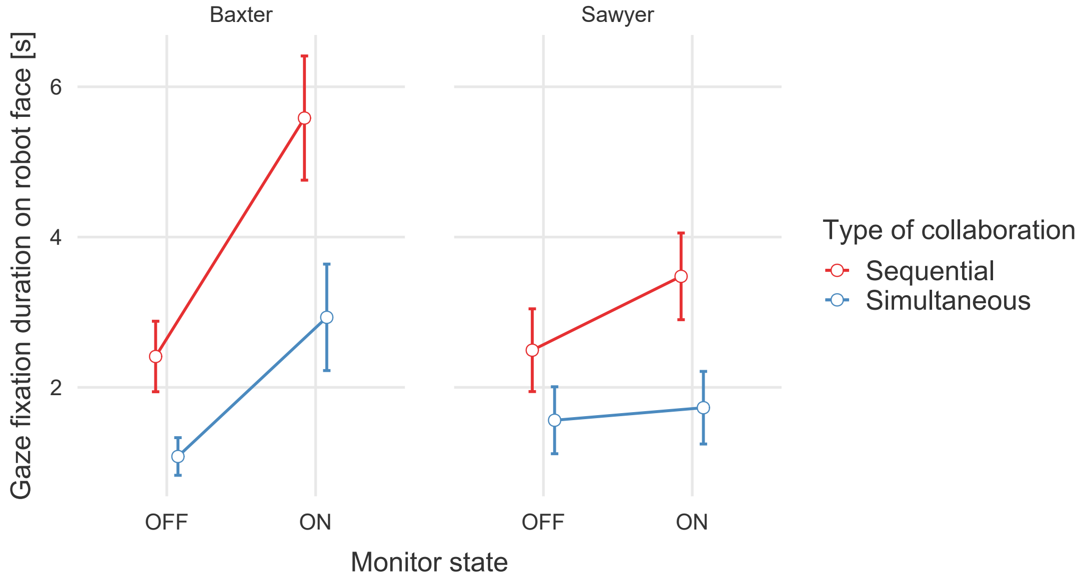
					95% Confidence Interval	
Type of robot	Type of collaboration	Monitor state	Mean	SE	Lower	Upper
Baxter	Sequential	OFF	0.296	0.0379	0.219	0.373
		ON	0.271	0.0313	0.207	0.335
	Simultaneous	OFF	0.304	0.0416	0.219	0.388
		ON	0.330	0.0393	0.250	0.410
Sawyer	Sequential	OFF	0.321	0.0360	0.247	0.394
		ON	0.320	0.0360	0.247	0.393
	Simultaneous	OFF	0.346	0.0375	0.269	0.422
		ON	0.329	0.0378	0.252	0.406

Figure 5.15: Blink rate in blinks-per-second data analysis for monitor state.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	17.9	1	17.87	2.66	0.120
Residual	120.9	18	6.71		
Type of collaboration	105.4	1	105.41	21.15	< .001
Residual	89.7	18	4.98		
Monitor state	90.5	1	90.52	9.37	0.007
Residual	173.9	18	9.66		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Monitor state ✻ Type of collaboration ✻ Type of robot

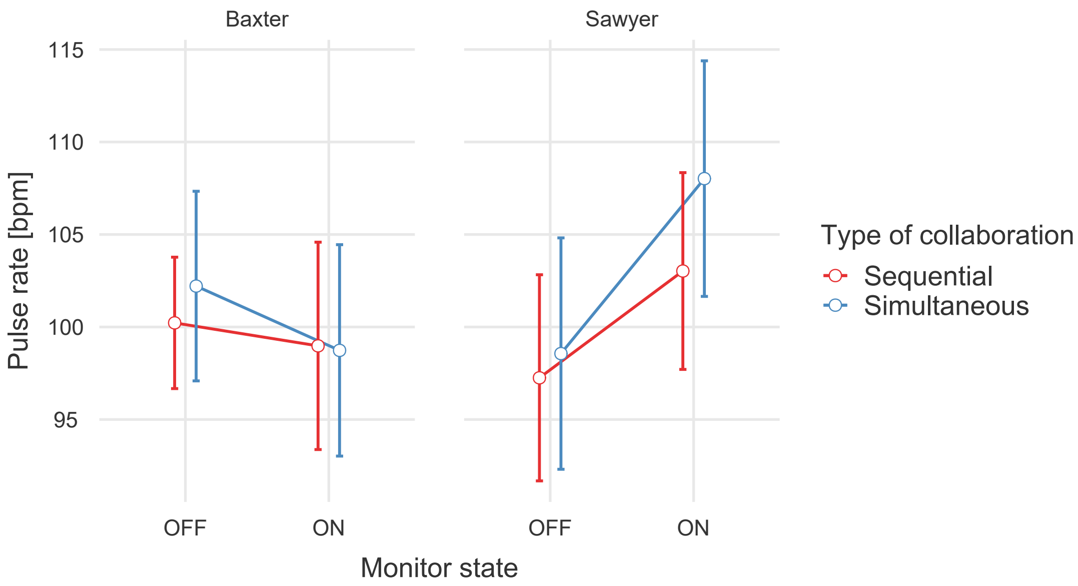
					95% Confidence Interval	
Type of robot	Type of collaboration	Monitor state	Mean	SE	Lower	Upper
Baxter	Sequential	OFF	2.41	0.469	1.425	3.40
		ON	5.58	0.826	3.848	7.32
	Simultaneous	OFF	1.08	0.250	0.554	1.61
		ON	2.93	0.708	1.444	4.42
Sawyer	Sequential	OFF	2.49	0.550	1.338	3.65
		ON	3.48	0.577	2.266	4.69
	Simultaneous	OFF	1.56	0.446	0.627	2.50
		ON	1.73	0.483	0.715	2.74

Figure 5.16: Average gaze fixation duration per robot animated face in seconds (s) data analysis for monitor state.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	175	1	175	0.183	0.672
Residual	28593	30	953		
Type of collaboration	251	1	251	1.107	0.301
Residual	6805	30	227		
Monitor state	428	1	428	0.422	0.521
Residual	30470	30	1016		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Monitor state ✻ Type of collaboration ✻ Type of robot

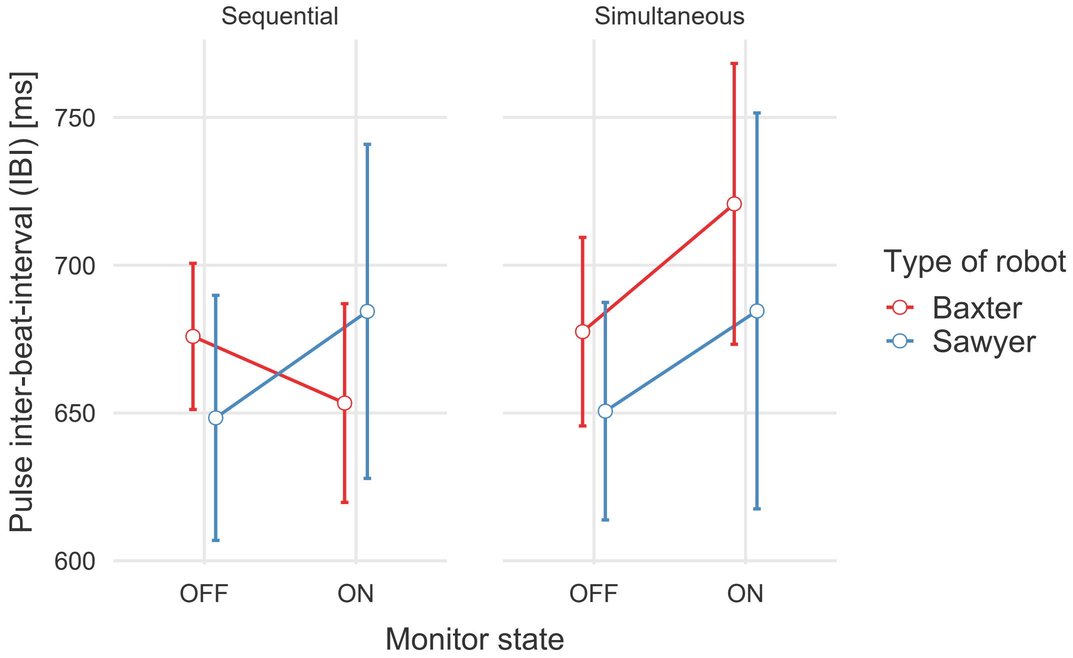
					95% Confidence Interval	
Type of robot	Type of collaboration	Monitor state	Mean	SE	Lower	Upper
Baxter	Sequential	OFF	100.2	3.55	93.0	107
		ON	99.0	5.61	87.5	110
	Simultaneous	OFF	102.2	5.13	91.7	113
		ON	98.7	5.71	87.1	110
Sawyer	Sequential	OFF	97.3	5.57	85.9	109
		ON	103.0	5.32	92.2	114
	Simultaneous	OFF	98.6	6.25	85.8	111
		ON	108.0	6.37	95.0	121

Figure 5.17: Pulse rate in beats-per-minute (BPM) data analysis for monitor state.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	13785	1	13785	0.188	0.668
Residual	2.20e+6	30	73382		
Type of collaboration	19751	1	19751	1.224	0.277
Residual	484138	30	16138		
Monitor state	31815	1	31815	0.630	0.434
Residual	1.52e+6	30	50525		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Monitor state ✻ Type of robot ✻ Type of collaboration

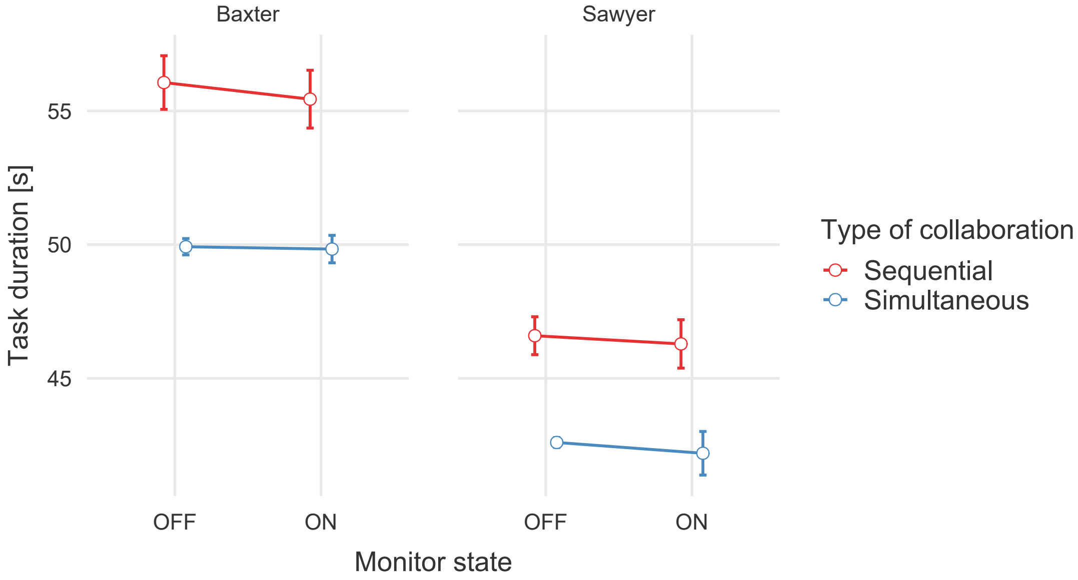
					95% Confidence Interval	
Type of collaboration	Type of robot	Monitor state	Mean	SE	Lower	Upper
Sequential	Baxter	OFF	676	24.7	625	726
		ON	653	33.6	585	722
	Sawyer	OFF	648	41.5	564	733
		ON	684	56.5	569	800
Simultaneous	Baxter	OFF	678	31.9	612	743
		ON	721	47.5	624	818
	Sawyer	OFF	651	36.8	575	726
		ON	685	67.0	548	821

Figure 5.18: Pulse IBI in milliseconds (ms) data analysis for monitor state.

Within Subjects Effects

	Sum of Squares	df	Mean Square	F	p
Type of robot	5071.50	1	5071.50	342.527	< .001
Residual	518.21	35	14.81		
Type of collaboration	1768.89	1	1768.89	58.064	< .001
Residual	1066.25	35	30.46		
Monitor state	9.07	1	9.07	0.559	0.460
Residual	568.13	35	16.23		

Note. Type 3 Sums of Squares



Estimated Marginal Means - Monitor state ✻ Type of collaboration ✻ Type of robot

					95% Confidence Interval	
Type of robot	**Type of collaboration**	**Monitor state**	**Mean**	**SE**	**Lower**	**Upper**
Baxter	Sequential	OFF	56.1	1.002	54.0	58.1
		ON	55.4	1.081	53.2	57.6
	Simultaneous	OFF	49.9	0.305	49.3	50.5
		ON	49.8	0.514	48.8	50.9
Sawyer	Sequential	OFF	46.6	0.709	45.2	48.0
		ON	46.3	0.905	44.5	48.1
	Simultaneous	OFF	42.6	0.184	42.2	43.0
		ON	42.2	0.816	40.5	43.9

Figure 5.19: Task duration in seconds (s) data analysis for monitor state.

5.5.2 Human Performance

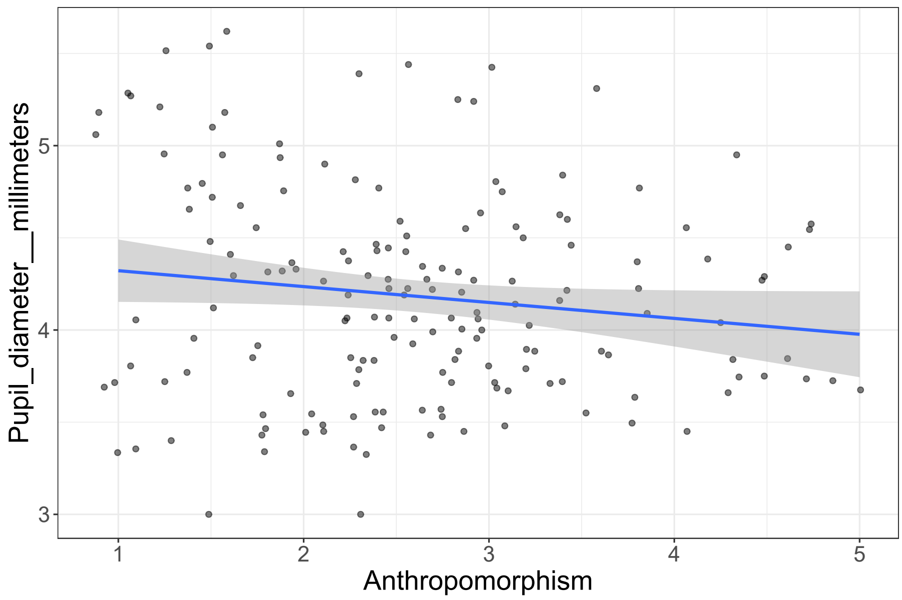
In the following section, the general linear models between the physiological and psychological data are presented in more detail. The scatter diagram linear trend lines show patterns in the data, the grey areas represent the standard errors, and the dots represent the data points.

The **pupil diameter** and the anthropomorphism, as displayed in figure 5.20, shows a negative trend that is a nearing statistically significant difference ($p = 0.062$), indicating that the pupil diameter decreases as the anthropomorphism increases.

The **blink rate** and the anthropomorphism, displayed in figure 5.21, indicate no linear relation and no statistical significance ($p = 0.862$).

The **pulse rate** and the anthropomorphism, displayed in figure 5.22, indicate no linear relation and no statistical significance ($p = 0.401$).

The **task duration** and the anthropomorphism, displayed in figure 5.23, indicate no linear relation and no statistical significance ($p = 0.208$).



Model Info

Info	
Estimate	Linear model fit by OLS
Call	Pupil_diameter__millimeters ~ 1 + Anthropomorphism
R-squared	0.0207
Adj. R-squared	0.0149

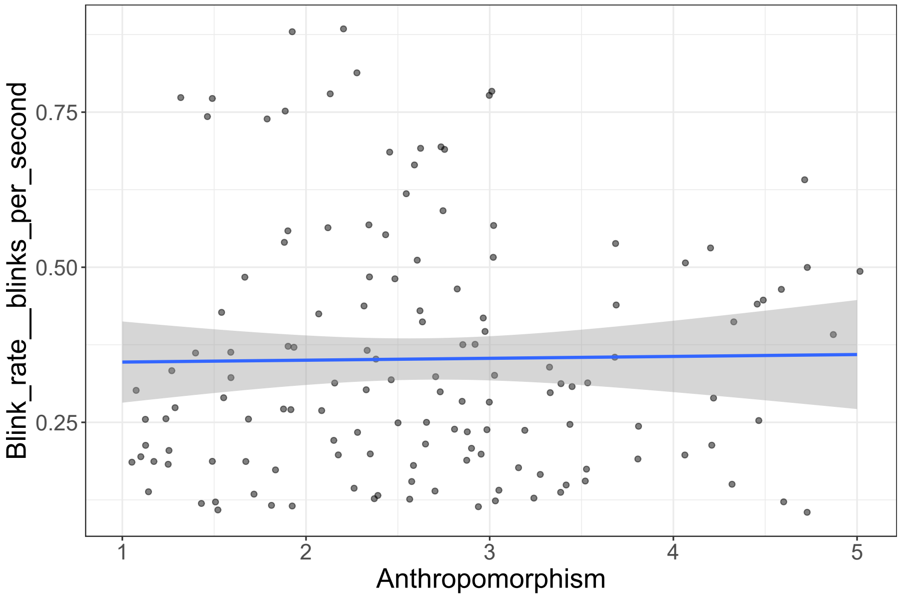
ANOVA Omnibus tests

	SS	df	F	p	η^2p
Model	1.13	1	3.53	0.062	0.021
Anthropomorphism	1.13	1	3.53	0.062	0.021
Residuals	53.33	167			
Total	54.45	168			

Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	**Estimate**	**SE**	**Lower**	**Upper**	**β**	**df**	**t**	**p**
(Intercept)	4.1829	0.0435	4.097	4.26875	0.000	167	96.23	< .001
Anthropomorphism	-0.0862	0.0459	-0.177	0.00433	-0.144	167	-1.88	0.062

Figure 5.20: Pupil diameter (vertical axis) in millimetres (mm) relation with anthropomorphism (horizontal axis) group scores.



Model Info

Info	
Estimate	Linear model fit by OLS
Call	Blink_rate__blinks_per_second ~ 1 + Anthropomorphism
R-squared	2.22e-4
Adj. R-squared	-0.00708

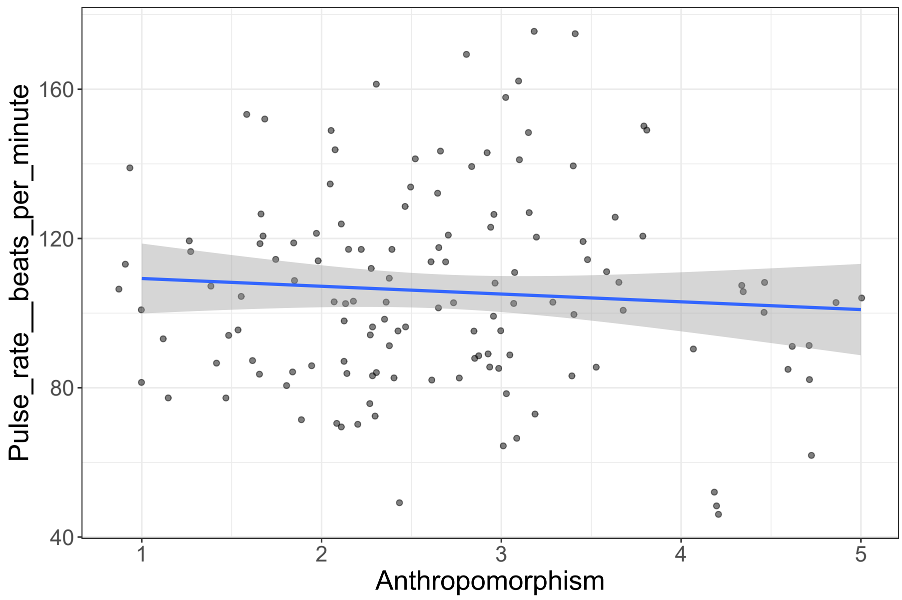
ANOVA Omnibus tests

	SS	df	F	p	η^2p
Model	0.00120	1	0.0304	0.862	0.000
Anthropomorphism	0.00120	1	0.0304	0.862	0.000
Residuals	5.42312	137			
Total	5.42433	138			

Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	Estimate	SE	Lower	Upper	β	df	t	p
(Intercept)	0.35211	0.0169	0.3187	0.3855	0.0000	137	20.865	< .001
Anthropomorphism	0.00303	0.0174	-0.0314	0.0374	0.0149	137	0.174	0.862

Figure 5.21: Blink rate (vertical axis) in blinks per second relation with anthropomorphism (horizontal axis) group scores.



Model Info

Info	
Estimate	Linear model fit by OLS
Call	Pulse_rate__beats_per_minute ~ 1 + Anthropomorphism
R-squared	0.00535
Adj. R-squared	-0.00218

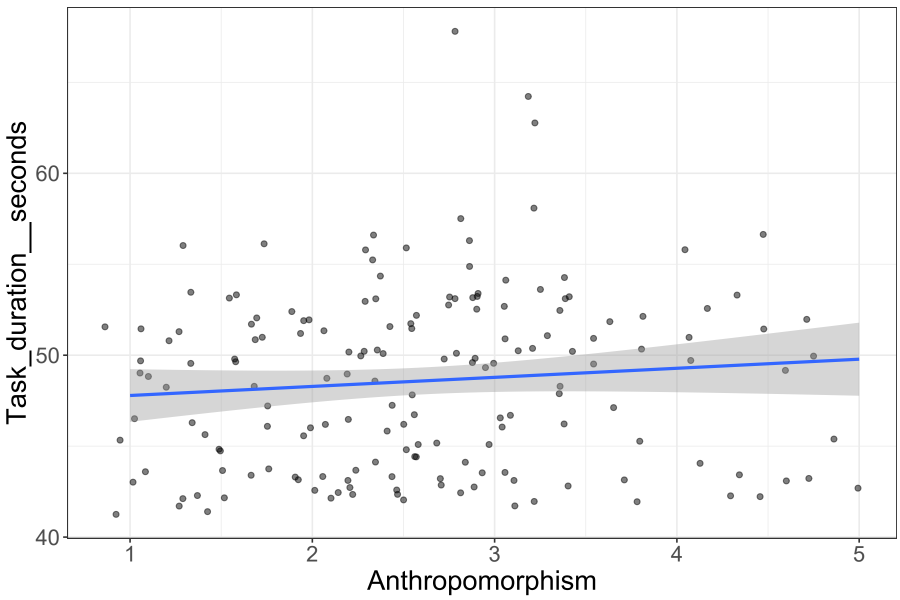
ANOVA Omnibus tests

	SS	df	F	p	η^2p
Model	505	1	0.710	0.401	0.005
Anthropomorphism	505	1	0.710	0.401	0.005
Residuals	93896	132			
Total	94401	133			

Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	Estimate	SE	Lower	Upper	β	df	t	p
(Intercept)	105.79	2.30	101.23	110.35	0.0000	132	45.916	< .001
Anthropomorphism	-2.08	2.47	-6.97	2.80	-0.0732	132	-0.843	0.401

Figure 5.22: Pulse rate (vertical axis) in beats per minute (bpm) relation with anthropomorphism (horizontal axis) group scores.



Model Info

Info	
Estimate	Linear model fit by OLS
Call	Task_duration__seconds ~ 1 + Anthropomorphism
R-squared	0.00943
Adj. R-squared	0.00354

ANOVA Omnibus tests

	SS	df	F	p	η^2p
Model	38.3	1	1.60	0.208	0.009
Anthropomorphism	38.3	1	1.60	0.208	0.009
Residuals	4021.8	168			
Total	4060.1	169			

Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	Estimate	SE	Lower	Upper	β	df	t	p
(Intercept)	48.583	0.375	47.842	49.32	0.0000	168	129.47	< .001
Anthropomorphism	0.499	0.394	-0.280	1.28	0.0971	168	1.26	0.208

Figure 5.23: Task duration (vertical axis) in seconds (s) relation with anthropomorphism (horizontal axis) group scores.

5.5.3 Godspeed Questionnaire Series

This section presents a brief overview of the results of the Godspeed Questionnaire Series. In the end, it also discusses the uncanny valley effect.

Description of the questionnaire series dimensions

As described before, the Godspeed Questionnaire is composed of the following dimensions: anthropomorphism, animacy, likeability, intelligence and safety. In the **anthropomorphism** questionnaire group, the participants rate the different conditions concerning their perception of the robots' anthropomorphism, i.e. how closely the robots resemble the human morphology. In the **animacy** questionnaire group, participants rate the different conditions regarding their perception of the robots' animacy, i.e. the robots' expression of life through movement and appearance. In the **likeability** questionnaire group, the participants rate the different conditions regarding their perception of the robots' likeability, i.e. their perceived sympathy, kindness and friendliness. In the **intelligence** questionnaire group, the participants rate the different conditions regarding their perception of the robots' intelligence, i.e. how knowledgeable, competent and intelligent the robots' are perceived to be. Lastly, in the **safety** questionnaire group, the participants rate the different conditions regarding their perception of the robots' safety, i.e. how safe the participants feel when interacting with the different robots.

To verify the internal consistency reliability of the items in a questionnaire group, Cronbach's α was calculated for each condition and questionnaire group, as suggested by Bartneck et al., the questionnaire authors. The results obtained from the calculations are presented in the table 5.5 and are generally good, i.e. ≤ 0.7, except for the "Baxter On" condition on the Perceived Safety questionnaire group, where the lowest value inferred was 0.507. Although lower, the value still deems the results acceptable (Bartneck et al., 2009b).

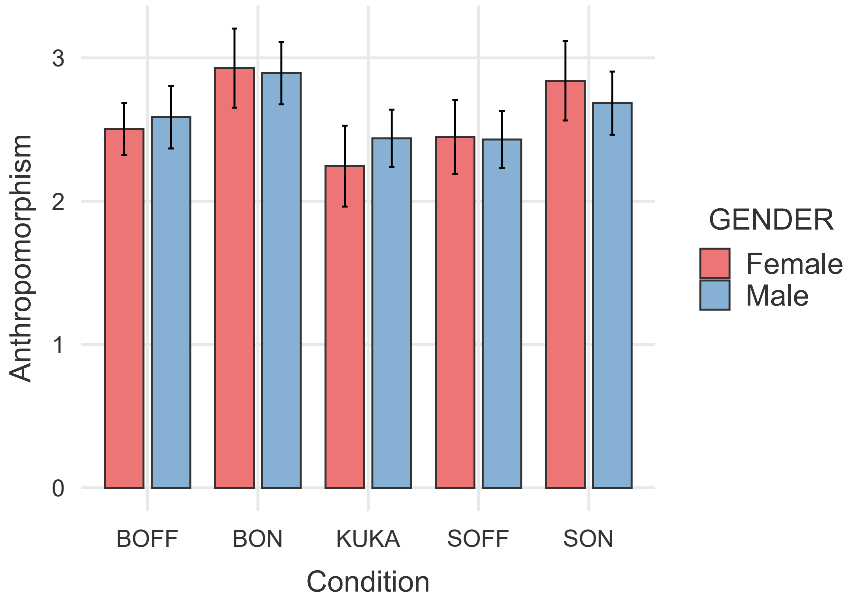
	Questionnaire Group					
Condition	Anthropomorphism	Animacy	Likeability	Perceived Intelligence	Perceived Safety	
Kuka	0.873	0.886	0.923	0.884	0.695	*Cronbach's α*
Baxter Off	0.820	0.853	0.916	0.863	0.761	
Baxter On	0.892	0.884	0.920	0.863	0.507	
Sawyer Off	0.886	0.906	0.915	0.864	0.668	
Sawyer On	0.914	0.903	0.954	0.884	0.720	

Table 5.5: Cronbach's Alpha for each test condition and questionnaire group.

For easier presentation and readability of the psychological results, the names of the conditions were abbreviated in the following graphs and tables. In the description below, the abbreviations and their correspondent condition are shown: **BOFF** - Baxter with monitor turned off, **BON** - Baxter with monitor turned on, **KUKA** - Kuka LBR iiwa, **SOFF** - Sawyer with monitor turned off, **SON** - Sawyer with monitor turned on.

Anthropomorphism

The figure in 5.24 shows a graph with participants' average anthropomorphism score, an higher score value indicates that the participants perceive that condition to be more anthropomorphic. The condition with the highest score is BON, the robot Baxter with the monitor on, with a value of 2.91, followed by SON, the robot Sawyer with the display monitor turned on. The condition with the lowest average rating is KUKA with a score of 2.35. The results reveal that the robot with the highest perceived anthropomorphic level is Baxter, followed by Saywer and Kuka LBR iiwa. There is also a significant difference between the robots Baxter and Sawyer having their monitors turned on or off, with the cases where the monitors were turned on being determined as more anthropomorphic. Concerning different genders' perceptions of the robots' anthropomorphism, there are no notable differences or trends between the females' and the males' attributed scores.



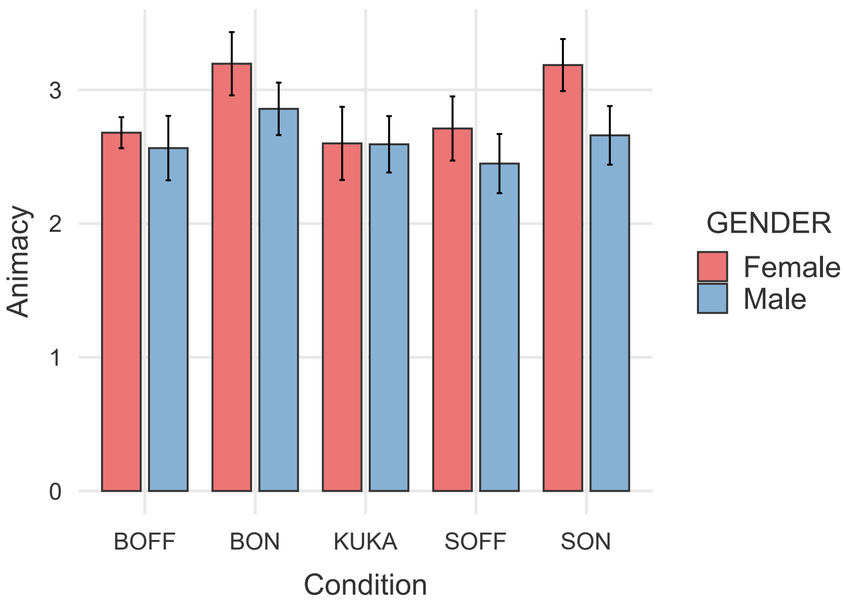
Descriptives

	Condition	Mean	SD
Anthropomorphism	BOFF	2.55	0.841
	BON	2.91	0.988
	KUKA	2.35	0.966
	SOFF	2.44	0.913
	SON	2.75	0.998

Figure 5.24: Anthropomorphism group mean scores for separate genders and their standard error bars. The table displays the anthropomorphism group mean scores for combined genders and their standard deviation (SD).

Animacy

The animacy questionnaire scores, as shown in figure 5.25, present similar results to the anthropomorphism questionnaire, which as mentioned in (Bartneck et al., 2009b), the two questionnaire groups have question items that overlap. In the animacy group, the highest rated condition is BON for the Baxter robot with the monitor on with a score of 3.01, followed by SON for the Sawyer robot with the monitor on. The lowest-rated condition in the animacy group is SOFF, the Sawyer robot with its monitor turned off, with a score of 2.56. However, the average scores for the remainder of the conditions, KUKA and BOFF, all exhibit very close values of 2.60 and 2.62, respectively, indicating that the participants do not perceive significant differences with regard to their animacy. The robots with the monitors turned on were perceived as having more animacy. This indicates that the monitor factor is important for the perception of animacy. Overall, the female participants perceive the robots as holding more animacy than the male participants, except for Kuka LBR iiwa, where both genders perceive the robot animacy similarly.



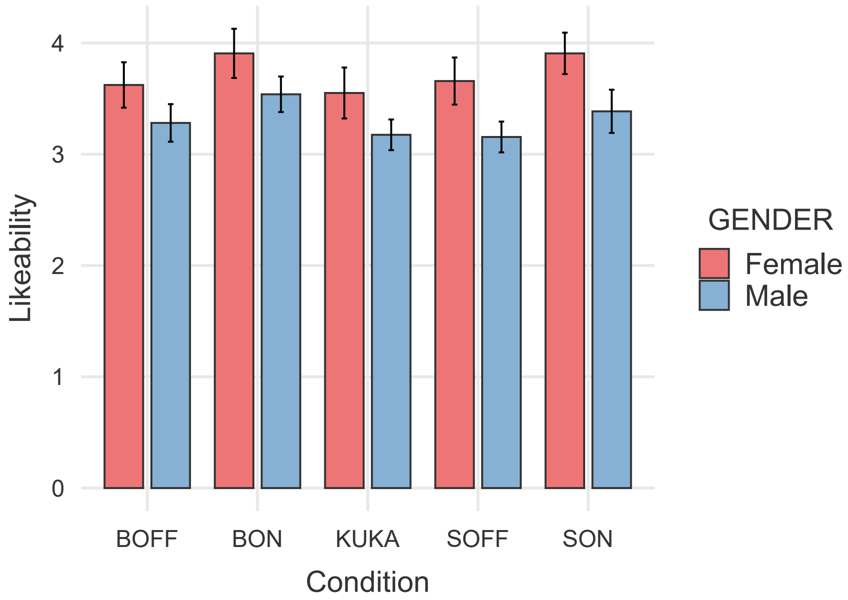
Descriptives

	Condition	Mean	SD
Animacy	BOFF	2.62	0.831
	BON	3.01	0.885
	KUKA	2.60	0.969
	SOFF	2.56	0.944
	SON	2.89	0.899

Figure 5.25: Animacy group mean scores for separate genders and their standard error bars. The table displays the animacy group mean scores for combined genders and their standard deviation (SD).

Likeability

In the likeability questionnaire group scores, as shown in figure 5.26, it is possible to observe that BON, the Baxter robot with his monitor turned on, is the condition which the participants deemed to be more likeable with a score of 3.70, closely followed by the condition SON, the Sawyer robot with his monitor turned on with a score of 3.62. The most disliked robot, as rated by the participants, was Kuka LBR iiwa, with the lowest group score of 3.34, followed by SOFF and BOFF in the respective order. In line with the previously presented results, the most liked robots are both Baxter and Sawyer with their monitors turned on, meaning the participants seem to be more comfortable with the latter. A difference in ratings across the two genders is also visible, with the female participants rating all of the conditions with a higher score than the male participants.



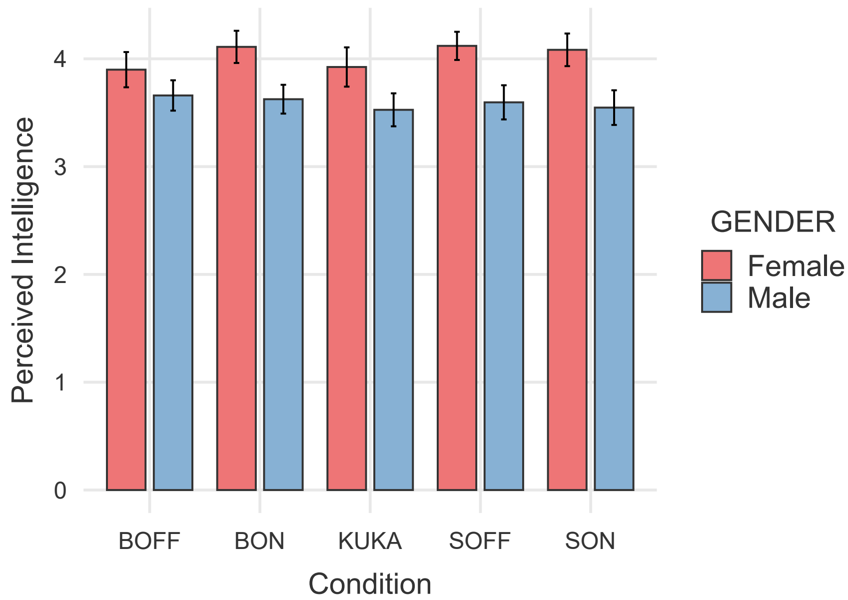
Descriptives

	Condition	Mean	SD
Likeability	BOFF	3.43	0.768
	BON	3.70	0.781
	KUKA	3.34	0.752
	SOFF	3.38	0.740
	SON	3.62	0.825

Figure 5.26: Likeability group mean scores for separate genders and their standard error bars. The table displays the likeability group mean scores for combined genders and their standard deviation (SD).

Intelligence

In the perceived intelligence group scores, the condition with the highest score is BON with a value of 3.84, very closely followed by SOFF with a value of 3.83 and by SON with a value of 3.78. This suggests that the robot the participants perceive as most intelligent is the robot Baxter with his monitor turned on, followed by the robot Sawyer with his monitor turned off, contradicting the previously shown results, where the robots with the monitor turned on were ranked higher in anthropomorphism, animacy and likeability. The robot with the lowest perceived intelligence was Kuka LBR iiwa, with a score of 3.70, followed by Baxter, with the monitor turned off, with a score of 3.77. This indicates that the monitors, whether on or off, were not an essential factor regarding the robots' perceived intelligence. The male participants overall perceived all the robots, independently of the monitor status, to be less intelligent than the female participants perceived the robots to be.



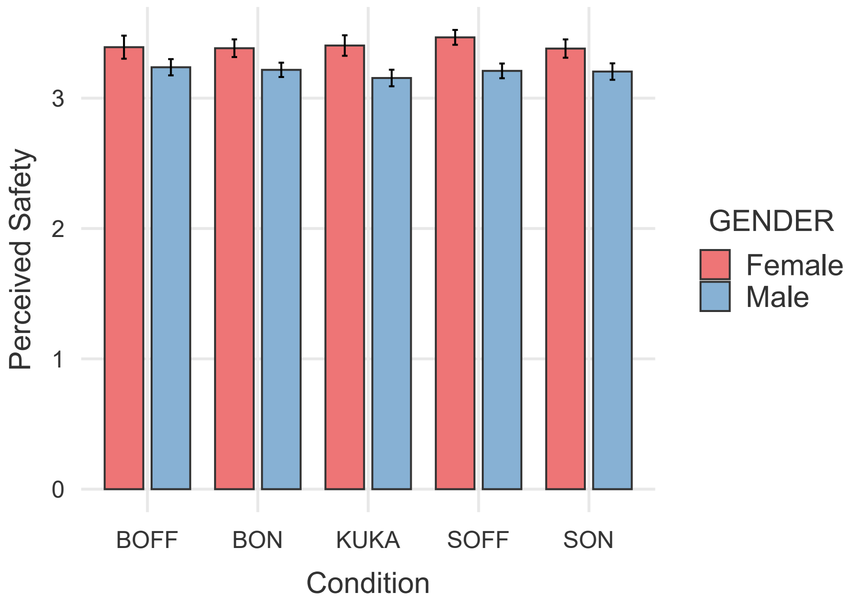
Descriptives

	Condition	Mean	SD
Perceived Intelligence	BOFF	3.77	0.625
	BON	3.84	0.622
	KUKA	3.70	0.703
	SOFF	3.83	0.662
	SON	3.78	0.698

Figure 5.27: Intelligence group mean scores for separate genders and their standard error bars. The table displays the intelligence group mean scores for combined genders and their standard deviation (SD).

Safety

In the perceived safety questionnaire group scores, the group with the lowest *Cronbach* α values, the condition perceived as safer is SOFF with a value of 3.32, meaning that the participants felt safer performing the collaborative task with the robot Sawyer with his monitor turned off, closely followed by BOFF with a value of 3.31. The lowest-scoring condition was for KUKA with a value of 3.26, meaning Kuka LBR iiwa was the robot perceived as less safe by the participants. There are no differences in perceived safety among the two robots with monitors, since they both average 3.30. When comparing the monitor states, on and off, the participants, on average, seemed to feel safer with both robots, Baxter and Sawyer, when their monitors were turned off during the collaboration. The female participants consistently perceived all conditions as safer than the male participants perceived them to be.



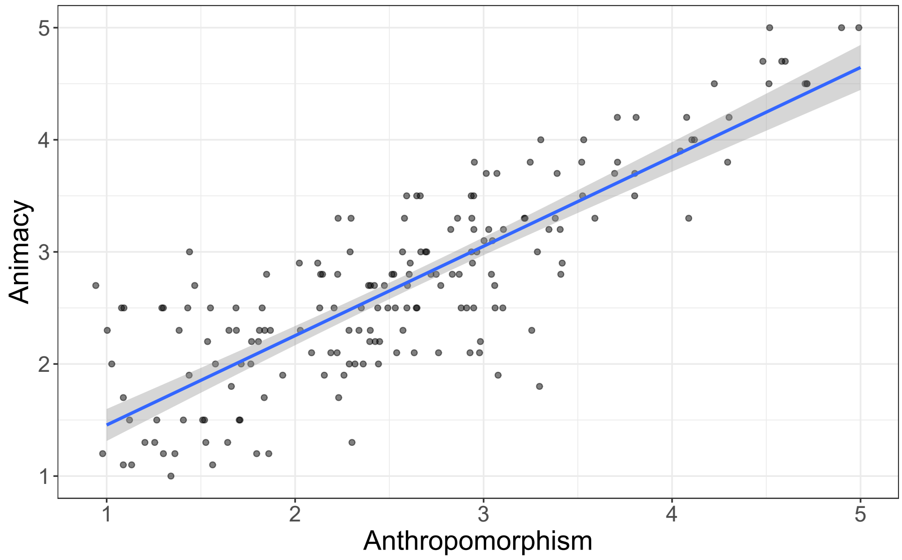
Descriptives

	Condition	Mean	SD
Perceived Safety	BOFF	3.31	0.310
	BON	3.29	0.260
	KUKA	3.26	0.311
	SOFF	3.32	0.266
	SON	3.28	0.284

Figure 5.28: Safety group mean scores for separate genders and their standard error bars. The table displays the safety group mean scores for combined genders and their standard deviation (SD).

Uncanny Valley

The uncanny valley effect observations were obtained from the psychological data acquired in the GQS responses. By comparing the participants' answers to the different questionnaire groups, it is possible to observe statistically significant ($p < 0.001$) positive trends between the perceived anthropomorphism and the remainder of the questionnaire groups: animacy, likeability, intelligence and safety. The results show that, as the perceived **anthropomorphism** increases, the perceived **animacy**, **intelligence**, and **safety** also increase, as displayed in figures 5.29, 5.30 and 5.31, respectively. This indicates that robots perceived as more anthropomorphic are also perceived as more animate, intelligent and safe than robots perceived as less anthropomorphic robots.



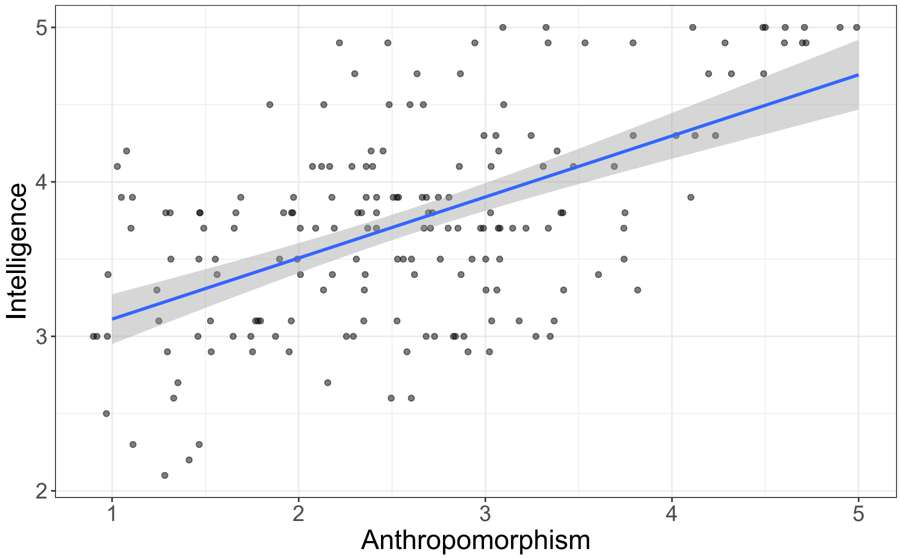
ANOVA Omnibus tests

	SS	df	F	p	$\eta^2 p$
Model	100.8	1	413	< .001	0.699
Anthropomorphism	100.8	1	413	< .001	0.699
Residuals	43.5	178			
Total	144.3	179			

Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	Estimate	SE	Lower	Upper	β	df	t	p
(Intercept)	2.719	0.0368	2.647	2.792	0.000	178	73.8	< .001
Anthropomorphism	0.797	0.0393	0.720	0.875	0.836	178	20.3	< .001

Figure 5.29: Animacy (vertical axis) and anthropomorphism (horizontal axis) group scores relation.



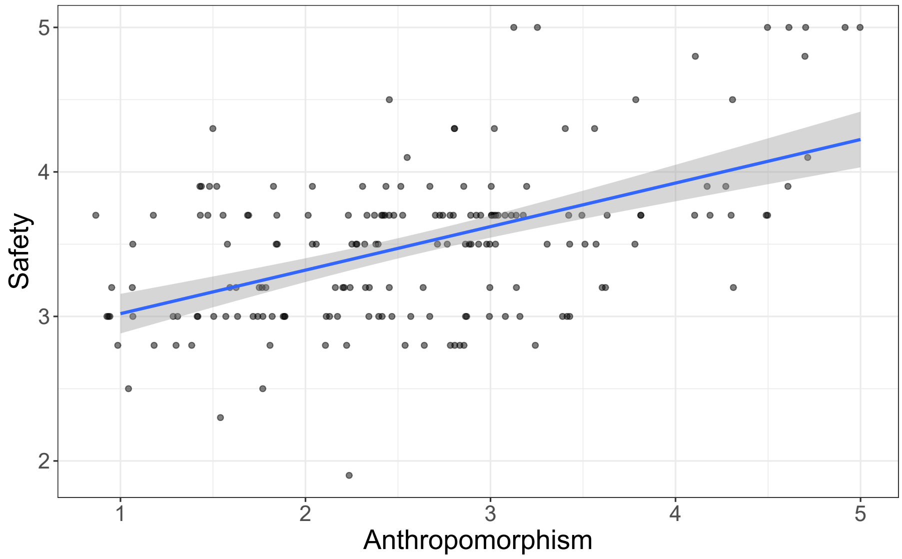
ANOVA Omnibus tests

	SS	df	F	p	η^2p
Model	24.8	1	79.7	< .001	0.309
Anthropomorphism	24.8	1	79.7	< .001	0.309
Residuals	55.3	178			
Total	80.1	179			

Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	**Estimate**	**SE**	**Lower**	**Upper**	**β**	**df**	**t**	**p**
(Intercept)	3.738	0.0416	3.656	3.820	0.000	178	89.95	< .001
Anthropomorphism	0.395	0.0443	0.308	0.483	0.556	178	8.93	< .001

Figure 5.30: Intelligence (vertical axis) and anthropomorphism (horizontal axis) group scores relation.



ANOVA Omnibus tests

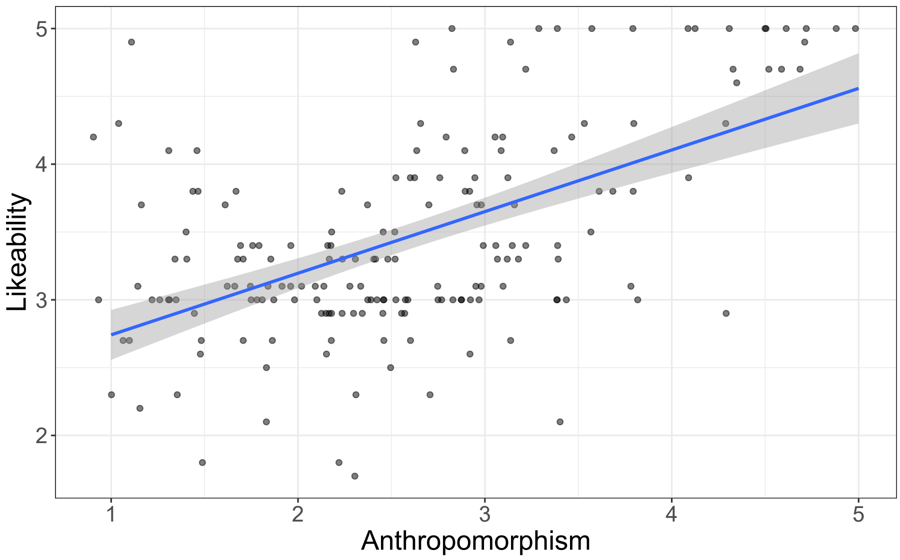
	SS	df	F	p	η^2p
Model	14.4	1	63.9	< .001	0.264
Anthropomorphism	14.4	1	63.9	< .001	0.264
Residuals	40.1	178			
Total	54.5	179			

Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	**Estimate**	**SE**	**Lower**	**Upper**	**β**	**df**	**t**	**p**
(Intercept)	3.497	0.0354	3.427	3.566	0.000	178	98.88	< .001
Anthropomorphism	0.301	0.0377	0.227	0.376	0.514	178	8.00	< .001

Figure 5.31: Safety (vertical axis) and anthropomorphism (horizontal axis) group scores relation.

The perceived **anthropomorphism** and **animacy** of the robots also correlated positively with the perceived **likeability** and are statistically significant ($p < 0.001$), as displayed in figures 5.32 and 5.33 respectively. This indicates that robots perceived as more anthropomorphic and animate are also perceived as more likeable by the participants.



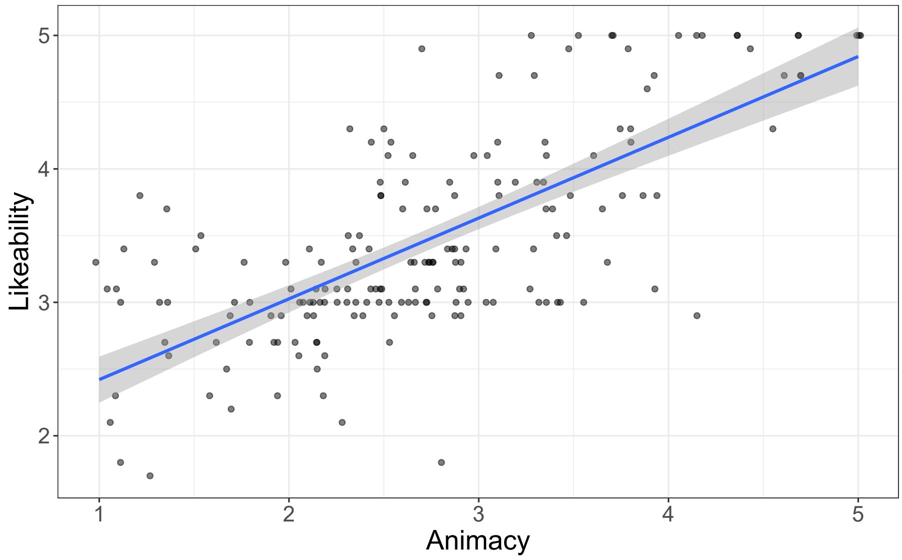
ANOVA Omnibus tests

	SS	df	F	p	η^2p
Model	32.7	1	80.5	< .001	0.311
Anthropomorphism	32.7	1	80.5	< .001	0.311
Residuals	72.3	178			
Total	105.1	179			

Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	**Estimate**	**SE**	**Lower**	**Upper**	**β**	**df**	**t**	**p**
(Intercept)	3.462	0.0475	3.368	3.555	0.000	178	72.85	< .001
Anthropomorphism	0.454	0.0506	0.354	0.554	0.558	178	8.97	< .001

Figure 5.32: Likeability (vertical axis) and anthropomorphism (horizontal axis) group scores relation.



ANOVA Omnibus tests

	SS	df	F	p	η^2p
Model	52.9	1	180	< .001	0.503
Animacy	52.9	1	180	< .001	0.503
Residuals	52.2	178			
Total	105.1	179			

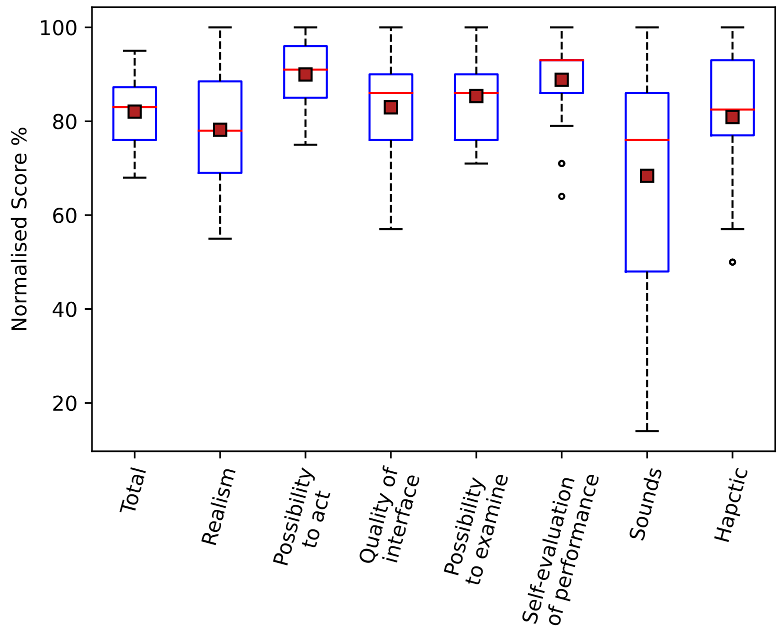
Fixed Effects Parameter Estimates

			95% Confidence Interval					
Names	**Estimate**	**SE**	**Lower**	**Upper**	**β**	**df**	**t**	**p**
(Intercept)	3.462	0.0404	3.382	3.541	0.000	178	85.8	< .001
Animacy	0.605	0.0451	0.516	0.694	0.709	178	13.4	< .001

Figure 5.33: Likeability (vertical axis) and animacy (horizontal axis) group scores relation.

5.5.4 Virtual Reality Presence Questionnaire

This section presents the results of the VR PQ, which covers the following dimensions: realism, possibility to act, quality of interface, possibility to examine, self-evaluation of performance, sounds and haptic. After completing all test conditions, the study participants were asked to report their sense of presence by answering the VR PQ through a digital form on a computer. The Cronbach's α for the questionnaire reported a value of 0.868. The normalised total score of the PQ shows a general acceptance of the interactive VR simulation with an average of 82.1% and a standard deviation of 8.4. Five of the seven evaluated domains show a score above 80%: **haptics**, **quality of the interface**, **possibility to examine** and the **possibility to act**, the latter the highest score of 90%. The **realism** domain of the VR task scored lower, with a value of 78.2%, followed by **sounds**, with an even lower score of 68.4% and a standard deviation of 24.4, indicating there is a lot of variance around the mean for the sounds domain. Although the realism and sounds domains scored the lowest, the total normalised score indicates that most participants reported a good quality VR experience and an overall sense of presence in the interactive VR simulation.



Descriptives

	Domain	Mean	SD
Score	Total	82.1	8.37
	Realism	78.2	13.19
	Possibility to act	90.0	7.95
	Quality of interface	83.0	12.15
	Possibility to examine	85.4	8.93
	Self-evaluation of performance	88.8	9.25
	Sounds	68.4	24.39
	Haptic	80.9	14.75

Figure 5.34: VR PQ normalised scores. The square shapes represent the mean score. The table presents the mean scores and their standard deviation (SD).

5.5.5 Summary of Results

Concerning the three different robots used in this study, results show that **pupil diameter** increased and the **pulse rate** decreased with the increase of the robot's anthropomorphism. This may indicate that participants are under a higher cognitive load but less stressed when performing a collaborative task with robots of higher anthropomorphic levels like Baxter. The participants **gaze fixation duration** on the robots was longer for Kuka LBR iiwa, followed by Baxter and Sawyer. The **task duration** was smaller for the robot Sawyer and was greater for the robot Baxter. The robot that participants spent less time looking at is also the one with the shortest task duration. This shows that Sawyer was the robot with whom the participants collaborated more efficiently, having performed the assembly task faster.

Concerning the two collaborative approaches analysed, sequential and simultaneous collaboration, it seems that the sequential collaboration induced a higher cognitive load and awareness as suggested by the lower **blink rate**. While the pulse data did not show any significant differences in types of collaboration. The participants **gaze fixation duration** on the robots was greater for the sequential collaboration and smaller for the simultaneous collaboration. The **task duration** was longer for the sequential collaboration, indicating that the participants are less efficient when collaborating sequentially and more efficient when collaborating simultaneously. However, Sawyer, the robot with the shortest task duration, does not show a significant difference in the task durations regarding the two types of collaboration.

Regarding the monitor states for the robots Baxter and Sawyer, there are no statistically significant differences for the pupil diameter, blink rate, pulse rate and IBI. The only measure that had statistical significance for the factor monitor state was the **gaze fixation duration** on the robots' animated faces. Indicating that the participants spend more time looking at the robots' faces when turned on, and less time when turned off. A summary of the different variables' results averages is displayed in table 5.6.

		Mean					
		Pupil diameter [mm]	**Blink rate [blinks/s]**	**Gaze fixation duration [s]**	**Pulse rate [bpm]**	**Pulse IBI [ms]**	**Task duration [s]**
Sequential	Kuka	4.00	0.294	14.26	106.4	643	47.7
	Baxter	4.21	0.284	11.16	99.6	665	51.1
	Sawyer	4.10	0.320	9.92	100.1	666	42.1
Simultaneous	Kuka	4.01	0.352	8.25	109.0	633	45.2
	Baxter	4.18	0.317	7.25	100.5	699	49.2
	Sawyer	4.01	0.337	5.79	103.3	668	41.5

Table 5.6: A summary of the results with the physiological measures and the task duration.

The results obtained can explain a part of the **uncanny valley** theory but are not reflective of the actual proposed valley, where the affinity is lower for a higher human likeness. According to the results presented in section 5.5.3, the perception of likeability also increases as the participants' perception of robots' human likeness increases, indicating a similar effect to what Mori et al. propose.

In figure 5.35, the robots are classified according to their human likeness, as proposed originally in the uncanny valley. The robot Kuka LBR iiwa, classified as having a low human likeness due to its industrial look, is positioned where (Mori et al., 2012) originally pointed industrial robots to be. Baxter, classified as having a moderate human likeness due to its humanoid appearance, is positioned where (Mori et al., 2012) originally pointed humanoid robots to be. Lastly, Sawyer is classified as having a low-moderate human likeness positioned between Kuka LBR iiwa and Baxter due to its industrial and humanoid appearance.

This classification can indicate that Baxter might not have the human likeness required to be perceived as uncanny, an even more anthropomorphic robot would be needed to cause such an emotional response.

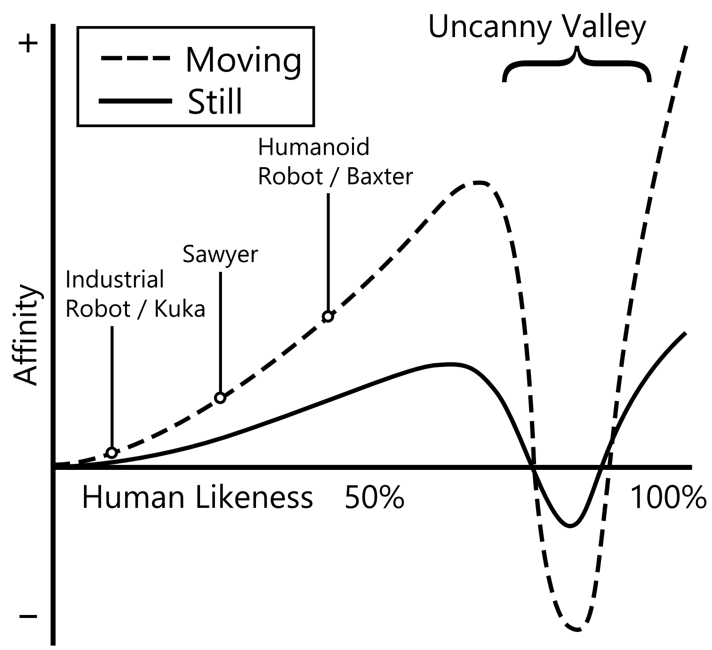


Figure 5.35: The robots Kuka LBR iiwa, Baxter and Sawyer are classified according to their human likeness in the original uncanny valley graph.

The work done in this chapter allowed evaluating HRC with a focus on human performance and the subsequent interpretation and presentation of the results. The following chapter shows a discussion of the results, the limitations and future work.

Chapter 6

Discussion & Future Work

The sixth chapter is divided into three sections. The first section covers the discussion of the results, followed by the research limitations and lastly the future work section.

6.1 Discussion

Concerning the main research question, the results indicate that human performance in a VR human-robot collaborative assembly task scenario is affected by both the robot and the type of collaboration used. We showed that for the assembly task employed, a robot with lower human likeness might be preferable to execute the task in less time, thus having higher efficiency. However, robots with lower human likeness can induce lower cognitive load and higher stress than robots with a higher human likeness, indicating that the use of a collaborative robot should depend on the task to be performed. For certain applications, using a collaborative robot in conjunction with humans can help improve the task at hand. Even though in a collaborative assembly task such as the one presented, the stress induced by robots with low anthropomorphism may not have a big impact on human performance, the same cannot be said if the task was more critical such as in surgery, where stress could negatively affect the human performance. It is necessary to be prudent when implementing robots collaborative robots so as not to affect the quality and the outcome of the task being performed.

Furthermore, albeit the results indicate that a simultaneous collaborative approach can be more efficient from a task completion time perspective, particularly with Kuka and Baxter, when compared to the sequential collaborative approach, for the robot with the shortest task completion time, Sawyer, there is no significant time difference between the collaborative approaches. Nonetheless, the sequential approach can induce higher cognitive load and awareness when compared to the simultaneous approach, thus indicating that simultaneous collaboration might be the preferable approach.

While the results can explain a portion of the uncanny valley theory, that is, the increase of perceived likeability as perceived human likeness increases, the presence of a negative emotional response or uncanniness and its effect on human performance could not be demonstrated. The robot with the highest anthropomorphism used in this study, Baxter, which can be classified as a humanoid robot, does not have the even higher human likeness required to elicit the uncanniness or eeriness triggered by having an almost human anthropomorphism.

Regarding the evaluation of HRC through VR scenarios, although the VR system and the developed interactive VR simulation have their limitations, the physiological and psychological data suggests that to some extent, it is possible to evaluate a human-robot collaborative assembly task scenario performed in a VR simulation. This is demonstrated by the distinct participants' physiological and psychological responses to the various stimuli presented. This indicates that VR can be a powerful tool to evaluate HRC without compromising safety.

The results presented in this study are adequate and pertinent, however, they also exhibited minor inconsistencies. When comparing the data for the three robots implemented, the eyes' data shows that the pupil diameter increases as the robots' human likeness increases. However, when comparing the eyes' physiological responses data to the psychological responses data, the inverse trend is shown, with the pupil diameter decreasing as the perception of robots' anthropomorphism increases.

In part, this can be explained by the comparison being made, the results of the physiological responses to the stimulus presented with the results of the physiological responses to the participants' perceptions of the robots. This comparison might not be entirely correct due to the participants' perceptions of the robots being subjective to each participant's age, culture and previous experiences and the physiological responses to the stimulus presented being objective measures. However, this inconsistency is negligible since it is not statistically significant and it compares an objective measurement, the physiological responses, with a subjective one, the psychological responses.

6.2 Limitations

The present research was found to have limiting factors across some domains. Namely, concerning the hardware tools used, the interactive VR simulation developed as a possible solution to answer the research questions, and finally, the data processing approach. The hardware tools used in the study, specified in subsection 3.2.1, are susceptible to their limitations. In the case of the VR system, albeit a higher-end system with haptic feedback, the physical component of the interactions between the human and the VR simulation is still limited by the technology.

The VR system embedded eye-tracker sensor also has its limitations. Albeit calibration was performed successfully for every participant, this does not mean there are no eye data quality issues. Besides this, the eye-tracking system's performance can be affected by eye disease, heavy makeup and high myopia.

Restrictions were identified regarding the pulse monitoring system in the sensor and the OpenBCI Cyton board to which the sensor connects. Initially, in the pilot study, participants would carry the Cyton board on themselves in a pouch, but this, too, triggered an increase in signal noise due to the constant movement. To get around this problem, the Cyton board was placed on a bench near the participants instead of on the pouch, improving the sensor readings but not wholly solving the noise introduced by the hand movements. The pulse sensor caused some difficulties during test sessions due to erroneous readings. Factors like moisture, movement and having the sensor wrapped too tightly or loosely to the participants' fingers can affect the sensor readings.

The wires solder where they connect to the printed circuit board (PCB) broke three different times due to being placed on the participants' fingers. Although not placed on the dominant hand, the participants moved their hands naturally while performing the collaborative assembly task in VR, leading to wear and breaking of the wire connections and causing increased signal noise in the sensor readings due to their movements.

Furthermore, the robots used in this research and implemented in the interactive VR simulation present some limiting factors. The robots have varying anthropomorphism levels or human-likeness; however, they only range from low anthropomorphic levels, in the case of Kuka, to moderate anthropomorphic levels, as in the case of Baxter, whereas Sawyer is situated in a low-moderate level, between the robot Kuka and the robot Baxter. To show more accurate results regarding the uncanny emotional response elicited by the uncanny valley theory, a robot with an even higher human likeness would need to be implemented. Although the robots' movement significantly improved from integrating the BioIK inverse kinematics solver, the robots do not use the Robot Operating System (ROS) software, which can affect their movement fidelity.

Moreover, the robots being designed for collaborative applications should have implemented collision detection and other safety and handling features that were not present in this implementation. The end-effector gripper devices of the robots lacked realistic movement when grabbing and inserting the interlocking building blocks and the natural physical interactions that would be expected with the remainder of the objects present in the scenario. Lastly, albeit the animated faces of both Baxter and Sawyer were made to look as close as possible to the originals, they lack the eyebrows and other animated facial expressions, only moving the eyes to follow their end-effector.

6.3 Future Work

To further understand the implications of the results obtained in this study, future research should implement multiple human performance measures, such as human movement hesitations, collisions with robots, task errors and various tasks with different difficulty tiers. In addition, future research should consider using non-intrusive and high-fidelity biological sensors. For example, the presence of cables in the OpenBCI Cyton pulse sensor interferes with natural human movement, reduces VR immersion and can introduce noise in the sensor readings. Instead, non-invasive wireless electrocardiography (ECG) devices should be considered to achieve higher fidelity results without negatively impacting human movement and immersion. Furthermore, other physiological data could be collected, such as electrodermal activity (EDA) and brain electrical activity through electroencephalography (EEG), providing further insight into human physiological responses.

Further research is needed to understand the effects of the uncanny valley on human performance. Studies should consider using specifically made questionnaires to infer psychological responses to the uncanny valley. The use of robots with near-human likeness and higher fidelity in terms of behaviour and interactions should also be taken into account when attempting to reproduce the negative emotional response proposed by Mori et al..

Concerning the development of interactive VR simulations to evaluate HRC, future research endeavours should integrate Robot Operating System (ROS) in their applications when there is compatibility, as it helps to simulate robots' behaviours more accurately. Additionally, more research is needed to assess the impact of a virtual versus a real stimulus on humans' physiological and psychological responses. This could be achieved by having humans perform the same collaborative task with the same robots in similar virtual and real environments.

Finally, future work concerning this dissertation includes the implementation of final adjustments and preparation of the developed interactive VR simulation Unity project and its documentation, to later share with the scientific community through an open repository.

The current chapter presented a discussion of the results, the research limitations and the future work. The next chapter presents the conclusion.

Chapter 7

Conclusion

This research aimed to understand the effects of different robot characteristics and collaborative approaches on human performance while performing a collaborative assembly task through VR. The analysis of the eyes and pulse data indicates that humans are under a higher cognitive load and awareness, but less stressed, when performing a collaborative assembly task with robots of higher human likeness, i.e. higher anthropomorphism. Additionally, results indicate that the use of higher robot anthropomorphism levels, like Baxter, is not always preferable to perform a collaborative assembly task, when compared to lower robot anthropomorphism levels, such as Kuka LBR iiwa and Sawyer, as shown by the task completion times experimental results.

Regarding the two collaborative approaches investigated, the physiological data shows that the sequential approach caused higher cognitive load and awareness. In turn, the analysis of the task completion times shows that humans are more efficient when collaborating simultaneously with robots, thus indicating that a simultaneous collaboration is not only more efficient from a task completion time standpoint, but also induces a lower cognitive load and awareness in humans, thus making it the preferable approach.

This work is relevant to the field of HRC and for the Neurocobots project because it shows that human performance in collaborative assembly tasks executed through VR is affected by the type of robot and the type of collaboration used. This research also demonstrates that VR is useful to perform the evaluation of HRC, however, further research can help to specify the differences between using real and virtual stimuli. This work also contributes to the fields of HRC and VR by sharing the conceptualised, designed and developed interactive VR simulation used to evaluate human performance in HRC. The tasks of making this simulation available to the scientific community through an open repository are ongoing. Once fully available, this work will allow future researchers to build upon the project and employ it in their studies.

References

Alenljung, B., Lindblom, J., Andreasson, R., and Ziemke, T. (2017). User Experience in Social Human-Robot Interaction. *International Journal of Ambient Computing and Intelligence*, 8(2):12–31.

Badia, S. B. i., Silva, P. A., Branco, D., Pinto, A., Carvalho, C., Menezes, P., Almeida, J., and Pilacinski, A. (2022). Virtual Reality for Safe Testing and Development in Collaborative Robotics: Challenges and Perspectives. *Electronics*, 11(11):1726. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.

Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2009a). My robotic doppelgänger - A critical look at the Uncanny Valley. pages 269–276.

Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009b). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics*, 1(1):71–81.

Bauer, A., Wollherr, D., and Buss, M. (2008). Human-Robot Collaboration: a Survey. *I. J. Humanoid Robotics*, 5:47–66.

Bieller, D. S. (2021). Director Statistical Dpt. Vice Chair IFR Service Robot Group. page 43.

Bornemann, J. (2021). KUKA LBR iiwa - GrabCAD Community.

Cambridge Dictionary (2022a). Definition of collaboration from the Cambridge Advanced Learner's Dictionary & Thesaurus.

Cambridge Dictionary (2022b). Definition of interaction from the Cambridge Advanced Learner's Dictionary & Thesaurus.

Cambridge Dictionary (2022c). Definition of user experience from the Cambridge Advanced Learner's Dictionary & Thesaurus.

Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S., and Sheikh, Y. A. (2019). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*.

Castro, A., Silva, F., and Santos, V. (2021). Trends of Human-Robot Collaboration in Industry Contexts: Handover, Learning, and Metrics. *Sensors*, 21(12):4113.

Cohen, Y., Shoval, S., Faccio, M., and Minto, R. (2021). Deploying Cobots in Collaborative Systems: Major Considerations and Productivity Analysis. *International Journal of Production Research*.

Dautenhahn, K. (2007a). Methodology & Themes of Human-Robot Interaction: A Growing Research Field. *International Journal of Advanced Robotic Systems*, 4.

Dautenhahn, K. (2007b). Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):679–704.

Dautenhahn, K. (2013). Human-Robot Interaction. *Human-Robot Interaction*, (The Encyclopedia of Human-Computer Interaction, 2nd Ed.).

Dautenhahn, K. (2018). Some Brief Thoughts on the Past and Future of Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction*, 7(1):4:1–4:3.

Demir, K. A., Döven, G., and Sezen, B. (2019). Industry 5.0 and Human-Robot Co-working. *Procedia Computer Science*, 158:688–695.

Dianatfar, M., Latokartano, J., and Lanz, M. (2021). Review on existing VR/AR solutions in human–robot collaboration. *Procedia CIRP*, 97:407–411.

DiSalvo, C. F., Gemperle, F., Forlizzi, J., and Kiesler, S. (2002). All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, DIS '02, pages 321–326, New York, NY, USA. Association for Computing Machinery.

Falter, D., Babcock, S., Burks, B., Lloyd, P., Randolph, J., Rutenber, J., and Van Hoesen, S. (1995). Remote systems for waste retrieval from the Oak Ridge National Laboratory gunite tanks. Technical report, United States. CONF-951006–33 INIS Reference Number: 27047832.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3):143–166.

Freedy, A., DeVisser, E., Weltman, G., and Coeyman, N. (2007). Measurement of trust in human-robot collaboration. In *2007 International Symposium on Collaborative Technologies and Systems*, pages 106–114.

Gallucci, M. (2019). GAMLj: General analyses for linear models. [jamovi module].

Gee, F., Browne, W., and Kawamura, K. (2005). Uncanny valley revisited. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 151–157. ISSN: 1944-9437.

Goetz, J., Kiesler, S., and Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.*, pages 55–60.

Goodrich, M. and Schultz, A. (2007). Human-Robot Interaction: A Survey. *Foundations and Trends in Human-Computer Interaction*, 1:203–275.

Gray, K. and Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1):125–130.

Guizzo, E. (2020). Types of robots - robots: Your guide to the world of robotics.

Hassenzahl, M. and Roto, V. (2007). Being and doing: A perspective on user experience and its measurement. *Interfaces*, 72:10–12.

Hassenzahl, M. and Tractinsky, N. (2006). User experience - A research agenda. *Behaviour and Information Technology*, 25:91 – 97.

IFR (2020). Demystifying Collaborative Industrial Robots.

ISO/TC 299 (2016). ISO/TS 15066:2016.

Kolbeinsson, A., Lagerstedt, E., and Lindblom, J. (2019). Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing. *Production & Manufacturing Research*, 7(1):448–471. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21693277.2019.1645628.

Kwak, S. S. (2014). The Impact of the Robot Appearance Types on Social Interaction with a Robot and Service Evaluation of a Robot. *Archives of Design Research*.

LaValle, S. (2020). Chapter 1: Introduction. In *Virtual Reality*. Cambridge University Press.

Lindblom, J. and Cort, R. (2016). Current Challenges for UX Evaluation of Human-Robot Interaction. volume 490, pages 267–277.

MacDorman, K., Green, R., Ho, C.-C., and Koch, C. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, 25:695–710.

Malawey, D. (2021). Rethink Robotics Sawyer - GrabCAD Community.

Marr, B. (2017). The 4 Ds Of Robotization: Dull, Dirty, Dangerous And Dear. Section: Tech.

Matheson, E., Minto, R., Zampieri, E., Faccio, M., and Rosati, G. (2019). Human–Robot Collaboration in Manufacturing Applications: A Review. *Robotics*, 8:100.

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine*, 19(2):98–100. Conference Name: IEEE Robotics Automation Magazine.

Oxford Advanced Learner's Dictionary (2022a). collaboration noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com.

Oxford Advanced Learner's Dictionary (2022b). interaction noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com.

Oxford Advanced Learner's Dictionary (2022c). user-experience noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com.

Powers, A. and Kiesler, S. (2006). The Advisor Robot: Tracing People's Mental Model from a Robot's Physical Attributes. volume 2006, pages 218–225.

R Core Team (2021). R: A Language and environment for statistical computing. (Version 4.1) [Computer software].

Rosenthal-von der Pütten, A. M. and Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, 36:422–439.

Scholtz, J. (2003). Theory and Evaluation of Human Robot Interactions.

Shea, R. (2016). Collaborative robot technical specification iso/ts 15066 update.

Simon, T., Joo, H., Matthews, I., and Sheikh, Y. (2017). Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*.

Starke, S., Hendrich, N., Krupke, D., and Zhang, J. (2017a). Evolutionary multi-objective inverse kinematics on highly articulated and humanoid robots. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6959–6966. ISSN: 2153-0866.

Starke, S., Hendrich, N., Magg, S., and Zhang, J. (2016). An efficient hybridization of Genetic Algorithms and Particle Swarm Optimization for inverse kinematics. In *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1782–1789.

Starke, S., Hendrich, N., and Zhang, J. (2017b). A memetic evolutionary algorithm for real-time articulated kinematic motion. In *2017 IEEE Congress on Evolutionary Computation (CEC)*, pages 2473–2479.

Stein, J.-P. and Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition*, 160:43–50.

Terveen, L. G. (1995). Overview of human-computer collaboration. *Knowledge-Based Systems*, 8(2):67–81.

The jamovi project (2022). jamovi - open statistical software for the desktop and cloud (Version 2.3) [Computer Software].

Thrun, S. (2004). Toward a Framework for Human-Robot Interaction. *Human-computer Interaction*, 19:9–24.

Thrun, S., Hahnel, D., Ferguson, D., Montemerlo, M., Triebel, R., Burgard, W., Baker, C., Omohundro, Z., Thayer, S., and Whittaker, W. (2003). A system for volumetric robotic mapping of abandoned mines. In *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, volume 3, pages 4270–4275 vol.3. ISSN: 1050-4729.

Wei, S.-E., Ramakrishna, V., Kanade, T., and Sheikh, Y. (2016). Convolutional pose machines. In *CVPR*.

Weiss, A., Wortmeier, A.-K., and Kubicek, B. (2021). Cobots in Industry 4.0: A Roadmap for Future Practice Studies on Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems*, 51(4):335–345. Conference Name: IEEE Transactions on Human-Machine Systems.

Weistroffer, V., Paljic, A., Callebert, L., and Fuchs, P. (2013). A methodology to assess the acceptability of human-robot collaboration using virtual reality. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, VRST '13, pages 39–48, New York, NY, USA. Association for Computing Machinery.

Wilhelm, B., Manfred, B., Braun, M., Rally, P., and Scholtz, O. (2016). *Lightweight robots in manual assembly – best to start simply! Examining companies' initial experiences with lightweight robots*.

Witmer, B. G., Jerome, C. J., and Singer, M. J. (2005). The Factor Structure of the Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 14(3):298–312.

Yanco, H. and Drury, J. (2004). Classifying human-robot interaction: an updated taxonomy. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 3, pages 2841–2846, The Hague, Netherlands. IEEE.

Appendices

Appendix A

Collaborative robots review

Cobot	Manufacturer	Collaboration	Standard	Industries	Applications
ABB IRB 14000 YUM/ ABB IRB 14050 YUMI [1]	ABB	Cooperation, Responsive	ISO 10218, ISO 13849, ISO/TS 15066	AS, ET	A
AUBO i-series [2]	AUBO	Sequential	ISO 10218	AS, ET, MC, MM	A, D, MT, QI
Automata Eva [345]	Automata	Cooperation, Sequential	ISO 13849, ISO/TS 15066	MC	QI
Cobotta[6]	Denso	Cooperation, Sequential	ISO 10218, ISO 13849, ISO/TS 15066	ES, MC	A, MH, QI
P3/P5	Elephant Robotics	Cooperation, Sequential	ISO 10218, ISO 13849	AS, C, ES, ET, MC, MM	A, MT, W
M/H series[7]	Doosan Robotics	Sequential	ISO 13849	FB, MC, PP	A, D, MT, MH, MR, QI, W
CS series	Elibot	Sequential	ISO 10218	AS, ES, ET, FB, MM	A, D, F, MH, QI
CR series[8]	Fanuc	Sequential	ISO 10218, ISO 13849	AS, ET, FB, MC, MM, PP	A, D, MT, MH, QI
Rizon[9]	Flexiv	Cooperation, Sequential	ISO 10218, ISO 13849, ISO/TS 15066	A, AS, ES, ET, MC	A, F, MT
LBR iiwa/Med[10]	Kuka	Cooperation, Sequential	ISO 13849	AS, ET, FB, MC, MM	A, D, F, MH, QI
CZ10[11]	Nachi Robotics	Cooperation, Sequential	ISO 10218, ISO/TS 15066	AS, ET	D, MT, MH
Baxter	Rethink Robotics	Cooperation, Sequential		ES, ET, MM, PP	MT, MH, QI
Sawyer Black[1213]	Rethink Robotics	Cooperation, Sequential	ISO 10218	AS, ET, MM, PP	MT, MH, QI
Nextage	Kawada Industries	Cooperation, Responsive			A
PAV series[14]	Precise Automation	Cooperation, Sequential	ISO 13849, ISO/TS 15066	ES, MC	MT, MH, QI
OB7 series	Productive Robotics	Cooperation, Sequential	ISO 10218	AS, ES, ET, FB, MM, PP	MT, MR
Armar	Karlsruhe IT	Cooperation, Responsive		ES	
UR series[15]	Universal Robots	Cooperation, Sequential	ISO 10218, ISO 13849	AS, ES, ET, FB, MC, MM, PP	A, D, F, MT, MH, MR, QI, W
HC series[1617]	Yaskawa	Cooperation, Sequential	ISO 10218, ISO 13849, ISO/TS 15066	AS, C, ET, FB, PP	A, D, F, MT, MH, MR, QI, W

Table A.1: A review summary of the collaborative robots studied, their respective industries and applications. The notes indicate some of the use cases reviewed.

The industries acronyms used are:

A - Aerospace

C - Chemical

AS - Automotive & Subcontractors

ES - Education & Science

ET - Electronics & Technology

FB - Food & Beverage

MM - Metal & Machining

MC - Medical & Cosmetics

PP - Plastic & Polymers

The applications acronyms used are:

A - Assembly (Screwdriver/Part Insertion)

D - Dispensing (Gluing/Sealing/Painting)

MT - Machine tending (CNC/Injection mold/ICT)

QI - Quality inspection (Testing/Inspecting/Measuring)

MH - Material handling (Packaging/Palletizing/Bin Picking/Kitting)

F - Finishing (Sanding/Polishing)

MR - Material removal (Grinding/Deburring/Milling/Routing/Drilling)

W - Welding (Arc/Soldering)

M - Massaging

[1] `https://library.abb.com/r?dk=movie&q=9AKK106713A8993`
[2] `https://www.aubo-cobot.com/public/demo3`
[3] `https://automata.tech/resources/case-studies/scaling-diagnostic-testing-with-lab-automation/`
[4] `https://automata.tech/solutions/diagnostics/nucleic-acid-testing/`
[5] `https://automata.tech/solutions/drug-discovery/screening/`
[6] `https://www.denso-wave.com/ja/robot/katsuyou/collabo.html`
[7] `https://www.doosanrobotics.com/en/Applications/Applications`
[8] `https://www.fanuc.eu/fi/en/customer-cases`
[9] `https://www.flexiv.com/en/assets/pdf/Brochure.pdf`
[10] `https://www.kuka.com/en-de/industries/solutions-database`
[11] `https://www.nachirobotics.com/collaborative-robots/`
[12] `https://www.rethinkrobotics.com/sawyer/industries`
[13] `https://www.rethinkrobotics.com/sawyer/applications`
[14] `http://preciseautomation.com/Collaborative.html`
[15] `https://www.universal-robots.com/case-stories/`
[16] `https://www.yaskawa.eu.com/application/type`
[17] `https://www.yaskawa.eu.com/application/industries`

Appendix B

Procedure Script

The full procedure script translated to English:

1. **Welcome, preferred language and overview**
   (a) Hello and welcome to the EHRCtVRS test. What language are you more comfortable with, Portuguese or English?
   (b) I will give you a general overview of the test you are about to participate in:
      i. First, I will ask you to sign a form to have your informed consent to record and store your data;
      ii. After having your consent, I will need you to answer some questions regarding yourself;
      iii. If everything is ok, I will help you to put on the necessary hardware to proceed with the test;
      iv. The test will be divided into three parts:
         A. Instructions: Where you will learn how to perform the task.
         B. Task: Where you will perform the collaborative task with real robots.
         C. Questionnaire: Where you will answer some questions regarding the task you just performed.
2. **Informed consent form and the questionnaire**
3. **Assist with equipping the hardware**
   (a) Help to put on the pulse sensor [not too loose/tight fit]
   (b) Help to put on the HMD (help to adjust it/fit it correctly) and the handheld controllers
4. **[Record unity session (audio and video with OBS)]**
5. **Start the tests**
   (a) Practice stage

i. Grab the blue cube and insert it into the highlighted blue slot.
ii. After filling the column, grab a new block and insert it in the next column.
iii. When ready, press the green button to start the task with the actual robots.
iv. The robot is responsible for inserting the red cubes, and when you hear a double beep sound, it means the robot is starting to move.
v. A timer on the wall will time how long you take to finish each trial, so try to be as fast as possible.

(b) Collaborative assembly task with robots x5 test conditions

(c) Godspeed Questionnaire Series x5 test conditions

6. **Assist with removing the hardware**

   (a) Help to take off the pulse sensor, the controllers and the HMD

7. VR Presence questionnaire

8. Thank you for your time and participation