


Doctoral Dissertation
Doctoral Program in Computer Engineering (31^{th} cycle)

Human-Machine Interfaces for Service Robotics

Federica Bazzano

* * * * * *

Supervisor
Prof. Fabrizio Lamberti

Doctoral Examination Committee:
Prof. Hanan Samet, Referee, University of Maryland
Prof. Giuseppe Serra, Referee, Università di Udine
Prof. Andrea Bottino, Politecnico di Torino
Prof. Bill Kapralos, University of Ontario Institute of Technology
Prof. Filippo Sanfilippo, University of South-Eastern Norway

Politecnico di Torino
April 1, 2019

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . . . .

Federica Bazzano
Turin, April 1, 2019

Summary

Service robots are rapidly advancing toward becoming fully autonomous and skilled entities in a wide range of environments, and it is likely that more and more people will soon be interacting with robots in their everyday lives. As this happens, it is crucial that robots are designed to be easy to use and understand, reducing the need for people and environments to adapt to the robot. To this aim, in this thesis, different aspects of Human-Robot Interaction (HRI) field in service robotics applications were investigated by spanning the space between semi-autonomous and fully tele-controlled solutions. Despite the broadness of the explored field, through these activities it was possible to realize that HRIs could be classified according to two important correlated dimensions: the spatial proximity between humans and robots (remote or physically collocated) and the type of interaction (i.e., indirect or direct), thus defining the remote and collocated spatial proximity patterns. Research activities reported in this thesis were focused on exploring these patterns to detect the open problems as well as the arising challenges, in order to identify those interfaces that could be regarded as more appropriate and effective. The approach pursued in this dissertation was to explicitly focus on and directly consider HRI in the context of some application domains selected as representative examples of the above proximity categories in order to build robotic interfaces aimed to address the identified research issues. Hence, a set of interfaces were designed, implemented and evaluated by means of several user studies in order to learn from them how people tend to interact with robots depending on spatial proximity patterns, how robots can leverage these findings, and what are the implications for both users and HRI designers. By building on the obtained results, it was possible to identify and outline a set of user interface design requirements that could be considered to improve the HRI and make it more effective in both the remote and collocated spatial proximity patterns. The ultimate goal of this dissertation was to provide researchers with some design recommendations to be used as simple tools allowing them to apply the overall lessons learned without having to investigate the domain in depth again.

Acknowledgements

I wish to give thanks to all people who in one way or another have helped, supported and inspired me during the last three years.

First of all, I would like to thank my academic supervisor, Professor Fabrizio Lamberti for his guidance throughout my PhD study, but also for the challenging questions which made me think about my research from a different perspective. He inspired me with his vision, motivated me to develop my ideas and pushed me further than I thought I could go. His supervision has gone well beyond expectations since he prompts me to reflect on many stages of my life.

My sincere thanks also go to my industrial tutors from TIM, Dr. Marco Gaspardone and Dr. Roberto Antonini, for having supported the development of the research reported in this thesis, for their advice, encouragement, and belief in me from the very beginning of this work.

I would like to gratefully acknowledge Federica Marra and Giovanni Giannandrea, for being great friends, for being my family in Turin, for always listening to me, supporting me even in the dark moments of my life; for our crazy travels and for giving me memories full of cheerful laughter and smiles that I will never forget.

I would like to thank Evelina Di Corso for being a great essential friend, for sharing with me many beautiful and hard times, for always encouraging and supporting me with smiles and sweet words; for her belief in me even when I did not believe in myself.

Then, my heartfelt appreciation goes to Stefano Esposito, for his friendship, for his advice, and for teaching me to find my hidden courage by following his lead. Thanks also to Serhiy Avramenko, for our funny chats about Sicilian dialect and Russian language and for always providing me with different perspectives and point of views.

My appreciation also extends to my friends Francesco Savarese, Vincenzo Gallifuoco, and Carmine Sacco. We shared the excitement and fun of the undergraduate study years. Thank you for keeping me going on, for your support and feedback, for the long chats about the future and for being close friends even at a distance.

My thanks also go to my friend Sara Khan, for the fun lunch chats and reflection on research and healthy food, for making me see the world through her eyes and her tales of Japan and England, and for being a source of artistic inspiration.

In memory of my dear friend Andrea Guerriero, his story changed my life, made me reflect on the important things in life; I thank him for the long conversations about

life and the music of his guitar; for his freedom of thought and joy even in the hardest moments, for his unmistakable laugh that still pulls me a smile.

Moreover, I would like to thank Angelo Grimaldi, for being first a very good MS thesis student and for later becoming a very dear friend; for always being available, even when away, for a good chat to offload stress; for always smiling and laughing at my incessant complaining about something or other; for being so supportive and positive, and always providing great advice.

I would also like to express my acknowledge to colleagues and friends of the Lab 5 & Co., and in particular to Luca Venturini, Stefano Proto, Alberto Grand, Federico Manuri, Fabrizio Finocchiaro, Teo Montanaro, Giulietta Vasciavevo, Francesco Ventura, Eliana Pastor, Andrea Pasini, and Elena Daraio. I truly believe that this is one of the best labs in the world. The communal support and the family-like atmosphere have made my times in Politecnico di Torino an amazing and unforgettable chapter of my life.

In addition, I would like to thank my uncle Salvatore, my aunt Teresa, and my wonderful cousins and nephews, for always being close to me even at distance, for having believed in me and supported me at all times. My appreciation also goes to Antonella, Franco, Peppe, Giusy and Chiara, for always giving me a smile, great comfort and for always believing in me. My sincere thanks also go to my aunt Rosa, my uncle Francesco, my cousins Andrea and Valeria for being my second family, for encouraging and supporting me and for making me believe that the best is yet to come. I wish to express also my gratitude to my grandmother Maria, for being one of the strongest women I know, for teaching me never to give up and that in life the efforts and sacrifices are made to get what you want and sometimes you can make it out. For passing me the passion for cooking and for reminding me that my roots are an essential part of me.

Finally, the greatest thanks are due to my family. I am forever grateful to my parents, Giusy and Vincenzo, for their unending love and encouragement that supported me through the many difficult and dark times.

To my father, for having always stimulated my curiosity since childhood, for the "spells" (as you called them) that you taught me, when with a switch and some wire you showed me how to bring the sun inside a light bulb; for the hours spent together during my childhood in our "magic lab" in the attic, for teaching me how to solve problems using ingenuity and for the books you lent me; if today I am an engineer, it is mainly due to you.

To my brother, for being my first supporter, for always understanding me at the first glance, for making me feel important for his life, for being my point of reference, for growing up together and for having always urged me to become the best version of me.

To my mother, a great woman... graduated in love... at any time you took charge of my wounds on the knees and on the heart, taking care of my pain with your love... my teacher of life and my confidant... you have always known how to appease my worries and enhance my efforts... Although sometimes you felt defeated, I have always seen in your eyes that the battles of life failed to bend you... if I am strong, it is because you showed me how to pick myself up... your love is responsible for my not complying with

anything, for always wanting more and for appreciating life every day the sun rises... you have preached with your example the most important values that I possess today and that I will keep forever: to love with all my heart, to have a hand to give and another to receive, to be humble and to feel proud of myself... thanks to you today I know that my successes belong to me and that my dreams do not have an expiration date... and the smaller I felt, the more you protected me, filling my gaps of despair with your hugs... thanks for making my life more beautiful, for giving me your heart full of true love and for giving me the strength to reach the unattainable...

Without all of you, I would not be who I am today.

To my mother,

every bit of me is a little bit of you.

Contents

List of Tables

List of Figures

Glossary

Acronyms / Abbreviations

AA Adjustable Autonomy

AAS Adjustable Autonomy System

AI Artificial Intelligence

ANN Artificial Neural Network

API Application Programming Interface

AR Augmented Reality

ASQ After-Scenario Questionnaire

ASR Artifact Subspace Reconstruction

BCI Brain Computer Interface

BGE Blender Game Engine

BLE Bluetooth Low Energy

BN Bayesian Network

CD Cognitive Demand

COTS Commercial Off-The-Shelf

CRAB Connected Robotics Applications LaB

CRP Cloud Robotic Platform

CV Computer Vision

DIY Do It Yourself

DOF Degrees of Freedom

ECA Embodied Conversational Agents

ECG ElectroCardioGraphic

EE End-Effector

EEG ElectroEncephaloGraphic

EOG Electro-OculoGraphic

FFOV Fisheye Field of View

FIR Finite Impulse Response

FOV Field of View

FPS First Person Shooter

GCS Ground Control Station

GUI Graphical User Interface

HP Human Player

HR Heart Rate

HRI Human-Robot Interaction

HRV Heart Rate Variability

IFR International Federation of Robotics

IK Inverse Kinematics

IM Interactive Map

IR Infrared Radiation

JOL Joint Open Lab

LOA Level of Autonomy

MAVROS Micro Air Vehicle Robot Operating System

MDA Mechanical, Dynamic and Aesthetic

MORSE Modular Open Robots Simulation Engine

MR Mixed Reality

MW Mental Workload

$NASA-TLX$ NASA Task Load Index

NAU Nielsen Attributes of Usability

$NFOV$ Narrow Field of View

NTR Network Transmission Rate

NUI Natural User Interface

PID Proportional–Integral–Derivative

$PIRG$ Physically Interactive Robotic Game

PR Protoman Revenge

PRG Physical Robot with Gestures

PRM Physical Robot with Map

RBF Radial Basis Function

ROS Robot Operating System

RPG Role Play Game

$RSSI$ Received Signal Strength Indicator

RT Reaction Time

RTL Return To Launch

SA Situation Awareness

SAR Socially Assistive Robotics

$SART$ Situation Awareness Rating Technique

$SASSI$ Subjective Assessment of Speech System Interfaces

SDK Software Development Kit

$SERI$ Samsung Economic Research Institute

SIR Socially Interactive Robot

$SITL$ Software-In-The-Loop

$SLAM$ Simultaneous Localization and Mapping

SPA Sense, Plan, Act

SVM Support Vector Machine

TUI Tangible User Interface

UAV Unmanned Aerial Vehicle

UI User Interface

UN United Nations

$USAR$ Urban Search and Rescue

USE Usefulness, Satisfaction, and Ease of use

VR Virtual Reality

VRG Virtual Robot with Gestures

VRM Virtual Robot with Map

$VRUSE$ Virtual Reality Usability

$WFOV\&PT$ Wide Field of View and Pan-Tilt

YAH You Are Here

Chapter 1

Introduction

From robotic vacuum cleaners which are now a common appliance in over 6.1 million homes [1] to robot receptionists and autonomous drones that will deliver items around cities in the near future, service robots are playing an increasingly relevant role in the society. These and other uses of robots in recent years strongly suggest that robotic systems will continue to permeate humans' life, and new challenges will arise in the Human-Robot Interaction (HRI) field. Consequently, identifying appropriate human-robot interfaces suitable also for ordinary people is of paramount importance.

The interaction between humans and robots represents a rich and complex kind of communication, since it emerges from the confluence of internal and external factors that shape the interaction itself. For instance, in home automation settings or entertainment applications, where the robot and the human user are co-located and share the same environment, special attention has to be devoted to develop robots' social attitudes and making their interaction skills as much natural as possible [2]. In search and rescue applications, where robots explore unsafe areas in search of victims, Situation Awareness (SA) represents an important element to be taken into consideration [3]. When the robot and the human user do not share a common physical environment and the latter performs tasks "through" the robot (exploiting its senses), the Mental Workload (MW) arising from the execution of remote actions represents a critical factor [4]. Similarly, the role that human users play during the communication with robots as well as their background knowledge and/or experience in the task to be performed have an important impact on HRI [5, 6, 7].

The next paragraphs will explore the topics mentioned above, by identifying the dimensions shaping the HRI and the relationships occurring among them.

1.1 The Five Dimensions of HRI

HRI represents a very wide research domain, because of the complex interactions that occur between the involved parties. Identifying these parties as well as their mutual

relationships will allow to classify and delineate HRI categories.

In [8], Wang et al. identified a five-dimensional space through which HRI can be classified along one (or more than one) of the following axes: human, robot, world, task and time (Fig. 1.1).

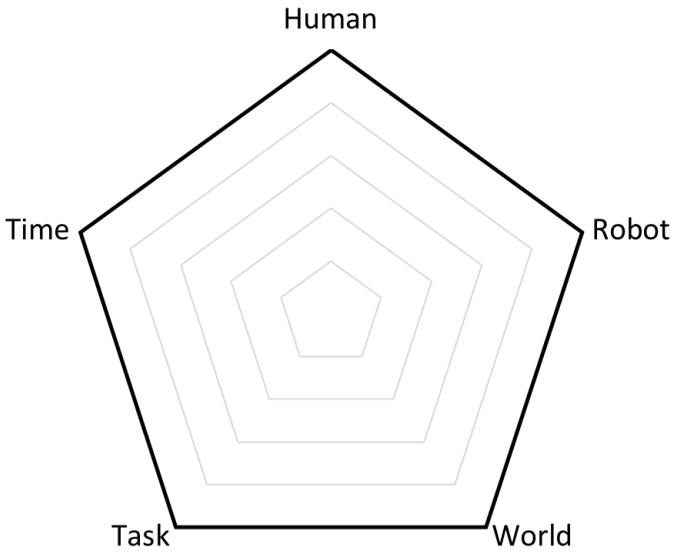


Figure 1.1: HRI dimensions according to [8].

1.1.1 Human

The human axis is considered a very important dimension in the HRI field. In fact, the role that human users play in human-robot teams as well as their proficiency arising from their personal skills and previous experience may influence the communication between people and robots. According to [8], humans can either take direct control of a robot or implicitly influence its operations by acting as decision-makers, communicators or inspectors. By spanning the spectrum of personal skills, a human user can be:

- novice;
- trained;
- expert.

As a matter of example, authors of [9] showed how different users may generally exhibit various levels of performance when controlling different robots and/or executing different tasks. A novice user may need to transfer the control to the robot even to

perform basic tasks in order to achieve the goal safely. On the contrary, a more experienced user may be able to easily perform these tasks without any robot's assistance, and may not even notice the lack of control in certain situations. Similarly, authors of [10] demonstrated the importance of developing HRI systems that meet users' needs and requirements while simultaneously developing the robotic systems.

1.1.2 Robot

Robotics is a broad discipline. The broadness of the field becomes evident by simply considering the different definitions that exist in the literature. For instance, according to the Robot Institute of America (1979) a robot is: "*a reprogrammable, multifunctional manipulator designed to move materials, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks.*" In contrast, the Webster dictionary defines a robot as: "*an automatic device that performs functions normally ascribed to humans or a machine in the form of a human.*" The United Nations (UN) together with the International Federation of Robotics (IFR), in their survey of robotics [11] define a robot as: "*an actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks.*" In this thesis, the latter definition will be used for referring to a robot.

From the mobility perspective, a robot can be static (stand in a fixed position) or mobile (able to move around within an environment). In terms of morphology, a robot can take three physical forms: *anthropomorphic* (human-like appearance), *zoomorphic* (animal-like appearance), and *functional* (neither human nor animal, but related to the robot's task) [12].

Concerning the autonomy dimension, which is defined in [11] as "*the ability to perform intended tasks based on current state and sensing, without human intervention*", ten Levels of Autonomy (LOAs) have been identified ranging from manual teleoperation to full autonomy [13]. Between these extremes, robots and human users can both have some level of control in order to achieve a given goal. This spectrum of control modes is shown in Figure 1.2.

Furthermore, when more than one robot is involved in a system, the robots form a group. According to the type of robots involved in the group, two team composition modes may be distinguished: *homogeneous* (all robots of the same type) and *heterogeneous* (different types of robots).

Depending on the area of application, two main categories can be identified: industrial and service robots [11]. The next two sections will present these categories in detail, by also providing the differences between them. A diagram of the all categorization criteria in the robot dimension is shown in Figure 1.3.

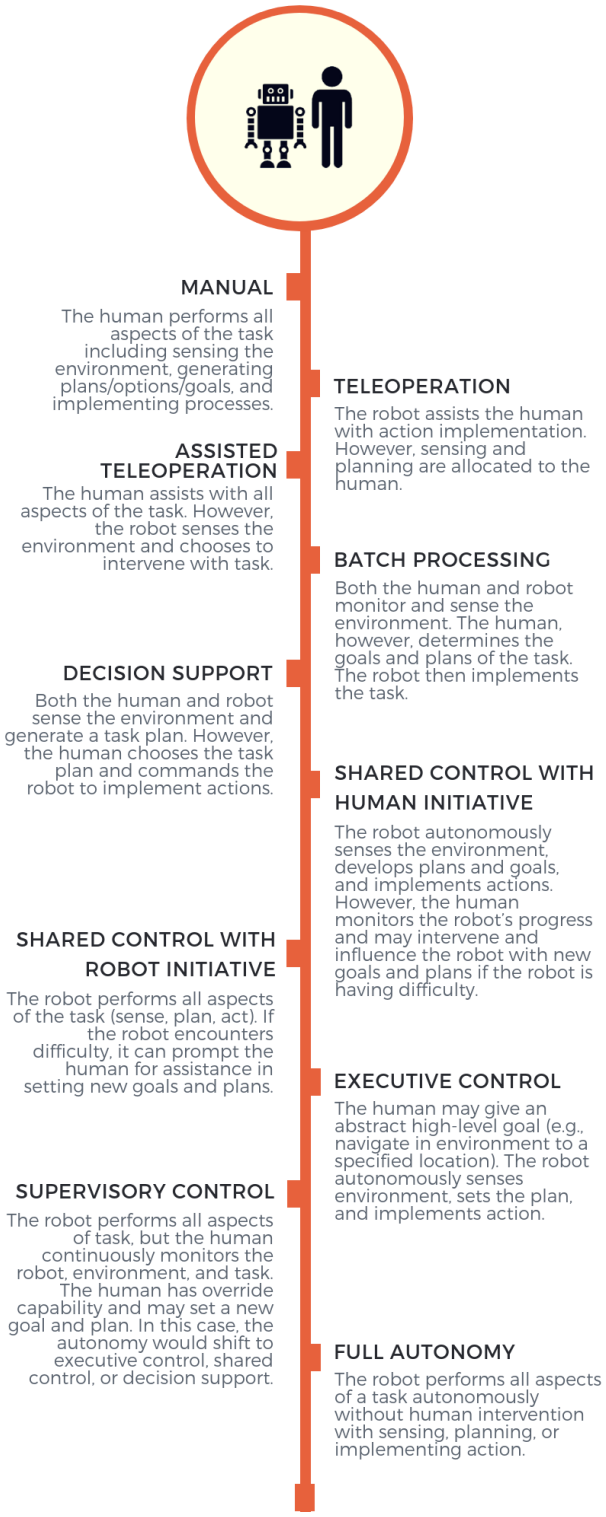


Figure 1.2: Levels of autonomy.

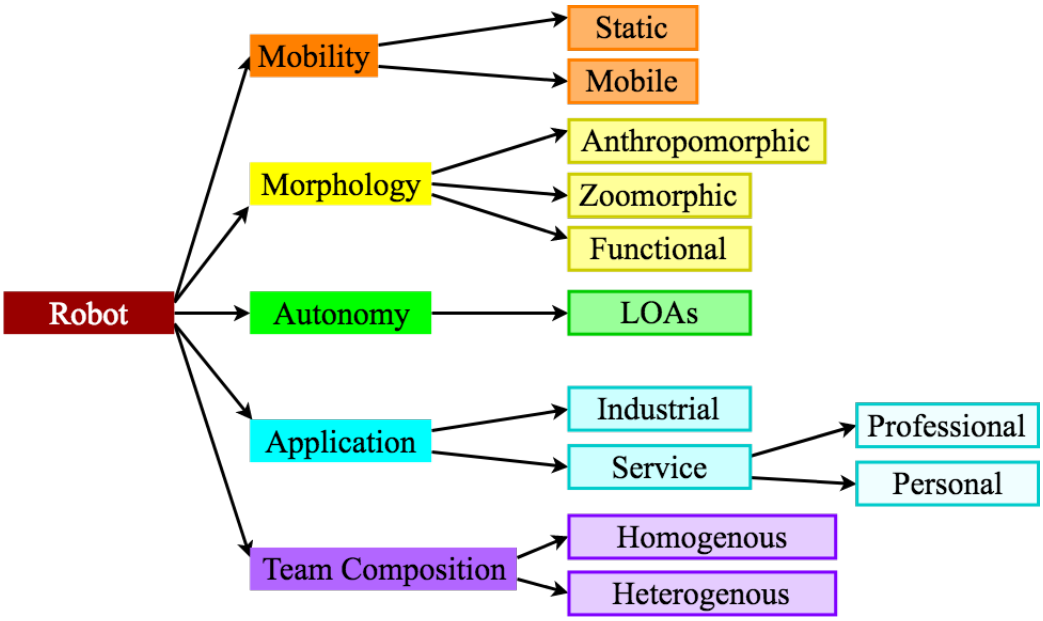


Figure 1.3: Categorization criteria of the robot dimension.

Industrial Robots

An industrial robot, very commonly a robotic arm (as illustrated in Fig. 1.4), is defined as "*an automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes, which can be either fixed in place or mobile for use in industrial automation applications*" [14].

Figure 1.4: Examples of industrial robots [14].

These types of robots can be classified according to different criteria, such as the mechanical structure (i.e., Degrees of Freedom or DOF), the application (manufacturing process) and the architecture (serial or parallel).

Depending on the mechanical structure, industrial robots belong to one of the following categories:

- cartesian robots, are robots that can do 3 translations using linear slides;

- scara robots, are robots that can do 3 translations plus a rotation around a vertical axis;
- 6-axis robots, are robots that can fully position their End-Effector (EE) in a given position (3 translations) and orientation (3 orientations);
- redundant robots, can also fully position their EE in a given position, but while 6-axis robots can only have one posture for one given EE position, redundant robots can accommodate a given EE position under different postures (like the human arm that can hold a fixed handle while moving the shoulder and elbow joints);
- dual-arm robots, are composed of two arms that can work together on a given workpiece. The type of movement is dictated by the arrangement of joints (placement and type) and linkages.

According to the type of applications, the robot will have to meet different requirements. As a matter of example, a painting robot will need to manage a small payload but a large movement range, whereas an assembly robot will have a small workspace but will have to be very precise and fast. Below, some types of applications are presented:

- welding robots, produce precise welds by managing different parameters such as power, wire feed and gas flow;
- material handling robots, capable of manipulating different products (from car doors to eggs) and picking and placing them from conveyor lines to packaging;
- palletizing robots, load corrugated cartons or other packaged items onto a pallet in a defined pattern;
- painting robots, capable to spray solvent-based paints and coatings to minimize human contact;
- assembly robot, used to lean industrial processes and expand production capabilities in the assembly line.

Considering the architecture, industrial robots can be classified in:

- serial robots, are composed of a series of joints and linkages that go from the base to the robot EE;
- parallel robots, also called spider robots, have concurrent prismatic or rotary joints creating a closed kinematic chain from the base to the EE and back to the base again (like many arms working together with the robot EE).

Service Robots

A service robot is defined as "*a robot that performs useful tasks for humans or equipment excluding industrial automation application*" [11]. According to this definition, service robots are envisioned to coexist with humans and to fulfill various kinds of tasks, as illustrated in Figure 1.5.

Figure 1.5: Examples of service robots [15].

In the last few years, a substantial progress in the field of service robots was recorded, and a variety of robotic systems designed to operate in environments populated by humans were developed. For instance, there are service robots deployed in hospitals [16], office buildings [17], department stores [18], museums [19], etc. As illustrated in the study performed by the Samsung Economic Research Institute (SERI)[1], the service robotics market is growing fast (Fig. 1.6).

Service robots are already able to perform various tasks (e.g., delivery, education, cleaning, etc.) and according to the type of these activities and the working environment, the following two categories can be identified:

- professional robots, operate in public or inaccessible settings (e.g., hospitals or nuclear waste sites) with the main function to manipulate or navigate their surrounding environment;

[1] http://www.seriworld.org

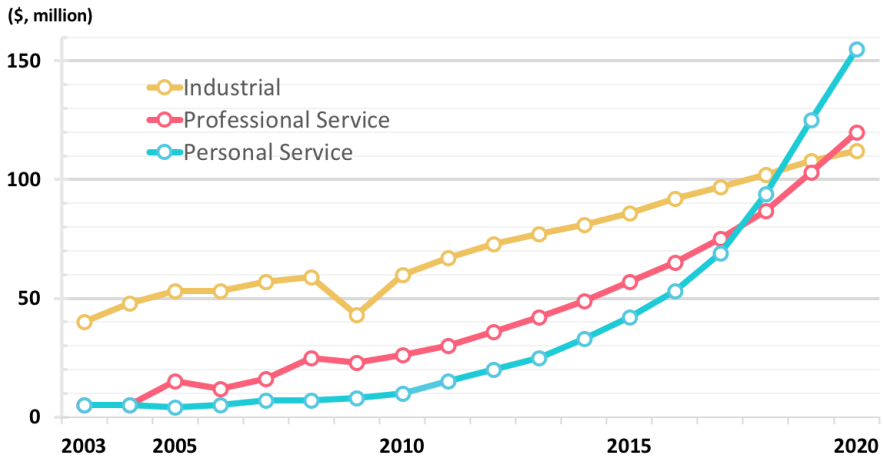


Figure 1.6: Worldwide robot industry market forecast.

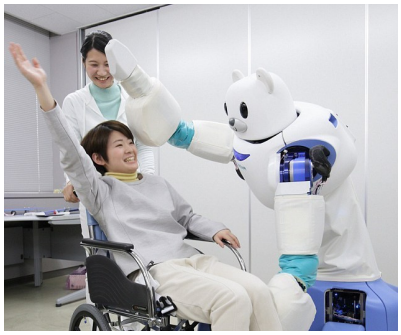
- personal robots, work in domestic and institutional settings with the main goal to assist or entertain people in everyday life.

Figures 1.7 and 1.8 depict some examples of existing robotic applications.

(a) Domestic [20].

(b) Home security and surveillance [21].

(c) Entertainment [22].



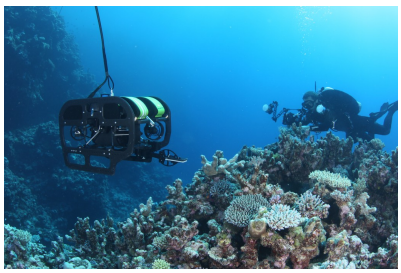
(d) Elderly assistance [23].

Figure 1.7: Examples of personal service robots applications.

(a) Field robotics [24].

(b) Professional cleaning [25].

(c) Search and rescue [26].



(d) Underwater [27].

Figure 1.8: Examples of professional service robots applications.

Industrial Robots vs Service Robots

Industrial robots are very different from service robots. In particular, industrial robots are specifically designed to be used in manufacturing operations with a specific number of tasks to be performed within workspaces that are generally isolated from human beings. Furthermore, due to operational requirements, industrial robots reach extreme operational accuracy, generally in the order of millimeters or fractions of millimeters. Reliability represents a key feature achieved thanks to their limited computing capacity and the establishment of precise tasks.

Service robots, differently from industrial robots, are devised to have an impact on human life and to satisfy requirements established by humans. To this aim, they are designed for being able to adapt to changes occurring in the environment and to accomplish tasks by exploiting real-time feedback. Flexibility represents a key aspect for allowing robots to adjust their behaviors and/or take decisions while interacting with human users. For this reason, the complexity of such a robot is much higher than an industrial robot in terms of designing and appearance, implemented intelligence and processing power required.

By moving from the above considerations, it becomes clear that industrial and service robots differ in many respects, also including the robotic market trends. In fact, as illustrated in Figure 1.6, the service robotics market is growing exponentially compared to the industrial one, strongly suggesting that service robots will continue to permeate

society in the near future by becoming ever more popular than industrial robots. Therefore, people will soon find themselves interacting with service robots in a wide variety of contexts and scenarios and new challenges will arise in HRI field. For this reason, many aspects have to be considered, ranging from the technological advances necessary to actually build appropriate hardware and software systems, to the more complex aspects arising when the results of technological developments are encountered by their potential users. Hence, many interesting questions may be posed, including for instance, how should service robots move around, communicate and adapt to both the environment and human users? How should common people interact with, work with, and understand service robots? What kind of communication paradigm is needed and better suited to allow humans and robots interacting in appropriate and effective ways? How can the service robots integrate into people's social spaces?

With the aim of addressing the above questions and the interest to study the research field in collaboration with TIM (the largest Italian telecommunication provider) JOL Connected Robotics Applications LaB[2] (CRAB), the focus of this thesis will be on service robots.

1.1.3 Task

A task in HRI field can be described as any activity or action that a human user has to accomplish within an environment through a robotic system interface. According to the authors of [28], five main tasks can be identified in HRI domain, as illustrated in Figure 1.9.

Navigation is one of the major tasks performed in robotics, and is composed of three related subtasks: localization, wayfinding and movement. A second task linked to navigation is related to the perception of the environment. The focus of this task is on perceiving and processing sensor data to understand the surrounding world. Management is defined as the activity to coordinate the actions of multiple robots, either acting as a group or independently from each other, whereas manipulation consists of all the actions performed to interact with the environment (e.g., grasping, pushing, etc.). Lastly, the social task encompasses all the actions requiring social and interaction skills as key features.

Many robotics applications can be described by combining the aforementioned tasks. Considering the five most representative applications in the HRI domain according to [29], i.e., urban search and rescue (USAR), personal assistance, museum guide, fleets, and physical therapy, they can be described within the task dimension as illustrated in Figure 1.10. As a matter of example, the USAR task consists of carrying out navigation, perception and manipulation activities without any social involvement. On the contrary, physical therapy robots working closely with human beings should be able to

[2]https://www.telecomitalia.com/tit/en/innovazione/archivio/jol-crab-torino.html

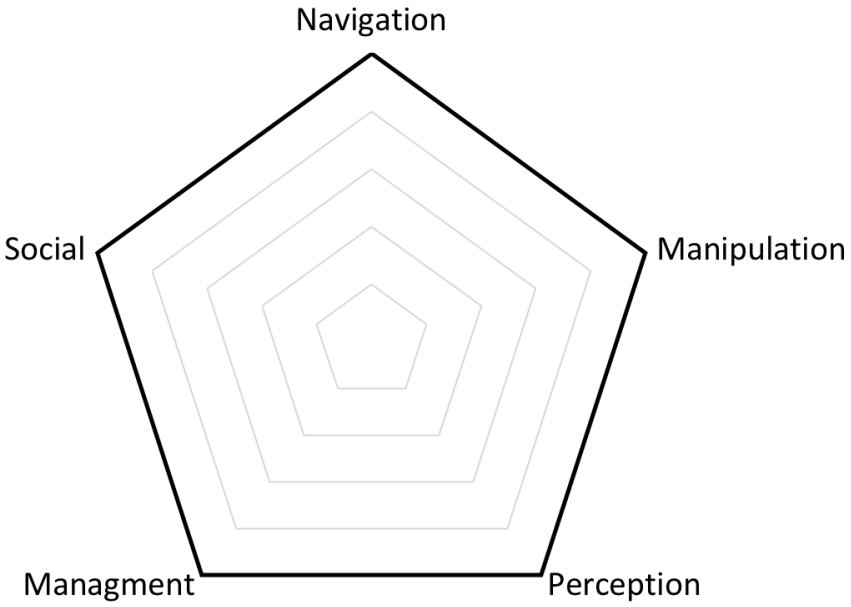


Figure 1.9: Components of the task dimension according to [28].

interpret and respond to users' social stimuli.

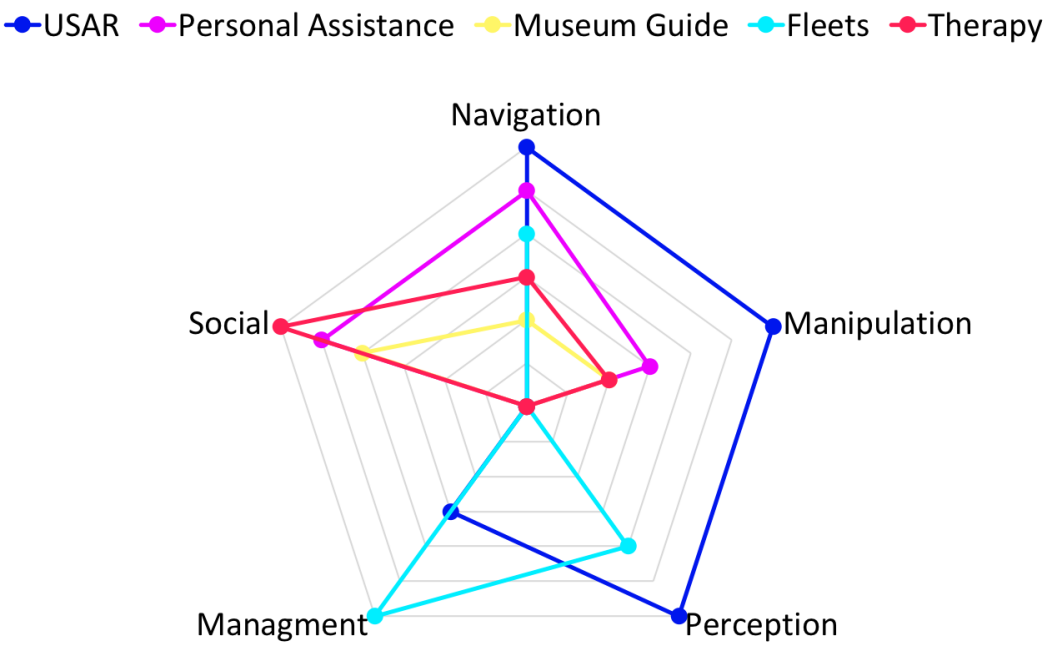


Figure 1.10: Robotics applications described through the five components of the task dimension [29].

Moreover, based on external features, a task can be classified according to its level of risk as high, medium, and low urgency; based on its frequency, it can then be unique, periodical, and routine [30].

1.1.4 World

The world, intended as the robots' workspace, represents a very important HRI dimension, strictly correlated with the task dimension. Performing a task in different environments, such as navigating in an office, space or agricultural scenario involves different challenges and, thus, different requirements related to the complexity of the world. The difficulty of the environment can be estimated in different ways. For instance, in [31, 32], the arrangement of obstacles and the terrain traversability were respectively used. According to environmental variables (e.g., weather, temperature), the world can be, for example, hot or cold or dark or bright, etc. Depending on the purpose, the world may be indoor or outdoor, open or closed, static or dynamic, etc.

1.1.5 The Relational Spaces

The relational spaces involving the four dimensions (human, robot, task, world) in the time domain (the fifth dimension), define the interactions in HRI. In the next paragraphs, the three main relational spaces are presented.

Human-Robot

This bi-dimensional space, which is obtained by ignoring the task and world dimensions, describes the relationship between humans and robots in terms of numeric ratio, i.e., the number of humans over the number of robots. This ratio defines the control and collaboration paradigms between human users and robots both at team and individual level [12]. Figure 1.11 illustrates the possible combinations with a maximum of two robots and two humans (the same concepts holding for "two" are valid for "many"). The flow of information is displayed through bidirectional arrows, whereas the collaboration and coordination of team members is depicted through the bracket symbol.

Figure 1.11.a depicts the one-to-one situation in which a human user provides commands to a robot, which sends back a feedback to the human. In Figure 1.11.b, the one-to-team case is depicted with a human user controlling a team of robots. The team members then coordinate among each other for identifying the robot(s) responsible to perform the task. Figure 1.11.c illustrates the one-to-many situation with a human user controlling two robots working independently. The team-to-one case is shown in Figure 1.11.d, where human users grouped in a team collaborate to issue a command to the robot. Figure 1.11.e describes the many-to-one situation, where users within a team act independently and issue different commands to a single robot. The team-to-team relation, which is shown in Figure 1.11.f, depicts a team of human users controlling a team

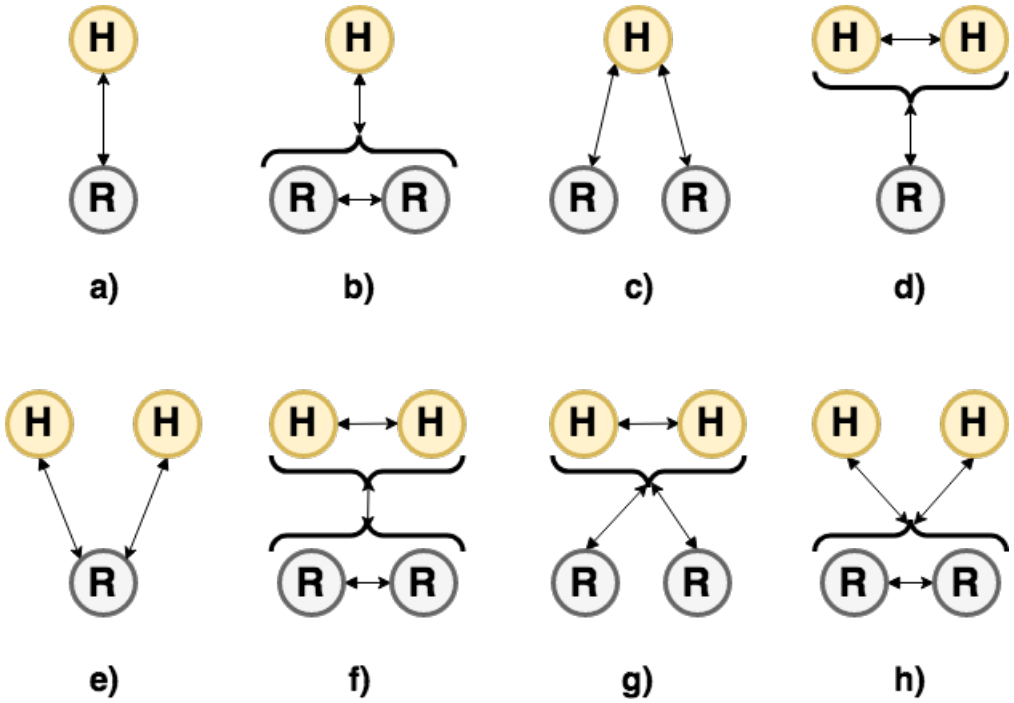


Figure 1.11: Relationship between human users and robots in terms of numeric ratio, where the human user is displayed through the letter "H", whereas the robot is depicted by the letter "R".

of robots. The humans collaborate to issue a command while the robotic team members cooperate to select the robot(s) in charge for performing the task. In Figure 1.11.g, a team of human users provides commands to different independent robots thus depicting the team-to-many situation. Lastly, Figure 1.11.h depicts the many-to-team situation, where human users do not collaborate and send different commands to a team of robots. The robots prioritize the commands and distribute them among themselves before carrying them out.

Human-Robot-Task

Considering not only the ratio of humans to robots but also the task they have to perform, a set of human roles can be identified as shown in Figure 1.12 [33]. Roles are described below.

- The *supervisor* is a person who monitors and controls the overall situation. In this case, the *supervisor* is responsible for evaluating the actions based on the perception of the system, and for ensuring that these actions will lead to the achievement of higher-level goals.
- The *operator*'s role is to directly interact with the robot. This interaction may

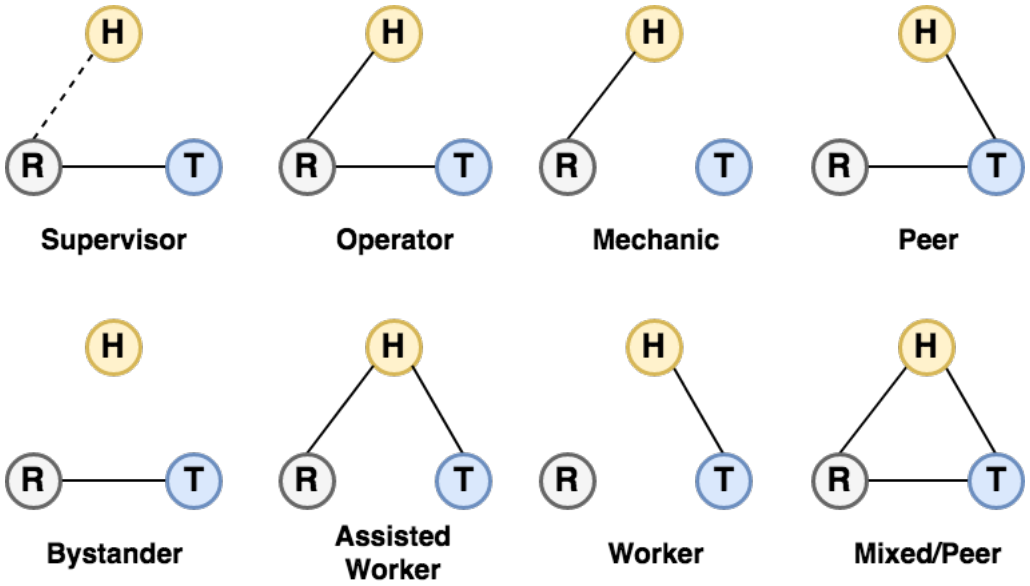


Figure 1.12: Human roles, where the human user is displayed through the letter "H", the robot is depicted by the letter "R" and the task with the letter "T".

be expressed as a simple change in the robot's parameters, or through a direct manipulation of its steering and pose.

- A *mechanic* assists in the resolution of hardware and software issues that cannot be resolved by the operator. This implies remotely fixing low-level software problems, monitoring hardware, and physically replacing electronic and mechanical parts on-site.
- The *peer/teammate* represents a component of a human-robot team working together and cooperating with the robot, usually in a shared environment, to reach a goal.
- The *bystander* is a person who does not control the robot directly, but affects its actions and the way it performs the task. For example, a person walking in a room with a robotic vacuum cleaner affects the robot's action which needs to be able to avoid the human user safely.
- The *assisted worker* is a person who executes a task with the assistance of a robotic system.
- The *worker* role represents the same situation depicted in the *bystander* case, but with the robot implicitly affecting the human user.
- The *mixed/peer* describes the situation in which the human, the robot and the task create a loop of interaction and each component directly interacts with the

others.

Human-Robot-Task-World-Time

In this relational space, the four dimensions (human, robot, task, world) in the spectrum of time, constitute the interactions in HRI field.

As stated by Wang et al. in [8], HRI can be divided into four categories based on spatial proximity (Fig. 1.13).

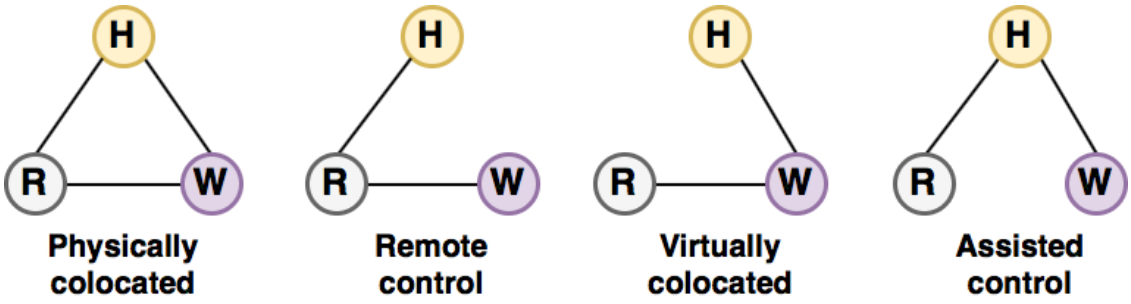


Figure 1.13: Proximity patterns.

The category named "*physically colocated*" shows the situation where the human user and the robot are physically in the same workspace; hence, both of them can directly interact with each other and with the world. For example, a robotic courier would interact with human users in such a way that both the robot and the users change the environment.

In the category named "*remote control*", the robot and the human are in different workspaces while all the changes (e.g., related to the environment or to the robot) happen in the robot's workspace. The human perceives the world through the robot's sensors and affects the world via the robot's interaction with the world. For example, a telepresence robot may be exploited in a remote teleconferencing environment when a human being cannot physically attend the meeting, though this person could remotely control the robot in the environment.

The third category named "*virtually colocated*", describes the situations in which the human and the robot can directly affect the world without any direct interaction. For example, a human can play a robotic game (game involving a physical robot) and the human and the robot affect each other only in the world created for the game.

The last category, named "*assisted control*", is similar to the situation depicted in the "*remote control*" configuration, but now the robot affects the world only via the human's interaction with the workplace. A human user interacting with a robot equipped with a decision support LOA (Section 1.1.2) represents an example of this type of situation.

According to the taxonomy defined in [34], HRI can be divided in other four categories depending on whether the interaction occurs in the same place (or not) and at the same time (or not), as illustrated in Figure 1.14.

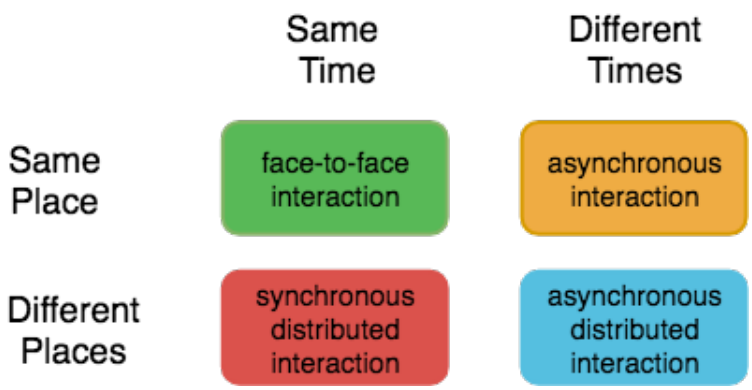


Figure 1.14: HRI categories according to temporal/spatial constraints [34].

As a matter of example, an assistant robot colocated with a human and controlled by him or her to perform a given task falls into the synchronous and colocated category. A telepresence robot located in a different environment with respect to the human user who remotely controls it is an example of the synchronous and non-colocated category. An example of asynchronous and colocated category is represented by robots performing tasks in manufacturing applications located in the same place as the humans who carry out the same tasks but in a different time. Lastly, a space robot may be an example of the asynchronous and non-colocated category, because it is an autonomous robot performing tasks in a different place with respect to humans' place.

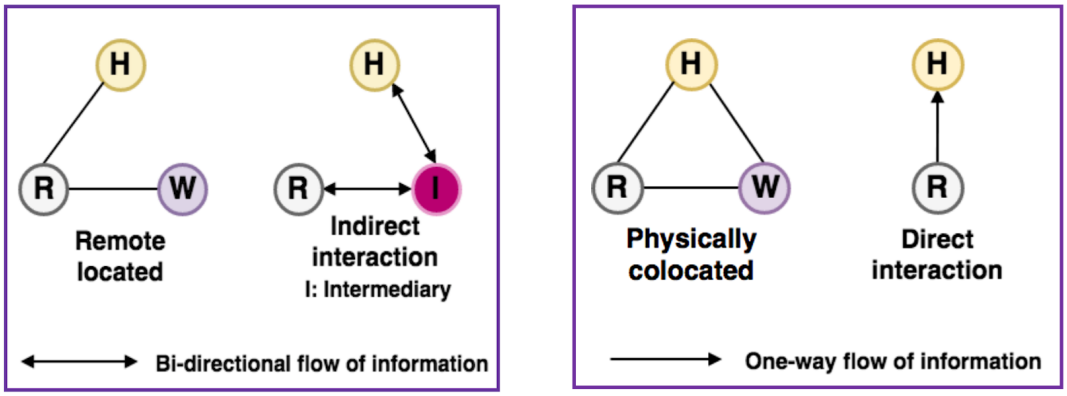
1.2 Human-Robot Interfaces

Robots, like many other technological machines, need user interfaces for letting people communicate with them. Though, these interfaces differ depending on whether they are used for industrial or service robots.

In industrial robotics, robots perform tasks within workspaces that are generally isolated from human beings. Hence, the interaction between humans and robots is limited and mostly occurs through programming or simulation languages.

In service robotics, the interaction is richer and, as illustrated in Figure 1.15, can be classified in two different categories based on humans' roles, spatial proximity and flow of information [35].

In "*indirect interaction*", the human controls and/or supervises the robot to some extent. The interaction mostly occurs through a Graphical User Interface (GUI) and/or a control device (e.g., keyboard, mouse, etc.), and takes place within the remote proximity pattern. The flow of information can be considered as one-way, that is, the human sends control commands to the robot which communicates back some feedback to the human about its state, sensors, etc. In this category, two prominent styles of interfaces tend to be prevalent [36]:



(a) Remote pattern (b) Colocated pattern

Figure 1.15: HRI categories.

- *supervisory control interfaces*, that commonly use a map which is either predefined or built as the environment is explored; these types of interfaces are favored for their SA, the ability to control multiple robots and varying robot's LOAs; the human takes on the role of supervisor or defines tasks for the robot to perform at a high level [37];
- *teleoperation interfaces*, which focus more directly on a video feed provided by the robot's onboard sensors and enable the operator to move the robot and interact with the world based on such feedback; the human takes on the role of operator by mastering the robot with one or more control devices [38].

In "*direct interaction*", the communication with colocated robots asks for interaction paradigms different from those required in indirect interaction. The human-robot proximity affects the nature of this information exchange. The flow of communication can be considered as bi-directional: information is transferred between human users and robots in both directions, and the robots and the users interact as peers. This interaction occurs via sensor-mediated interfaces compared to desktop and GUI-based interfaces used in indirect interaction. Humans do not need devices to be assisted by robots: they can interact with them via human-like communication paradigms, such as speech, gestures, and gaze. These interfaces are generally referred to as Natural User Interfaces (NUIs) [39], since they facilitate intuitive interactions and require minimal or no users' special training [40].

A prominent approach in designing robotic interfaces for colocated robot operation has been to use graphics content displayed onto the physical world (via Mixed Reality, or MR) as a means to highlight the commands given and to provide feedback on the commands' progress [41, 42], or to use Tangible User Interfaces (TUIs) as a means to map gestures to robot commands, for both robot steering and for the robot's pose definition [43].

1.3 Open Problems

As depicted in the previous sections, HRI is a very broad field playing a pivotal role in service robotics [29]. In this research area, the interaction techniques are various and defined by the sub-spaces within the five HRI dimensions.

In this thesis, the focus will be on HRI interfaces used in service robotics applications (as discussed in Section 1.1.2) lying in the space between remote and colocated scenarios (Fig. 1.15) and their implications on HRI domain. In fact, what is interesting to study within these two scenarios is how the spatial proximity may affect the nature of the interaction. For instance, the interaction with a mobile robot in a remote scenario is generally referred to as "*indirect interaction*" (Section 1.2) and may assume the form of teleoperation or supervisory control; the interaction with a mobile robot in a colocated scenario is generally referred to as "*direct interaction*" and may assume the form of assistive robotics. Similarly, the remote/indirect interaction with a robot equipped with a physical arm is often referred to as tele-manipulation, whereas the proximate/direct interaction with a robot equipped with a physical arm is generally referred to as natural interaction.

By leveraging the above examples, it can be observed that depending on the considered scenario, different interaction paradigms, and consequently different users' needs and robots' functionalities (as discussed in Section 1.2) can be distinguished, as well as different problems and challenges can be identified. Hence, the primary purpose of this section will be to identify and characterize the open problems as well as the arising challenges in the aforementioned space in order to define those interfaces that could be regarded as appropriate and effective.

1.3.1 Remote Spatial Proximity Pattern

By focusing on remote spatial proximity pattern (Fig. 1.16), a general problem arises from the distance between the human and the robot, which do not share a common physical environment.

The human operator must control and/or monitor the robot remotely by maintaining a continuous real-time awareness of the robot's state and surrounding areas. The human perceives the remote environment through the robot's sensors and its feedback data. Examples of data include robot's speed, video feed, robot's distance from obstacles, etc. As a matter of example, the common scenario of controlling a robot via a pan-tilt camera could be considered; it can be quite easy for the human user to forget that the camera may not be pointing forward and, thus, to provide wrong direction commands to the robot based on the visual camera feed. This general issue has been referred to as a problem of SA between the controller and robot. The user needs to be aware of the overall robot's status, and the robot needs awareness of the person's perspective in order to properly interpret commands. In [44], SA was defined as: "*the understanding that the human has of the location, activities, status, and surroundings, of the robot; and the*

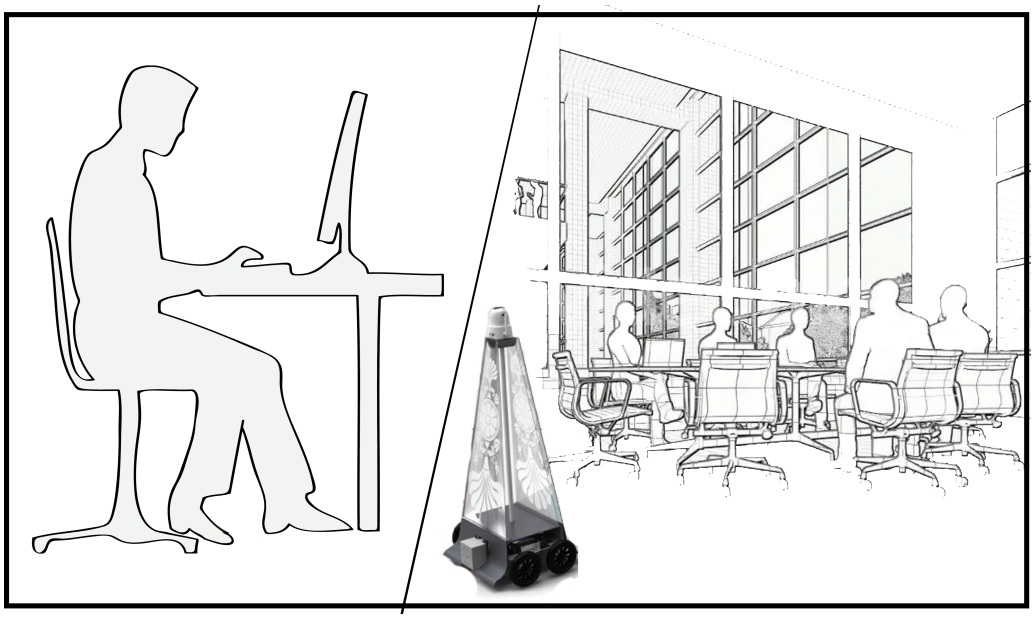


Figure 1.16: Remote spatial proximity concept.

knowledge that the robot has of the human's commands necessary to direct its activities and the constraints under which it must operate".

Another general problem arises from operating a robot from distance. The human user may both give instructions to the robot to make it perform a given task or monitor its actions in the remote environment. The operator's mental strain resulting from performing these tasks in possibly complex operational conditions is generally termed MW or Cognitive Demand (CD) [4].

1.3.2 Colocated Spatial Proximity Pattern

Concerning colocated spatial proximity pattern (Fig. 1.17), general problems as well as challenges arise from direct interaction between humans and robots as well as from humans' expectations.

Service robots, sharing a common environment with humans, are expected to be able to interact naturally with them, to be observant of their presence, recognize what they are doing and react appropriately to stimuli coming from them, such as gaze, voice or gestures. Consequently, user evaluations (e.g., in terms of satisfaction, ease of interaction, usefulness, etc.) define the quality of an interactive system, that is generally referred to as *usability* [45]. Furthermore, robots should be able to capture humans' attention and improve their *engagement*, defined as the process by which individuals initiate, maintain and terminate their perceived connection during an interaction [46].

Another general issue consists of how these robots are perceived and potentially accepted by ordinary people when integrated in the society. It is important to study

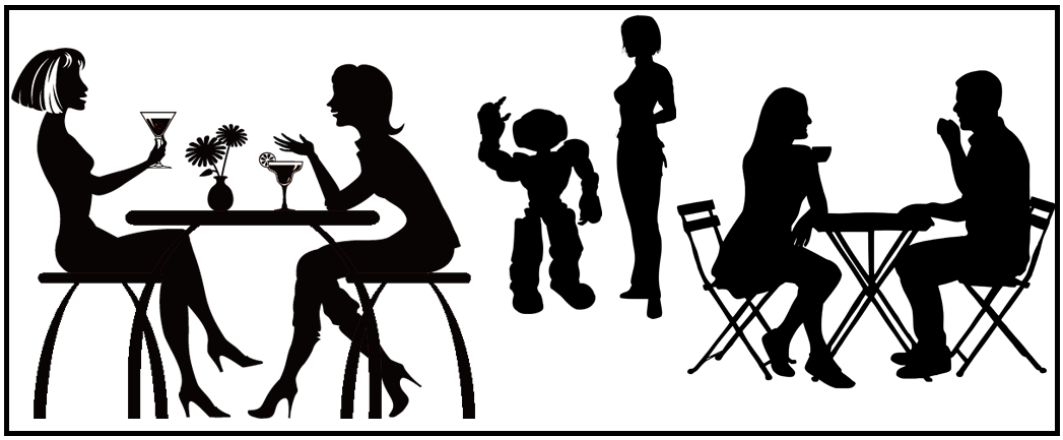


Figure 1.17: Colocated spatial proximity concept.

whether robots' behavior is appropriate according to the role they play as well as to investigate whether the execution of their job positively supports human beings in carrying out their activities. The extent to which a system satisfies users' needs by performing the expected tasks as well as in terms of physical design and system's functionalities is generally referred to as *user acceptability* [47].

A summary of the open problems identified in the domain investigated in this thesis is shown in Figure 1.18.

Issues in Colocated Pattern

- robot's understanding of humans' presence and stimuli;
- robot's ability to interact naturally with humans;
- how these robots are perceived by humans and whether they are accepted.

Issues in Remote Pattern

- poor human's Situation Awareness about the location, activities, status, and surroundings of the robot;
- high human user's Cognitive Demand for operating robot from distance.

Figure 1.18: Research issues in colocated and remote proximity patterns.

1.4 Thesis Goal

The research goal of this dissertation is to propose solutions to the open research issues discussed in Section 1.3 by first understanding strategies already pursued to tackle them and, then, identifying, designing and developing new HRI approaches capable to advance the state of the art in the field. In particular, the focus will be on devising HRI interfaces that could be regarded as appropriate and effective means for letting end-users interacting with robots both in remote and colocated spatial proximity patterns.

Service robotics is a broad domain that includes various types of applications ranging from remote and local HRI. In this thesis, some scenarios were selected and explored as representative examples.

Specifically, with the aim of addressing problems identified within the remote spatial proximity pattern and, at the same time, investigating the two different approaches that are prevalent in this model of interaction (i.e., *teleoperation* and *supervisory control interfaces* as discussed in Paragraph 1.2), two different application domains were explored:

- robotic telepresence;
- robotic aerial traffic monitoring.

The first application domain focuses on *teleoperation interfaces* (Paragraph 1.2) and human beings in the role of *operators* (Paragraph 1.1.5) and their needs. It aims to explore and study other possible interaction paradigms for robotic applications different from those investigated in previous works, able to improve the user experience in remote controlling a robot when a video feed is exploited. Specifically, operating a mobile robot from distance (generally via a front-facing camera) may be mentally challenging for the users when they do not possess a proper awareness of the environment (Paragraph 1.3.1). To this aim, a user study was performed to investigate through a comparative analysis, the usability of two major approaches used today for controlling telepresence robots, i.e., Keyboard Teleoperation and Point-and-Click Video Navigation. In particular, a telepresence framework implementing the above navigation modalities plus a combination of the two as third navigation modality was developed in this thesis. New modules were added to an existing cloud robotics platform for robotic telepresence applications and different telepresence interfaces merging the above modalities with a mixed map & video-based teleoperation system were devised and developed. Moreover, the impact on users' performance associated with the introduction of different augmented fields of view (FOVs) and/or a pan-tilt camera was also investigated. All of the above configurations were added to the developed interfaces, implemented as new modules into the telepresence framework and applied to the operation of both a real and a simulated mobile telepresence robot created at TIM JOL CRAB.

The second application domain focuses on *supervisory control interfaces* and human beings in the role of *supervisors* in an Unmanned Aerial Vehicle (UAV) traffic monitoring scenario. In this context, the human operators act as UAVs controllers (a professional profile that is expected to be required in the future) to deal with a possibly huge amount of flying objects. Consequently, supervisory systems including human in the control loop methods are required to both monitor UAV operations but also to assist UAVs controllers by predicting their MW changes when the number of UAVs to be monitored significantly increases. To this aim, a simulation framework for reproducing swarms of autonomous drones flying in urban environments and an Adjustable Autonomy (AA) system able to flexibly support different LOAs were designed and developed in this

thesis. The AA system is able to discriminate situations in which operators' abilities are sufficient to perform UAV supervision tasks from situations in which suggestions or automatic interventions may be required. An MW prediction model based on operators' cognitive demand was exploited and integrated in the framework to allow the AA system to infer the appropriate LOA accordingly. Different MW assessment techniques were used to train the prediction model, namely, subjective, physiological and performance–based. In particular, the NASA-Task Load Index or NASA TLX questionnaire was used as a subjective measure, the electroencephalographic (EEG) signals as physiological technique and users' performance in accomplishing the given task as an objective or performance-based measure. Furthermore, two different learning and classification methods, namely, a Bayesian Network (BN) classifier and a Support Vector Machine (SVM), were exploited and integrated in the framework as decision-making modules to both build the MW prediction model and to evaluate the classified patterns from the point of view of accuracy. Moreover, different supervisory control interfaces, one for each LOA considered in the thesis, were also devised and developed.

Concerning the colocated proximity pattern and with the aim of investigating the different interaction paradigms with proximate robots and users' perception about robotic systems, two different application domains were selected as representative examples:

- robotic gaming;
- assistive robotics.

The first application domain aims to investigate and evaluate human interaction with autonomous or semi-autonomous co-located robots during recreational and entertainment activities. To this purpose, two robotic games were developed. The first game was aimed to investigate how the introduction of emotional features (like happiness, anger and frustration) and autonomous behaviors in a robotic gaming scenario leveraging drones could affect users' experience and their engagement. In particular, a game architecture was designed and developed in this thesis implementing both the game logic and the drone's autonomous behaviors including object detection, localization & motion control, and emotional encoding. The second game was meant to show, through an autonomous toy robot and a set of tangible interfaces, how to favor an engaging interaction between players and real/virtual game elements in a floor-projected MR gaming concept. In particular, a game architecture implementing the robot's autonomous behaviors, the interaction with the tangible interfaces and the devised game concept was developed in this thesis.

Regarding the assistive robotics domain, two use cases involving two different robots, i.e., a mobile robotic assistant and a socially interactive robotic assistant, were considered. In particular, the first use case was aimed to study and assess the natural interaction paradigms between a human user and a robotic assistant within the specific context represented by a robotics-enabled office scenario. In particular, an immersive

Virtual Reality (VR)-based simulation framework was developed and exploited to guide the design and implementation of two different user interfaces. The proposed interfaces, called Augmented Reality (AR) and non-AR interfaces, were devised and developed to combine different natural interaction means (e.g., speech, gaze, arm gestures) and enable the execution of three possible robotic office tasks that the robot could be involved into. Specifically, the tasks differ in the way the selection interactions (required to select the robot or specify destination coordinates) are gathered and in the way information useful for controlling the robot (e.g., commands available, current status, etc.) are presented to the user. The second use case was aimed to investigate users' acceptability of a socially interactive humanoid robot in accomplishing tasks able to assist human users in carrying out their activities. In particular, the specific use case of the robotic receptionist was addressed to study the aforementioned considerations, given its implications in human living environments with the role of providing people with useful directions towards a place of interest through social interaction. In this context, a user study was performed to investigate how robots featuring or not a human-like appearance and endowed or not with social behaviors can be perceived by human users and can impact on their performance. Specifically, three different receptionist systems were considered in this thesis, namely, a physical robot, a virtual agent and an interactive audio-map. By digging more in detail, for the physical robotic receptionist new modules were developed and added to an existing robotic framework to implement the receptionist system considered, whereas the virtual robotic receptionist framework as well as the interactive audio-map were designed and developed in this thesis. It is worth noting that in the two applications domains described above, the human being plays the role of *peer/teammate* as defined in Paragraph 1.1.5.

In summary, within the whole thesis, a set of frameworks and interfaces were designed, implemented and evaluated by means of several user studies with the aim of learning from them how human users tend to interact with robots depending on spatial proximity patterns, how robots can leverage these findings, and what are the implications for both users and HRI designers. The performed analyses and the obtained results allowed to identify and outline a set of user interface design requirements and recommendations that could be considered and used by researchers and designers to improve the HRI, make it more effective in both the remote and colocated spatial proximity patterns by exploiting the overall lessons learned without having to investigate the domain in depth again.

1.5 Thesis Organization

The remainder of this dissertation is organized as follows. Chapters 2 and 3 present the application domains explored to address the research issues identified in the remote and colocated spatial proximity pattern, respectively, by investigating the related works defining the state of the art, providing details of the solutions proposed to overcome

the above problems (i.e., various interface designs, implementations, and evaluations) and illustrating the obtained results. Chapter 4 presents a discussion of these results in order to illustrate their significance and comprehend the overall contributions in the two considered proximity patterns. Chapter 5 concludes the thesis by providing the lessons learned and discusses future developments.

Chapter 2

Interaction in Remote Spatial Proximity Pattern

Part of the work described in this chapter has been previously published in [48, 49, 50, 51].

In the remote proximity pattern, two different interaction approaches have been identified, namely, the *teleoperation* and the *supervisory control interfaces* (Section 1.2). As discussed in Section 1.3, these types of interfaces are affected by problems such as the poor *operators*' or *supervisors*' SA and their high CD.

The purpose of this chapter is, on the one hand, to explore and study the research related to the works defining the state of the art of the two application domains considered in this thesis as representative examples of remote HRI, i.e., robotic telepresence and robotic aerial traffic monitoring (Section 1.4). On the other hand, chapter's goal is to exploit the acquired knowledge to develop and implement proper frameworks dealing with concrete use cases in order to identify and develop suitable HRI interfaces for addressing problems arising from the considered pattern.

2.1 Robotic Telepresence

The robotic telepresence application domain was selected as a possible representative example of applications featuring a *teleoperation interface* with the human user in the role of *operator* in remote spatial proximity pattern. The next sections will illustrate relevant works pertaining this research area as well as the robotic platforms considered in this domain. Moreover, the telepresence framework, the UIs, the methodology adopted to perform the experimental tests and obtained results will be also described.

2.1.1 Background

In a pioneering paper [52], the American scientist Marvin Minsky introduced for the first time the term "telepresence" as an adapted version of the older concept of "teleoperation" by focusing on letting people act and feel as they were physically present at a different location. When this experience is achieved through the use of a robot, it is generally referred to as "robotic telepresence" [53]. Well-known examples are teleconferencing, virtual tourism, health care, education, to name a few [54].

In this scenario, a human in the role of *operator* is located at a distance from the robot, and explores the environment based on feedback provided by the robot [55]. In these cases, human perceptual processes are disjoint from the physical world; therefore, the only sensory stimuli are represented by feedback information. In the remote environment, human actions may be compromised by the lack of or by feedback misperception, making even simple tasks incredibly difficult to manage [56]. The human's interpretation of this information should overcome this "decoupling effect" [57].

An effective design of teleoperation interfaces requires to identify the key elements that can improve the operator's ability to correctly perceive and understand the above information (SA), as well as lower the cognitive effort (CD) arising from the execution of the remote tasks while keeping the interaction with the robot as simple as possible [58].

Several guidelines for designing HRI techniques can be derived from the number and type of these key elements. For instance, information such as the position and orientation of the robot as well as the video feed and information on distance from obstacles represent key factors affecting humans' SA [59], introducing some pros and cons. As a matter of example, video information is largely exploited by human users to navigate remote environments via telepresence robots [55]; however, navigation performance can be strongly influenced by the way in which this information is presented as well as by the orientation, the point of view and the amount of information detected from the FOV of the robot's camera [56]. In fact, when a robot detects the environment through the use of a camera, the so-called "keyhole" effect is produced: only a portion of the remote environment is actually visible by the human operator – compared to direct vision – thus requiring him or her further effort to interpret it [60]. Furthermore, position and orientation represent useful information providing human operators with references for navigation, but communication delays and bandwidth limitations could introduce critical misalignments [36].

Based on key elements mentioned above, telepresence interfaces are often classified into two main categories: map-centric and video-centric [61]. In a map-centric interface, the map represents the most important feedback source the operator can rely on to supervise navigation. A large area of the operator's display is covered by the map showing all relevant information on it. In contrast, in a video-centric interface, the flow from the camera mounted on the robot represents the most important feedback.

Concerning the operator's MW, a key element that has to be taken into particular

account consists of the teleoperation paradigm used for navigating the remote robot. Historically, the main teleoperation paradigms have been grouped into four different categories, namely, direct, multimodal, supervisor and "novel" (i.e., not included in the previous categories) [62]. Examples of direct teleoperation paradigms include the use of traditional devices such as mouse, keyboard, joystick, etc., whereas the voice or gesture commands combined with the traditional inputs are examples of multimodal teleoperation [63]. Supervisory teleoperation paradigm, designed for robots with some LOA, are used to provide methods for reviewing results, monitoring and identifying anomalies, whereas novel paradigms use unconventional input methods, e.g., based on brainwave and muscle movement monitoring.

Many works in robotic telepresence and HRI domains have studied the design of interfaces for addressing the problems arising from possible shortcomings in the human perception of the remote environment as well as the high level of human operator's CD deriving from the remote teleoperation. As a matter of example, the authors of [64] presented a map-centric interface – developed by iRobot Corporation for the Ava 500 robot – for allowing human operators to set the robot's target position by clicking on a 2D map. Similarly, the MITRE Corporation has developed a map-centered system capable of mapping the environment through the use of multiple robots [65]. Robots receive navigation commands from the human operator who defines their destination using 2D coordinates. Small windows showing the video stream received from the robot's onboard camera are placed below the map. Another map-centric interface was proposed by authors of [66], in which information gathered from the environment is combined on a 3D semantic map. During a preliminary exploration of the environment, the operator defines symbols or icons that are used later for "augmenting" the 3D map and providing meanings to objects or places situated in it. A drawback of this type of interfaces is represented by the map itself. Invalid maps created due to faulty sensors or dynamic objects in the environment could lead to a reduction in the SA of the human operator.

In parallel to the aforementioned solutions, other works have oriented towards the development of video-centric interfaces by mainly focusing on two different aspects: the presentation of the information and the employment of different camera configurations. An example of the former aspect can be found in [67], where a 3D-display functionality is combined with a video-centric interface. The information about the robot and the remote environment is overlapped on the video stream of the interface and is made visible to the human operator through the use of VR glasses. Two serious limitations of the considered interface are the human's motion sickness due to the use of a wearable virtual display and the need to maintain the user's head orientation always pointing forward in order to ensure that robot's camera is correctly oriented. Authors of [68] developed a different video-centric interface. Here, the AR technology is exploited to show information such as the horizontal pan-tilt-zoom indicators and the robot's distance from obstacles on the video stream. The area on the screen sides displays other kinds of information, like the camera orientation, the location on the map, and distance

data from infrared and sonar sensors. The robot is controlled via the keyboard or by clicking the control buttons displayed on the interface. The main disadvantage of this interface is the way in which sensor data are shown. In fact, mixing them with the map information would allow the operator to better perceive where the robot is actually located [61].

Other works proposed different video capturing solutions. For example, in [69], remote robot teleoperation with six different FOV configurations (from narrow to omnidirectional) was investigated in both real and virtual environments. Obtained results showed better operators' performance when wide FOVs were used (from 120° up). The only drawbacks of this solution were the latency and the communication delay growing with the increase of the FOV. Similarly, in [70], the effectiveness of three different camera configurations with three diverse FOVs (45°, omnidirectional 360° and fisheye 180° FOV) in remote teleoperation applications was explored. According to experimental observations, the fisheye and omnidirectional configurations were preferred by users as they were allowed to have a clear view of the environment surrounding the robot. However, the distortions introduced in the omnidirectional images did not allow users to understand the position and orientation of the robot. In addition, the authors of [71] explored the use of two different control techniques (keyboard and through-the-screen, or TTS, which requires the user to define a path in the robot camera view using the mouse) combined with three different camera configurations (101° perspective camera, 185° wide-angle camera and 185° wide-angle camera with distortion-free center area). Obtained results showed a strong preference for the undistorted fisheye configuration with the TTS control mode also due to the perspective limitations, which did not allow users to move the camera to see obstacles on the floor.

Works discussed above were selected as representative examples of the progress made by the academics. From a market-oriented perspective, telepresence robots generally exhibit video-centric interfaces accompanied by a map and endowed with direct or supervisory teleoperation paradigms. Well-known examples include the Padbot robot [72], the VGo robot [73], the Beam Smart Presence robot [74] and the Ra.Ro robot [75], to name a few. These interfaces allow human operators to supervise robots via a real-time video feed and to set the intended destination by defining coordinates on a map. A different navigation method, called "smart drive", was integrated in the Ra.Ro robot allowing the human user to guide the robot by pointing and clicking a target destination directly on the video feed.

By moving from the review discussed above, it can be observed that a number of interfaces for remotely operating telepresence robots have been proposed by researchers from both industry and academy. Although the panorama of available interfaces is quite diversified, none of these solutions has emerged yet as the ultimate approach to robotic telepresence. For this reason, further analyses are actually needed to properly support next advancements in the field. In this direction, some studies have been already conducted. As a matter of example, the authors of [59] have compared video-centric and

map-centric interfaces by investigating three different configurations, namely, video-only, map-only and video plus map. Experimental results showed that merging map and video information improves overall users' performance since they complement each other. Concerning video capturing solutions, it can be observed that robots' onboard cameras generally exhibit limited FOVs that worsen operators' performance in remote navigation tasks. Hence, wide-angle FOVs represent the most used perspective modality in robot's camera, although they can suffer from distortion and no close-up view issues.

2.1.2 Robotic Platform

The robotic platform considered in this application domain (originally exploited for cultural heritage scenarios [76]) is named *Virgil*.

Virgil is a telepresence robot devised from the collaboration between TIM JOL CRAB and the Department of Architecture and Design of the Politecnico di Torino (Fig. 2.1).

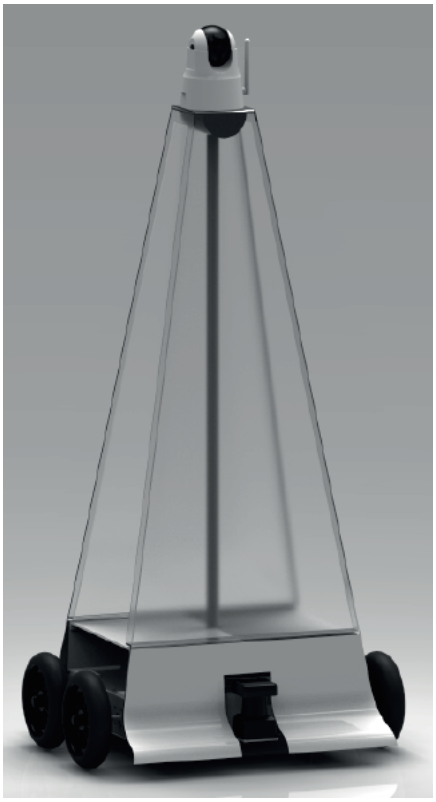


Figure 2.1: *Virgil*.

It is a wheeled mobile robot connected to a Robot Operating System[1] (ROS)-based

[1]https://www.ros.org

platform for cloud robotics created by TIM and equipped with a pan-tilt camera for HD shooting of environments and a laser sensor for obstacle avoidance. The LTE (Long-Term Evolution) technology – generally used for mobile broadband communications – was also exploited in order to both ensure low latencies in the connection with the cloud robotics platform and a safe robot navigation with short response times.

Virgil's cover is made of PMMA (PolyMethylMethAcrylate), a transparent material used for guarantee lightness and easy movements to the robot. The cover can also be customized, that means the robot can change coverage depending on the place in which it will be used. The mechanical structure of the robot's base was designed by Nuzoo Robotics [75], with the following characteristics (Fig. 2.2):

- a four-wheeled robot with a steel frame powered by two electric engines able to transmit the motion to the wheels with a series of gears and belts by providing traction on each side, in a separate way;
- a weight of about 14 Kg and a height of 120 cm;
- a docking station;
- a Li-Fe 12 V battery, with an autonomy of approximately 4 hours;
- a maximum velocity of 1 m/s;
- a pyramid-shaped cover with a rectangular base of 60x50 cm;
- WiFi and 4G-LTE connectivity;
- a maximum slope of 30°;
- a Hokuyo UTM-30LX as laser scanner sensor:
  - maximum detection distance: 30 m;
  - maximum angle: 270°;
  - environments: indoor/outdoor;
  - dimensions: W60xD60xH87mm;
  - power supply: 12 V;
  - weight: 370 g.
- a NUC DN2820FY mini PC with Intel CPU Celeron N2820@2.1 GHz (dual-core) onboard control unit equipped with a set of USB ports used to acquire data from the laser sensor, the Ethernet card, the camera, the LTE dongle and the commands for the pan-tilt system;

- advanced features: autonomous navigation (by using its local and global path planning functionalities, which rely on a map of the environment created in a preliminary exploration phase), obstacle detection and obstacle avoidance.

With respect to the original setup, for the purpose of this thesis, the camera on the robot's head was replaced by a tablet device on a pan-tilt support, which makes it possible to display at the remote site the face of the operator, thus enhancing the sense of presence.

Name	Number	Material
1- wire box	x1	perspex
2- cover	x1	trasparent perspex
3- vertical support structure	x1	perspex
4- hardware components cover	x2	perspex
5- support bracket	x1	perspex
6- bottom support brackets	x2	perspex
7- front bumper	x1	perspex
8- rear bumper	x1	perspex

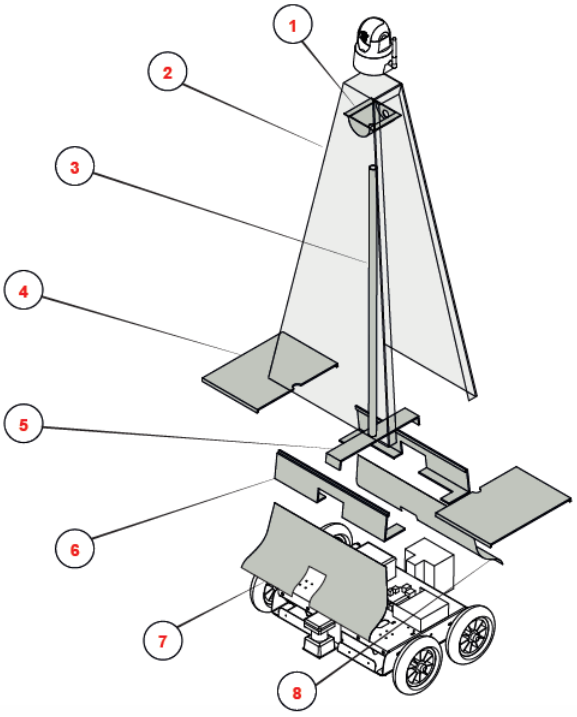


Figure 2.2: *Virgil* blueprint.

2.1.3 Telepresence Framework

By moving from the review discussed in Section 2.1.1, the goal of this paragraph is to build on results reported in [59] to explore and investigate a richer scenario. To this aim, a mixed map & video-based teleoperation system, and two major robot's navigation modalities used in recent solutions available on the market, i.e., *keyboard teleoperation* and *point-and-click video navigation*, were developed and compared from a usability perspective. Moreover, a combination of the two navigation modes was also considered. Afterwards, the impact of three different robot's camera configurations, i.e., a fixed camera with a narrow (45°) FOV, a perspective camera with a wide-angle (180°) horizontal FOV endowed with pan-tilt capabilities, as well as a fixed fisheye camera

with a wide-angle (180°) diagonal FOV and a no-distortion central area were also compared and studied. The analysis was performed by developing a telepresence framework implementing the above configurations and integrating them in the *Virgil* robot.

The logical components that were assembled and/or developed to implement the robotic telepresence system are illustrated in Figure 2.3. The architecture is made up of three different layers, namely, *Client*, *Cloud Robotic Platform* (CRP) and *Robot*.

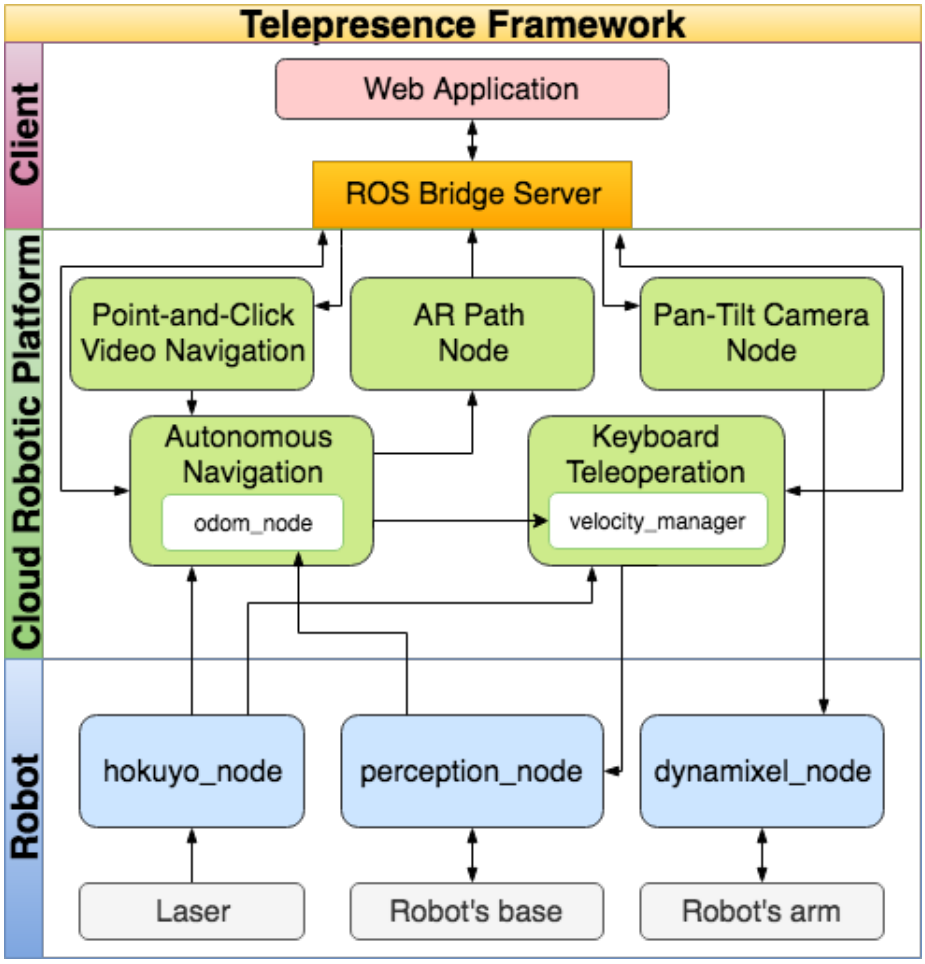


Figure 2.3: Telepresence framework.

The *Client* layer consists of a Web application, which communicates both with the *CRP* layer hosting the navigation algorithms and the *Robot* layer via RosBridge Server and `roslibjs`, a JavaScript-based library for using ROS on the Web [77].

The *CRP* layer represents the robot's "brain" providing all the functionalities needed by the robot to navigate the environment. It consists of two different navigation modalities, namely, *keyboard teleoperation* and *point-and-click video navigation*. In both the modalities, the input speed commands are sent and filtered by the *velocity_manager* node that is responsible for stopping the robot when the laser sensor (in the *Robot*

layer) detects an obstacle too close to the robot. The *AR Path Node* and the *Pan-Tilt Camera Node* are responsible for the visualization of the path to reach the goal in AR and to send pan and tilt commands to the robot's camera (mounted on top of the robot trough a robotic arm) by tracing the movement of the mouse, respectively. As illustrated in Figure 2.3, the *Pan-Tilt Camera Node* sends mouse movement commands to the *dynamixel_node* in the *Robot* layer.

The *Robot* layer is totally decoupled and independent from intelligent modules since the "brain" lays all in the CRP. It contains the robot's sensors and mechanical parts as well as the associated software nodes. In particular, the *hokuyo_node* is responsible for gathering the distance data from obstacles delivered by the laser sensor and transmitting them to both navigation modules in the *CRP* layer. The *dynamixel_node* is the module devoted to gain the mouse movement commands from the *Pan-Tilt Camera Node* and to transmit them to the robotic arm chain for making the robot's camera move. Lastly, the *perception_node* plays two different roles. On the one hand, it receives speed inputs from both the *Autonomous Navigation* and *Keyboard Teleoperation* modules through the *velocity_manager* node and converts them into commands for the robot's motors. On the other hand, it provides the space traveled by the robot's wheels to the *odom_node* in order to compute the odometry.

Details on two teleoperation interfaces as well as on the different camera configurations are reported in the next sections.

2.1.4 Teleoperation Interfaces

In this paragraph, the two navigation modalities, which differ in the way the operator can control the robot and in the type of feedback returned, will be introduced and described in detail.

In both modalities, the interfaces features a large video window showing the live stream from the robot's onboard camera as well as a smaller window showing the video captured by a local webcam (also displayed on the remote tablet mounted on the robot). A colored bar split in three regions (left, front, right) is placed below the video window to show the distances of the robot from obstacles. The color of each region changes from green to red passing through yellow according to the actual measurements of the laser sensor. On the left side of the interface, the position and orientation of the robot are shown on a map in real-time by using an orange triangle. In addition, in the *point-and-click video navigation*, a green triangle is used to depict the orientation and the location that the robot will assume once it will reach the clicked destination. It is worth noting that navigation algorithms exploited by the robot to move in the environment actually work on a different map created by using the robot's laser sensor and applying a Simultaneous Localization and Mapping (SLAM) strategy.

In the *keyboard teleoperation* modality, the operator manually controls the robot throughout the environment by using the directions keys (Fig. 2.4). The up and down arrow keys are used to make the robot move forward or backward in the environment

by changing its linear velocity, whereas the left and right arrow keys change its angular velocity by twisting it accordingly. When pressed together, the above keys can be used to merge the linear and angular velocities by making the robot move in the given direction while turning left or right. The same commands can be issued by clicking on the corresponding icons displayed on the interface. In this case, two icons are used to control the simultaneous use of direction and orientation commands. A local path planning algorithm is used by the robot to navigate the environment in this case.

The current robot's direction and the command issued by the user are shown on the video stream through AR arrows. The live video stream and the map represent the feedback provided to the operator.

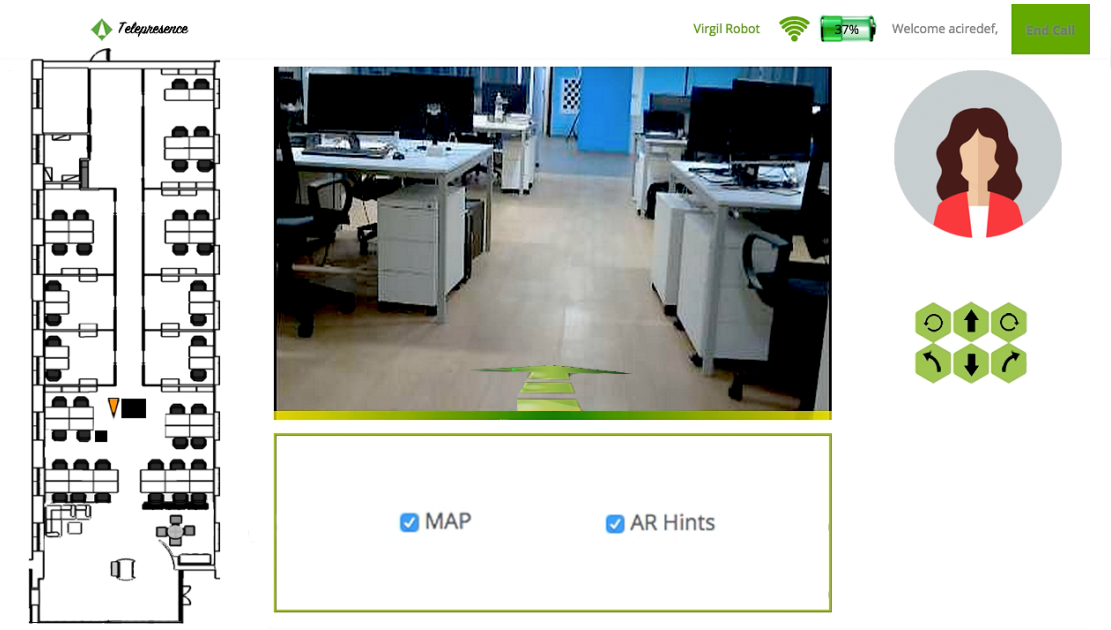


Figure 2.4: Keyboard teleoperation interface.

In the *point-and-click video navigation* modality, the operator defines a target destination for the robot by simply clicking it on the video stream received by the robot's onboard camera (Fig. 2.5).

Since the pan-tilt configuration and the intrinsic parameters of the camera are known, the coordinates of the clicked pixel can be converted to a point on the map by using ray-tracing. This latter point is sent as a goal to the global path planning algorithm, which is devised to move the robot towards the target location at a constant speed by avoiding both moving and fixed obstacles. With respect to the previous modality, at any point in time, the robot knows the path elaborated by the planning algorithm to reach the goal. As illustrated in Figure 2.6, this path is overlapped to the video stream in AR. It is worth observing that, when the whole FOV of the robot's onboard camera does not allow to click any point of the floor (e.g., the robot is too close to an obstacle or

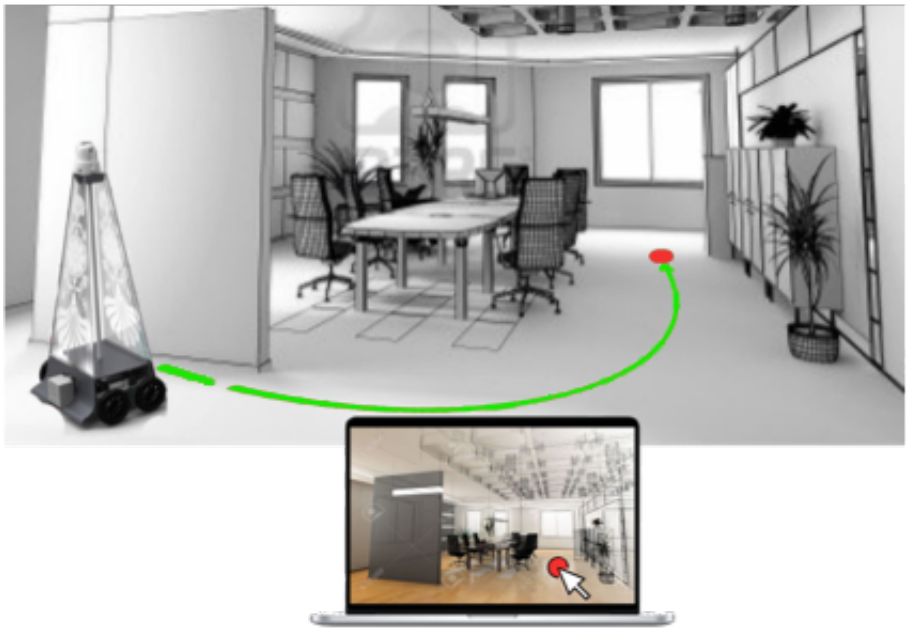


Figure 2.5: Point-and-click video navigation.

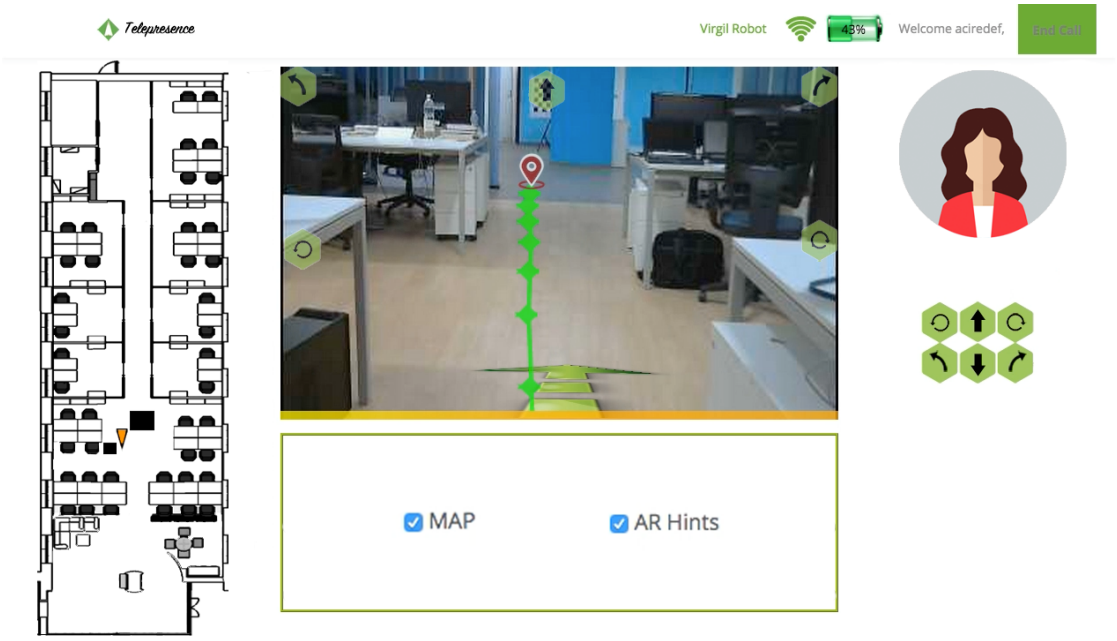


Figure 2.6: Point-and-click video navigation interface.

to the walls), an alternative control method can be used. With this method, the operator

can click on several active icons displayed on the edges and corners of the video window (like in Google Street View[2]) to directly control robot's angular or linear velocity. When such commands are issued, goals that might have been set for autonomous navigation are cleared. A new goal could then be specified once the target destination or an intermediate location are visible again in the camera's FOV. Furthermore, in order to make this modality comparable to the keyboard-based one, the possibility to pan-tilt the camera has not been considered in order to keep the focus on robot's navigation.

2.1.5 Camera Configurations

This section describes the three camera configurations that have been studied, which differ in the way the user can control the position and orientation of the camera mounted on top of the robot and the size of the FOV.

In the first configuration, later referred to as *Narrow FOV* (*NFOV*), the robot is endowed with a forward-facing camera characterized by a common 45° FOV, as illustrated in Figure 2.7.

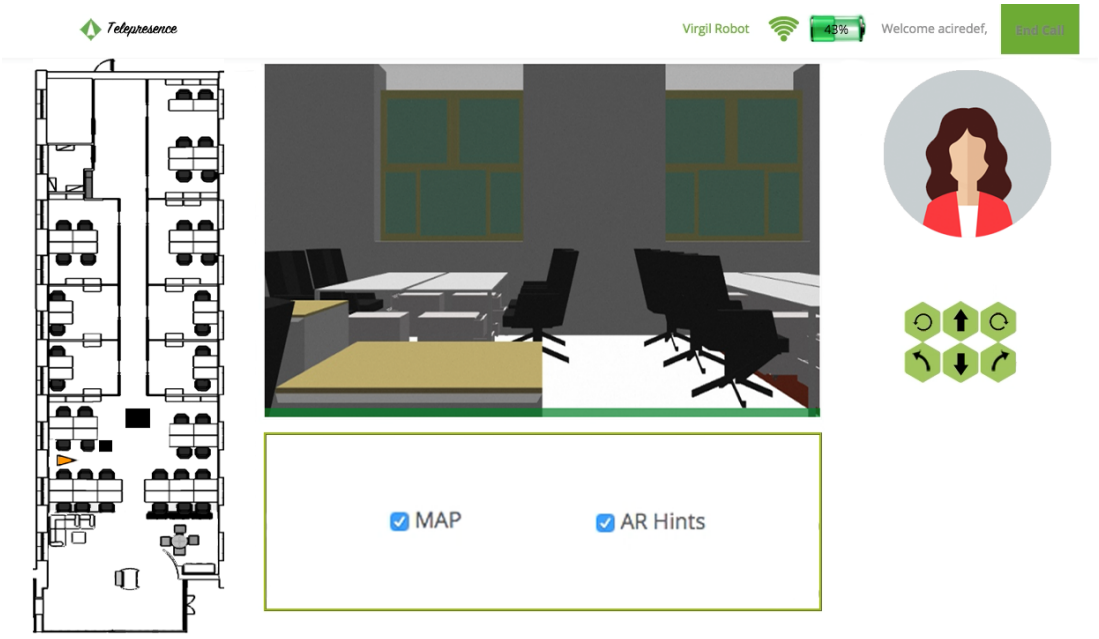


Figure 2.7: Narrow FOV (*NFOV*) configuration.

In the second configuration, a wide-angle perspective camera with a 180° horizontal FOV and pan-tilt capabilities (later abbreviated as *WFOV&PT*) was exploited, as shown in Figure 2.8. The pan-tilt was introduced to overcome the limitations experienced in

[2]https://www.google.com/streetview/

[71], which were because users were not able to move the camera to find a point to click or to see close obstacles on the floor in the perspective view. The user can issue vertical or horizontal orientation sliding commands by using the left button of the mouse and the three events *mouseDown*, *mouseMove*, *moveUp*. Specifically, the *mouseDown* event is used to hook the current view, the *mouseMove* event to drag it to the desired position (e.g., dragging the view to the left allows the user to look to the right, like in Google Street View) and the *mouseUp* event to leave the updated view in the desired position. Two pan-tilt position icons and a "Center Pan-Tilt" button were also introduced in this configuration, in order to let users be aware of the actual position of the camera (in terms of pan and tilt displacements) and to re-center it after a movement, respectively (Fig. 2.8).

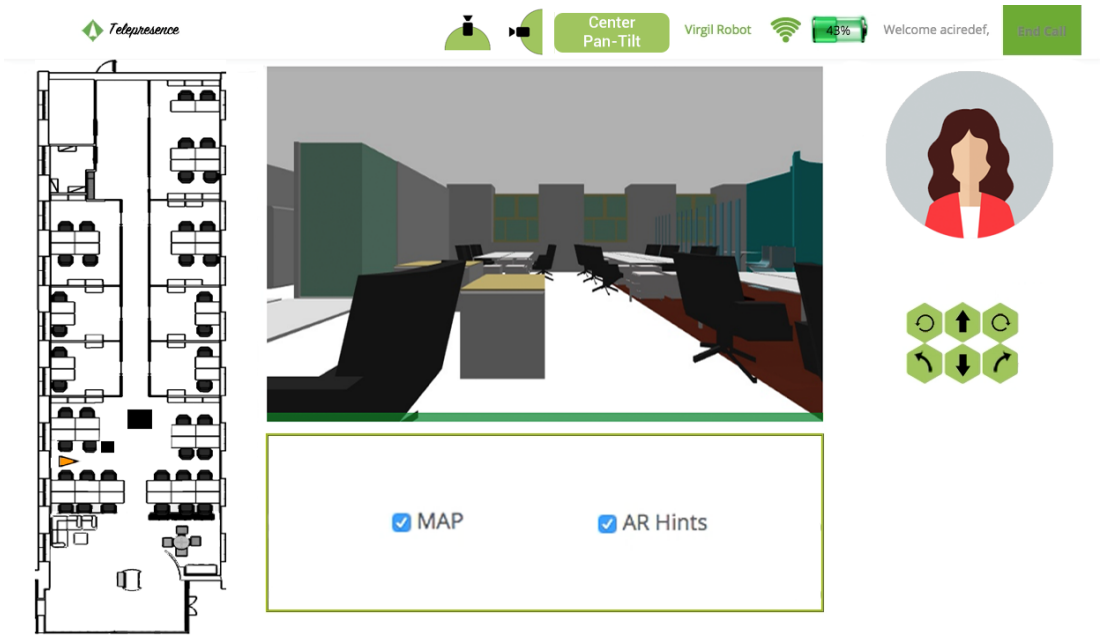


Figure 2.8: Wide FOV plus pan-tilt (*WFOV&PT*) configuration.

In the third configuration, a wide-angle fisheye camera with a 180° diagonal FOV (abbreviated as *FFOV*) was exploited, as illustrated in Figure 2.9. Since a fisheye lens suffers from radially symmetric distortions, a radially symmetric image remapping phase was implemented in order to obtain an undistorted (perspectively correct) circular region in the center of the view. As the intrinsic parameters of the camera, as well as its distortion vector, were known, the pixels in the circular area with a given radius could be undistorted, rectified and remapped on the image in order to generate a perspectively correct view inside the circle with the same radius. These steps (excluding the remapping) were also exploited in the *point-and-click video navigation* modality (described above). All the cameras transmitted a video stream at approximately 30 frames per second with a resolution of 1024x768 pixels.

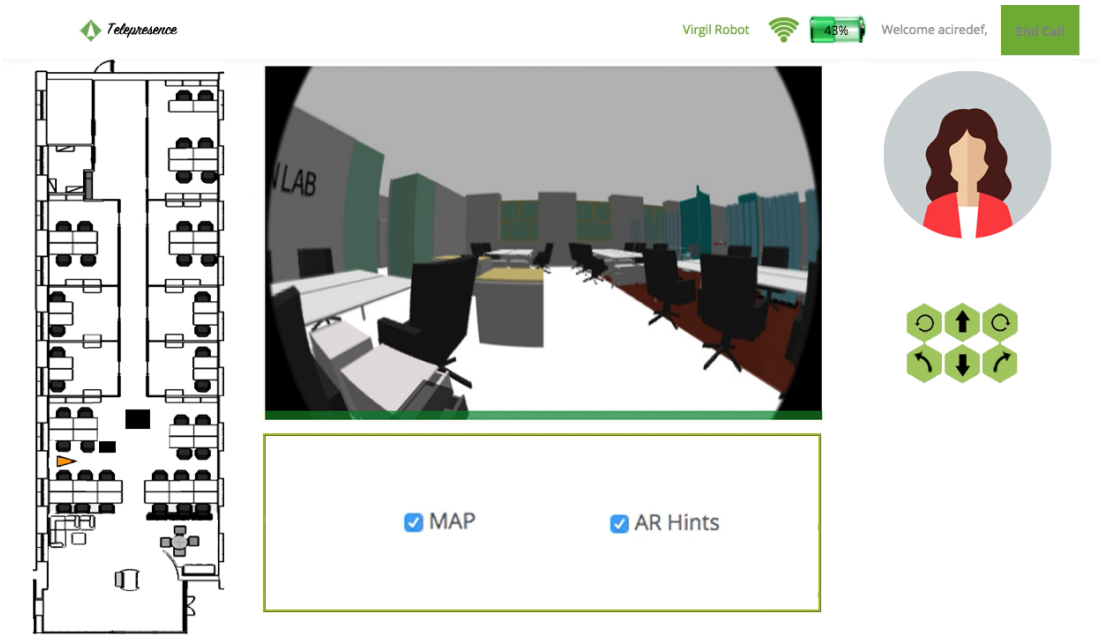


Figure 2.9: Fisheye FOV (*FFOV*) configuration.

It is worth observing that, in order to evaluate the above camera configuration alternatives before possibly moving to physically implementing them on a real robot, a simulation environment as well as a 3D version of the *Virgil* robot were used to experiment with the different setups. In particular, since all the modules of the telepresence framework were implemented with ROS, no further work was required to migrate them on the ROS-based Gazebo simulator[3].

2.1.6 Experimental Results

In this section, experimental observations that were carried out to assess the effectiveness and usability of the various control modalities for the execution of remote navigation tasks in the considered context, as well as the impact of different camera configurations are presented. Specifically, two user studies were performed.

The first study was aimed to identify the most suitable and appropriate teleoperation interface(s) by comparing the *keyboard teleoperation* (later abbreviated, *keyboard*, or *K*) and *point-and-click video navigation* (abbreviated *point-and-click*, or *P*) interfaces and the combination of the two (*combined*, or *C*).

The second study was aimed to investigate, through a comparative analysis involving three different camera configurations (i.e., *NFOV*, *WFOV&PT*, and *FFOV*), how an

[3]https://gazebosim.org

augmented FOV and/or a pan-tilt camera can affect users' performance and their SA in remote robot navigation tasks.

In both user studies, participants were told that they should reach the office of a person they were looking for by controlling the robot within a remote environment. In particular, each participant was invited to master the robot for accomplishing three different navigation tasks, namely, *T1 - Reach the corridor*, *T2 - Reach the room*, and *T3 - Enter/exit the room* (Fig. 2.10). Such tasks were specifically devised to assess the suitability of the proposed interfaces and camera configurations in a possible scenario the robot could be involved into when used in an office environment.

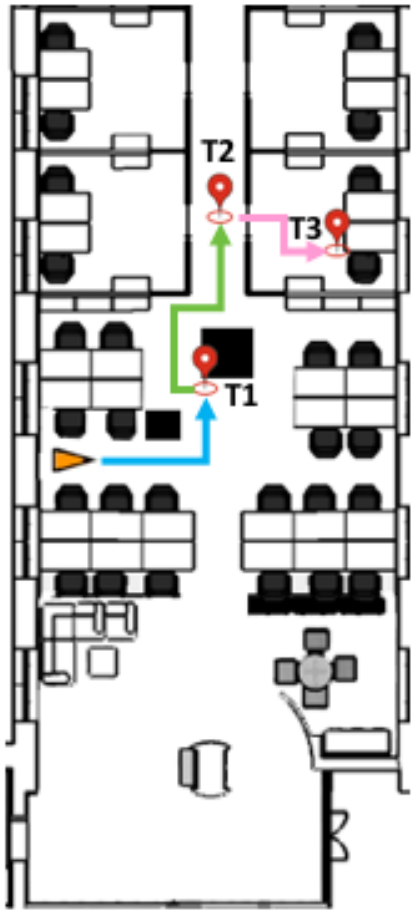


Figure 2.10: Map of the environment considered in the experiments, initial location of the robot, destinations to be reached in the tasks and possible paths.

In particular, *T1* was designed to test both the teleoperation interfaces and the camera configurations when controlling the robot towards a destination that is not framed in the camera's FOV. *T2* was designed to test the interfaces and users' performance with different FOVs in driving the robot when obstacles are to be avoided. Lastly, *T3*

was meant to study a scenario in which users had to drive the robot in constrained spaces.

At the beginning of the experiment, the robot is located in the open space of the office environment (on the left side of the corridor and oriented towards it) in order to guarantee that the camera cannot frame the point to be reached in the task *T1*. The robot's position and orientation are depicted by an orange triangle marker in the map, as shown in Figure 2.10. Then, participants have to drive the robot in order to reach the point labeled *T1*, by following a possible path like the one drawn in blue on the map. Afterwards, participants have to teleoperate the robot to the location labeled *T2* and positioned in front of the room they were looking for. During this task, the obstacle in the corridor (depicted as a black square) has to be avoided. A possible path to be followed is shown in green on the map. Lastly, in the third task, participants have to guide the robot for making it enter into the room, drive it close to a desk (indicated by label *T3*), twist and exit the room. A possible path is shown by the pink arrow on the map. Details about the two user studies conducted in this domain are illustrated below.

First User Study: Which Teleoperation Interface?

The first study involved 12 participants (8 males and 4 females) aged between 25 and 29 years (M = 26.58 SD = 1.24), recruited among university students from Politecnico di Torino. According to declarations collected, 55% had already used interfaces based on point-and-click, and 75% of them had previous experience with keyboard-based interfaces for issuing direction commands (e.g., in video-games).

At the beginning of the experiment, brief training was provided to participants for instructing them on the use of the various teleoperation interfaces (Section 2.1.4). Afterwards, participants were invited to carry out the three navigation tasks reported above by experimenting all the three interfaces. To compensate for possible learning effects, a random order was used for choosing the sequence of the interfaces to be experienced.

During each experiment, quantitative data about time required to complete the tasks and number of interactions (key presses and/or mouse clicks, depending on the interface considered) were measured. After having tested a given teleoperation interface, participants were asked to fill in a NASA TLX [78] questionnaire in order to evaluate their perceived MW on a six-dimension subjective scale, i.e., *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort*, and *frustration*. A score in the range [0;100] was assigned to each dimension and a global TLX score was then calculated by combining the six individual ratings via a weighting mechanism. At the end of the experiment, each participant was asked to compile a usability questionnaire in order to evaluate his or her experience.

The usability questionnaire was split in three parts. The first part was designed by considering the Nielsen's Attributes of Usability (NAU) [79]. NAU requested participants to assess the various teleoperation interfaces through five usability factors, i.e.,

learnability, *efficiency*, *memorability*, *errors* and *satisfaction*, by expressing their agreement on a 5-point Likert scale.

The second part was designed by considering the Subjective Assessment of Speech System Interfaces (SASSI) methodology [80] and adapting it to let participants judge the user experience with the given interaction means. The (adapted) SASSI questionnaire asked participants to evaluate the experimented interfaces through six usability factors, namely, *system response accuracy*, *likeability*, *cognitive demand*, *annoyance*, *habitability* and *speed*, by expressing their agreement on a 5-point Likert scale. For consistency reasons, scores for *cognitive demand* and *annoyance* factors, were inverted so that higher scores reflect positive opinions.

The third part requested participants to rank the experience made with the different teleoperation interfaces by providing their judgment both for the whole experiment as well as for the three individual tasks.

A one-way repeated measures ANOVA (Analysis of Variance) test (significance level of 0.05) was carried out first on collected data, in order to check for possible differences in participants' subjective evaluation and objective performance among the three teleoperation interfaces. A post-hoc analysis was then performed by applying a two-tailed paired t-test (significance level of 0.05) to compare the various interfaces and determine which of the three means are statistically different.

Results obtained in terms of completion time as well as number of interactions required to complete the tasks are reported in Figure 2.11 and Figure 2.12, respectively.

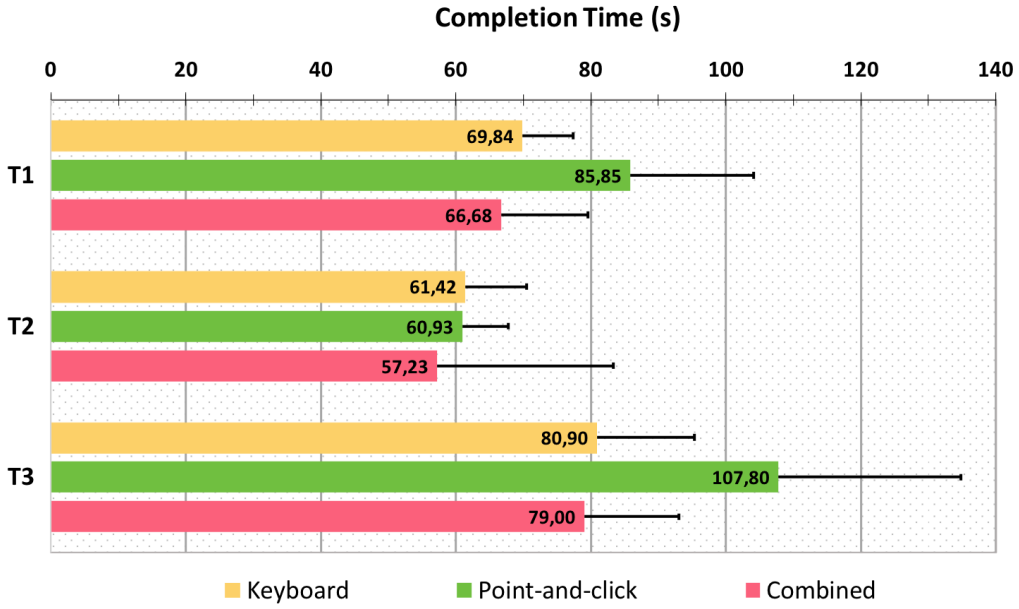


Figure 2.11: First user study: results in terms of completion time required to complete the tasks with the considered interfaces. Bar lengths report average values (lower is better), whereas whiskers report standard deviation.

At first sight, Figure 2.11 reveals that participants using the *keyboard* and the *combined* interfaces tended to complete the tasks *T1* (ANOVA: $p = 2.43 \times 10^{-3}$) and *T3* (ANOVA: $p = 2.08 \times 10^{-3}$) faster than those experimenting with the *point-and-click* interface. Post-hoc analysis revealed no statistically significant differences between the *keyboard* and the *combined* interfaces for all the three navigation tasks, as illustrated in Table 2.1. However, differences between *keyboard* and *point-and-click* interfaces, as well as between *combined* and *point-and-click* interfaces, were found to be significant both in *T1* and *T3*, as outlined in Table 2.1. Results for *T2* were not statistically significant.

Statistically significant differences in terms of completion time were also found for the users' gender classes. In particular, females were faster than males in accomplishing *T1* with the *combined* interface (ANOVA: $p < 0.05$) and *T3* with the *point-and-click* one (ANOVA: $p < 0.05$). Concerning users' previous experience with keyboard-based interfaces, obtained results shown that those who declared an everyday usage frequency were able to complete tasks in less time than those who indicated to be used working with this kind of interfaces once a week (ANOVA: $p < 0.01$). Results for prior knowledge about point-and-click interfaces were not statistically significant.

Table 2.1: First user study: post-hoc analysis on completion time results and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

	K vs. P	**C vs. P**	**K vs. C**
T1	t[11] = -2.97, $p = 1.28 \times 10^{-2}$ (+)	t[11] = 3.29, $p = 7.19 \times 10^{-3}$ (++)	t[11] = 0.79, $p = 0.4483$
T2	t[11] = 0.15, $p = 0.8810$	t[11] = 0.55, $p = 0.5953$	t[11] = 0.50, $p = 0.6302$
T3	t[11] = -2.80, $p = 1.73 \times 10^{-2}$ (+)	t[11] = 3.51, $p = 4.86 \times 10^{-3}$ (++)	t[11] = 0.36, $p = 0.7280$

Results obtained in terms of number of interactions (Fig. 2.12) indicate that, for all the tasks (*T1*: ANOVA: $p = 1.51 \times 10^{-4}$, *T2*: ANOVA: $p = 4.99 \times 10^{-6}$ and *T3*: ANOVA: $p = 5.10 \times 10^{-3}$), the *keyboard* interface required the highest number of interactions followed by the *point-and-click* interface and then the *combined* one. As illustrated in Figure 2.12, these differences become evident and much higher in the case of *T2* with respect to *T1* and *T3*. Post-hoc analysis confirmed this observation, as illustrated in Table 2.2, where statistical significance differences were found both between *K* and *P* interfaces as well as *K* and *C* interfaces in task *T2*. Furthermore, differences between all the interfaces were found to be also significant in *T1*, whereas in *T3*, no statistically significant difference was found between the *keyboard* and the *point-and-click* interfaces.

Concerning the users' gender classes, females required a lower number of interactions than males to complete *T3* with the *point-and-click* interface (ANOVA: $p < 0.05$). Results for prior knowledge about keyboard-based interfaces shown that, who stated to use this kind of interfaces with a daily usage frequency issued a lower number of teleoperation commands compared to those who declared to use them once a week (ANOVA: $p < 0.01$). Like for the completion time results, no statistically significant differences were found concerning prior knowledge about point-and-click interfaces.

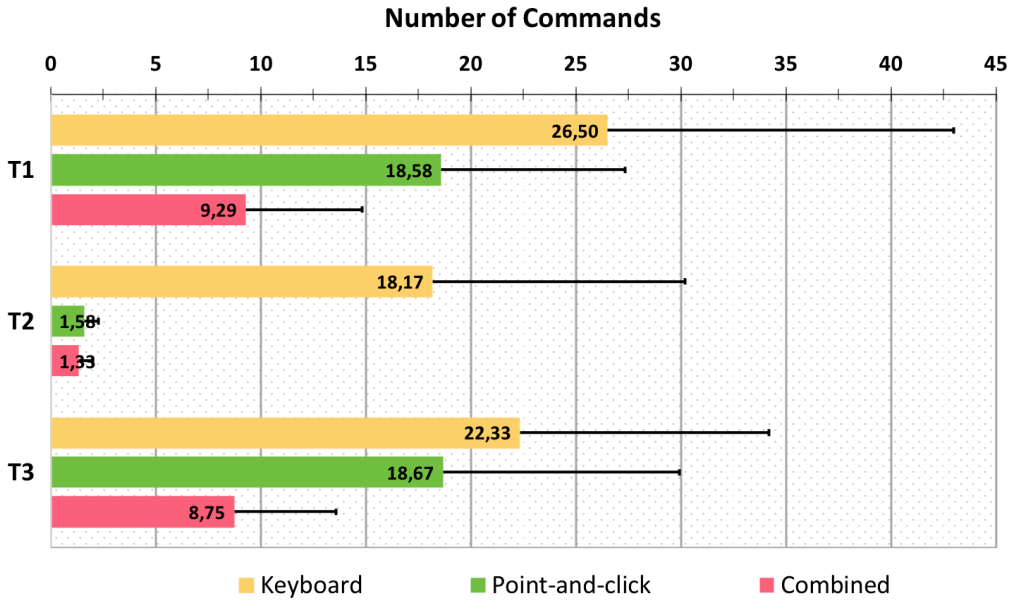


Figure 2.12: First user study: results in terms of number of interactions required to complete the tasks with the considered interfaces. Bar lengths report average values (lower is better), whereas whiskers report standard deviation.

Table 2.2: First user study: post-hoc analysis on number of interactions and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	K vs. P	C vs. P	K vs. C
T1	t[11] = 2.29, p = 4.31 $\times 10^{-2}$ (+)	t[11] = 4.13, p = 1.67 $\times 10^{-3}$ (++)	t[11] = 4.30, p = 1.25 $\times 10^{-3}$ (++)
T2	t[11] = 4.68, p = 6.77 $\times 10^{-4}$ (+++)	t[11] = 1.00, p = 0.3388	t[11] = 4.80, p = 5.56 $\times 10^{-4}$ (+++)
T3	t[11] = 0.78, p = 0.4533	t[11] = 2.92, p = 1.40 $\times 10^{-2}$ (+)	t[11] = 4.31, p = 1.23 $\times 10^{-3}$ (++)

Reduced number of interactions and lower completion time for the *combined* and the *keyboard* interfaces are observed as well when summing up results obtained for the three tasks, i.e., considering them altogether as a single experiment.

Results obtained in terms of participants' MW appear to describe an almost comparable situation. In fact, as illustrated in Figure 2.13, the *point-and-click* interface was judged by participants as the most cognitive demanding teleoperation interface, followed by the other two, thus confirming the fact that users were faster in accomplishing tasks with *keyboard* and *combined* interfaces.

Considering the results obtained with the subjective evaluation based on the NAU methodology, it can be observed from Figure 2.14, that participants judged the *combined* and *keyboard* interfaces more usable than the *point-and-click* one for all the five factors

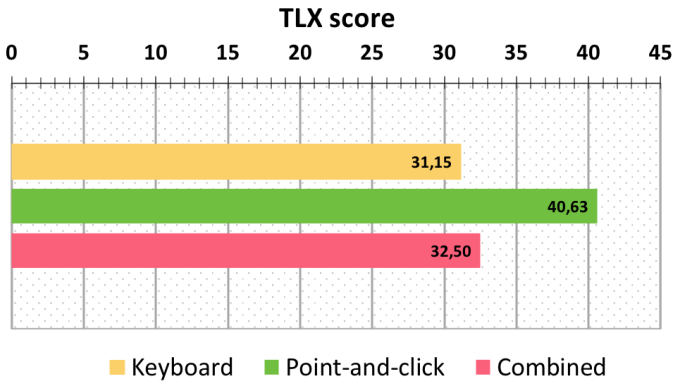


Figure 2.13: First user study: results concerning MW measurements for the three teleoperation interfaces during the whole experiment. Bar lengths report TLX score (lower is better).

considered. Statistical significance differences obtained with the ANOVA tests are reported in Figure 2.14 through the + symbols (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

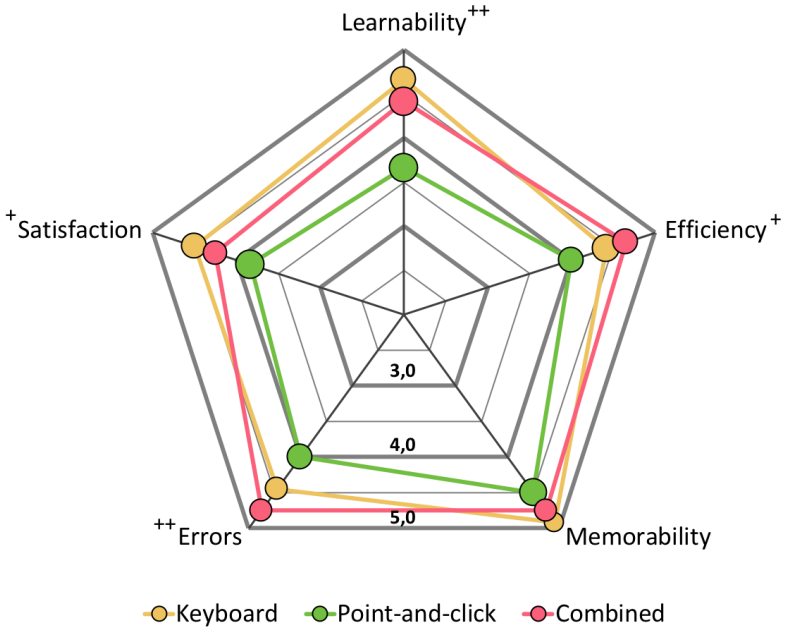


Figure 2.14: First user study: results concerning the usability of the three interfaces for the whole experiment based on NAU factors. Circle position reports average values (higher is better), circle dimension reports standard deviation, whereas + symbols report statistical significance determined with the ANOVA tests (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

The outcomes of the post-hoc analysis (Table 2.3) confirmed the above observations,

by showing no statistically significant differences between *K* and *C* interfaces for all the five usability attributes. By digging more in detail and considering the *learnability* and *error* factors, the differences between the *K* and *C* interfaces with the *P* one were pronounced and reached statistical significance. Similarly, the *C* interface was evaluated more positively than the *P* one in terms of *efficiency*, whereas the *K* interface was judged more satisfactory compared to the *P* one (Table 2.3).

Table 2.3: First user study: post-hoc analysis on NAU results and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	K vs. P	**C vs. P**	**K vs. C**
Learnability	t[11] = 3.32, $p = 6.87 \times 10^{-3}$ (++)	t[11] = -2.69, $p = 2.10 \times 10^{-2}$ (+)	t[11] = 1.39, $p = 0.1910$
Efficiency	t[11] = 2.16, $p = 0.0538$	t[11] = -2.57, $p = 2.61 \times 10^{-2}$ (+)	t[11] = -0.56, $p = 0.5862$
Memorability	t[11] = 1.60, $p = 0.1372$	t[11] = -1.39, $p = 0.1910$	t[11] = 1.00, $p = 0.3388$
Errors	t[11] = 2.57, $p = 2.61 \times 10^{-2}$ (+)	t[11] = -2.60, $p = 2.46 \times 10^{-2}$ (+)	t[11] = -1.48, $p = 0.1660$
Satisfaction	t[11] = 2.60, $p = 2.46 \times 10^{-2}$ (+)	t[11] = -1.60, $p = 0.1372$	t[11] = 1.39, $p = 0.1910$

Results concerning participants' evaluations gathered through the (adapted) SASSI methodology are illustrated in Figure 2.15 (+ symbols report statistical significance differences obtained with the ANOVA tests, i.e., $+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

At first sight, it appears that, participants' scores were higher for the *keyboard* and *combined* interfaces than the *point-and-click* one for five out of the six usability factors. In particular, the *point-and-click* interface was judged more positively only in terms of *annoyance*, i.e., how much the interface was evaluated repetitive and boring.

Results obtained with the post-hoc analysis revealed no statistically significant differences between the *K* and the *C* interfaces for all the six usability attributes (thus confirming the above results), as illustrated in Table 2.4. However, differences between the *K* and *P* interfaces as well as between the *C* and *P* ones, resulted to be pronounced (as illustrated in Fig. 2.15) and statistically significant both in terms of *system response accuracy*, *likeability* and *cognitive demand*. In particular the keyboard-based interface was evaluated to be more accurate compared to the *point-and-click* one. Furthermore, the *K* interface was also judged more positively than the *P* one in terms of *speed* of interactions with the robotic system.

Data collected in the third part of the questionnaire, regarding users' preferences in using the three teleoperation interfaces both for the whole experiment as well as the individual tasks, are illustrated in Figure 2.16. Considering overall rankings, it appears that the favorite interface is the *combined* one. When individual tasks are considered, it could be observed that the *combined* interface is the one that was preferred for performing *T3* and *T1* (see in particular yellow and green columns in Figure 2.16, where the number of times a given interface has been ranked 1st or 2nd is showed). Based on the feedback gathered during the tests, the preference seems to be mainly motivated by the

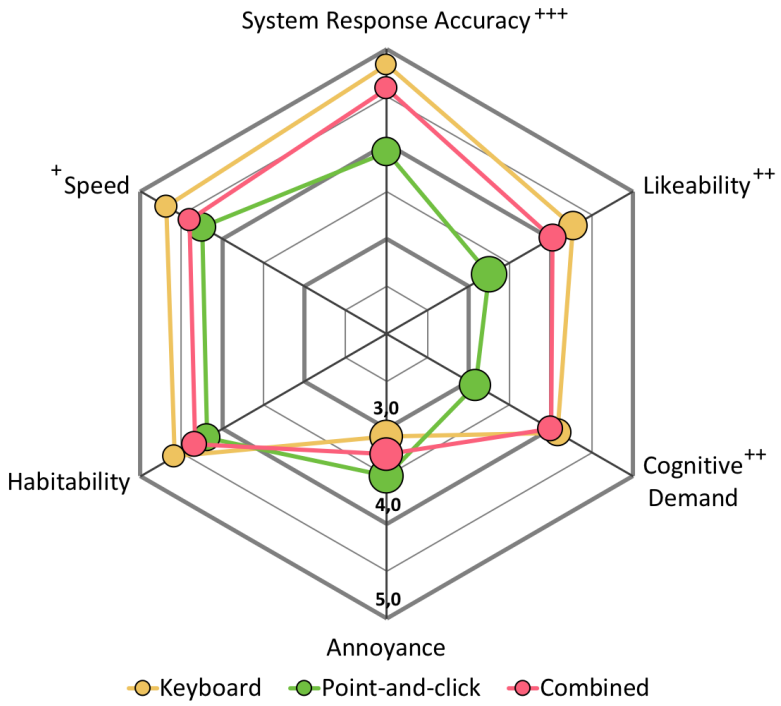


Figure 2.15: First user study: results concerning the usability of the three interfaces for the whole experiment based on (adapted) SASSI methodology. Circle position reports average values (higher is better), circle dimension reports standard deviation, whereas + symbols report statistical significance determined with the ANOVA tests (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

Table 2.4: First user study: post-hoc analysis on SASSI results and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

	K vs. P	**C vs. P**	**K vs. C**
System Response Accuracy	t[11] = 4.00, $p = 2.07 \times 10^{-3}$ (++)	t[11] = -2.60, $p = 2.46 \times 10^{-2}$ (+)	t[11] = 1.91, $p = 0.0818$
Likeability	t[11] = 2.56, $p = 2.65 \times 10^{-2}$ (+)	t[11] = -2.80, $p = 1.72 \times 10^{-2}$ (+)	t[11] = 0.56, $p = 0.5852$
Cognitive Demand	t[11] = 2.35, $p = 3.88 \times 10^{-2}$ (+)	t[11] = -2.93, $p = 1.37 \times 10^{-2}$ (+)	t[11] = 0.43, $p = 0.6742$
Annoyance	t[11] = -0.29, $p = 0.7773$	t[11] = -1.77, $p = 0.1039$	t[11] = -1.60, $p = 0.1372$
Habitability	t[11] = 1.77, $p = 0.1039$	t[11] = -0.29, $p = 0.7773$	t[11] = 1.39, $p = 0.1910$
Speed	t[11] = 2.35, $p = 3.88 \times 10^{-2}$ (+)	t[11] = -1.91, $p = 0.0818$	t[11] = 1.00, $p = 0.3388$

fact that, as it could be largely expected, participants were allowed to switch between the two interfaces when needed, thus benefiting from the advantages of both of them.

In summary, by combining the above results with those concerning users' interaction (Fig. 2.12), it is evident that the most used interface in performing the three navigation tasks resulted to be the *keyboard* one. The *point-and-click* interface was slightly

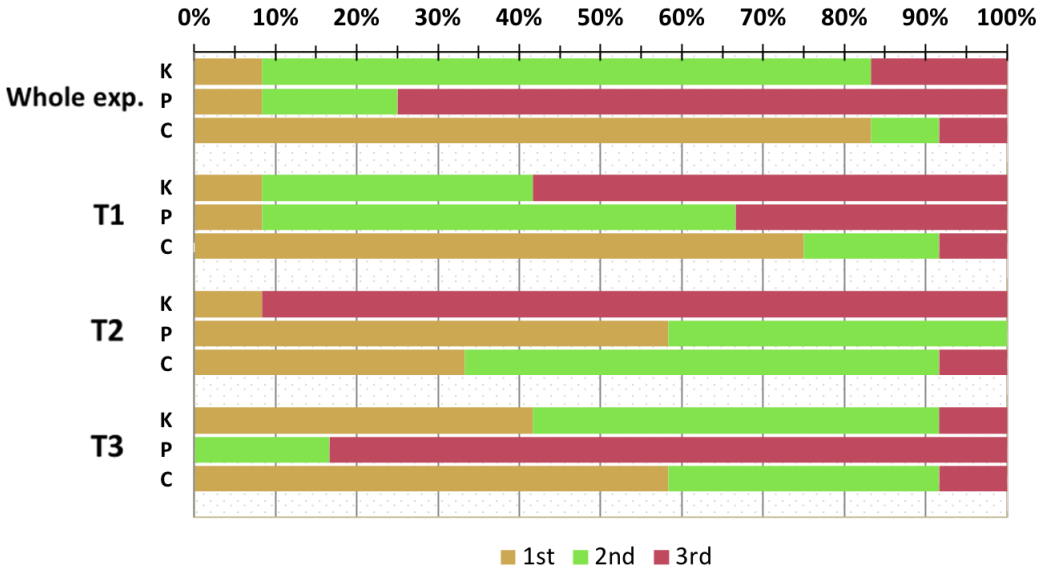


Figure 2.16: First user study: number of times the *keyboard* teleoperation (K), *point-and-click video navigation* (P) and *combined* (C) interfaces have been ranked 1st, 2nd and 3rd for the execution of the whole experiment and individual tasks.

preferred in the execution of *T1* and largely preferred in the execution of *T2*. This is reasonably due to the fact that, in these tasks, the robot had to move over long distances and the autonomous navigation capabilities could effectively limit the participants' MW, by requesting them to simply click on the destination to reach.

It is also worth noting that most of the concerns regarding the use of the *point-and-click* interface were related to the specific camera FOV, which often did not allow the user to immediately spot the destination to click. This limitation could be addressed by adding different camera configurations as well as the possibility for the various interfaces to control the pan-tilt of the camera.

Second User Study: Which Camera Configuration?

By moving from the results obtained in the first study, this section reports on a second user study that was designed to investigate whether the introduction of different camera configurations, as well as of a pan-tilt camera, may influence users' SA and their performance in remote robot navigation tasks.

Participants involved in the study (8 males and 2 females) aged between 24 and 32 years (M = 26.60 SD = 2.32), were recruited among university students from Politecnico di Torino.

At the beginning of the experiment, participants were provided with a brief training on the use of the teleoperation interface for controlling the robot considered in the study. In particular, the *combined* interface was experimented in this case by leveraging

the findings obtained in the first study. Afterwards, participants were invited to perform the navigation tasks *T1*, *T2* and *T3* (Fig. 2.10) in sequence, by using all the camera configurations. To compensate for possible learning effects, the latin square random order was used to select the camera configurations.

During the experiments, quantitative data about the number of navigation commands (key presses and/or mouse clicks) and time needed by the participants to complete the tasks were measured. For each camera configuration tested, participants were requested to evaluate their MW and SA by compiling a NASA-TLX [78] and a NASA Situation Awareness Rating Technique (SART) [81] questionnaire, respectively. More in detail, the first questionnaire evaluated the MW by exploiting the same dimension scale and weighting mechanism used in the first study. The second questionnaire assessed participants' SA on a 7-point scale regarding the *understanding of the situation* (information quantity, and information quality), the *demand of attentional resources* (complexity, variability, and instability of the situation), and the *supply of attentional resources* (division of attention, arousal, concentration, and spare mental capacity). Like for the first questionnaire, a global score was then calculated according to [81].

After completing the test, each participant was also asked to evaluate his or her experience with the different camera configurations through a usability questionnaire split in three parts.

The first part was designed by considering the Usefulness, Satisfaction, and Ease of use (USE) questionnaire [82]. USE requested participants to express their agreement with a number of questions/statements on a 5-point Likert scale (Table 2.5).

Table 2.5: Second user study: selection of statements in the USE questionnaire used for the subjective evaluation.

Evaluated Aspect	**Question/Statement**
Ease of use	
Q1	The system is easy to use
Q2	The system is simple to use
Q3	The system is user friendly
Satisfaction	
Q4	The system is pleasant to use
Q5	The system works the way I want it to work
Q6	The system is fun to use
Q7	I am satisfied with the system

The second part was designed by considering the (adapted) SASSI methodology [80] used in the first study. In the same way, scores for *cognitive demand* and *annoyance* usability factors were inverted (thus, higher scores have to be interpreted as being more positive).

The third part was created, as for the first study, with the aim to define a ranking between the three camera configurations according to the preferences expressed by participants both for the three individual tasks as well as for the task as a whole.

Data collected from the study were then analyzed using a one-way repeated measures ANOVA test (significance level of 0.05) and a two-tailed paired t-test (significance level of 0.05) in post-hoc analysis, in order to detect any overall differences between the three configurations and highlight exactly where these differences were actually occurring.

Objective data, in terms of task completion time and number of interactions for accomplishing the tasks, are reported in Figure 2.17 and Figure 2.18, respectively.

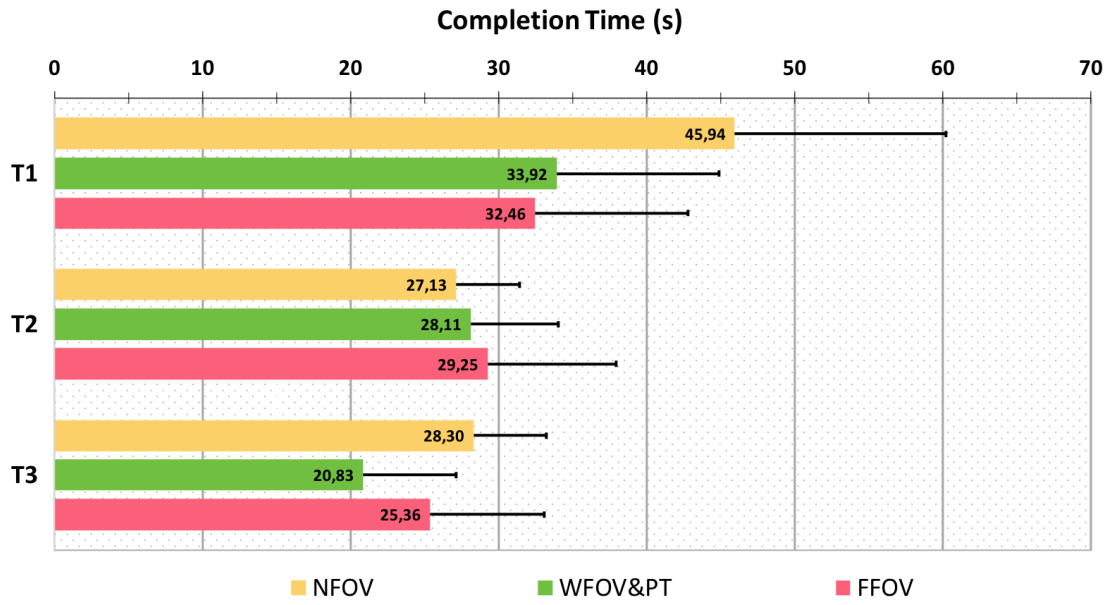


Figure 2.17: Second user study: results in terms of completion time required to complete the three tasks with each camera configuration. Bar lengths report average values (lower is better), whereas whiskers report standard deviation.

From Figure 2.17, it is clearly evident that participants using the *NFOV* configuration took more time to complete the task *T1* (ANOVA: $p = 2.67 \times 10^{-2}$) compared to the *WFOV&PT* and *FFOV* configurations. Similarly, completion time for task *T3* with the *NFOV* configuration was (slightly)higher compared to the (*FFOV*) *WFOV&PT* one. Post-hoc analysis, as illustrated in Table 2.6, validated the above observations by showing on one hand no statistically significant differences in *T2* and between the *WFOV&PT* and *FFOV* configurations. On the other hand, significant differences were found between the *NFOV* and *WFOV&PT* configurations both in *T1* and *T3*, whereas differences between *NFOV* and *FFOV* were statistically significant only in *T1*.

Concerning the number of interactions, it can be observed from Figure 2.18, that the *NFOV* configuration required a higher number of navigation commands compared to the *WFOV&PT* and *FFOV* ones, both in *T1* (ANOVA: $p = 8.91 \times 10^{-3}$) and *T3* (ANOVA: $p = 3.72 \times 10^{-3}$). Statistical significance validated by post-hoc analysis (Table 2.7) confirmed

Table 2.6: Second user study: post-hoc analysis on completion time results and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	NFOV vs. WFOV&PT	WFOV&PT vs. FFOV	NFOV vs. FFOV
T1	t[9] = 2.27, $p = 4.97 \times 10^{-2}$ (+)	t[9] = 0.42, $p = 0.6873$	t[9] = 2.34, $p = 4.40 \times 10^{-2}$ (+)
T2	t[9] = -0.43, $p = 0.6741$	t[9] = -0.44, $p = 0.6717$	t[9] = -0.61, $p = 0.5545$
T3	t[9] = 2.63, $p = 2.71 \times 10^{-2}$ (+)	t[9] = -1.24, $p = 0.2473$	t[9] = 1.03, $p = 0.3307$

a significant difference between the *NFOV* and both *WFOV&PT* and *FFOV* configurations in tasks *T1* and *T3*. However, no statistically significant differences were found between the *WFOV&PT* and *FFOV* configurations, expect for the task *T2*, where participants were able to complete the task by issuing a lower number of commands with the *FFOV* compared to *WFOV&PT*.

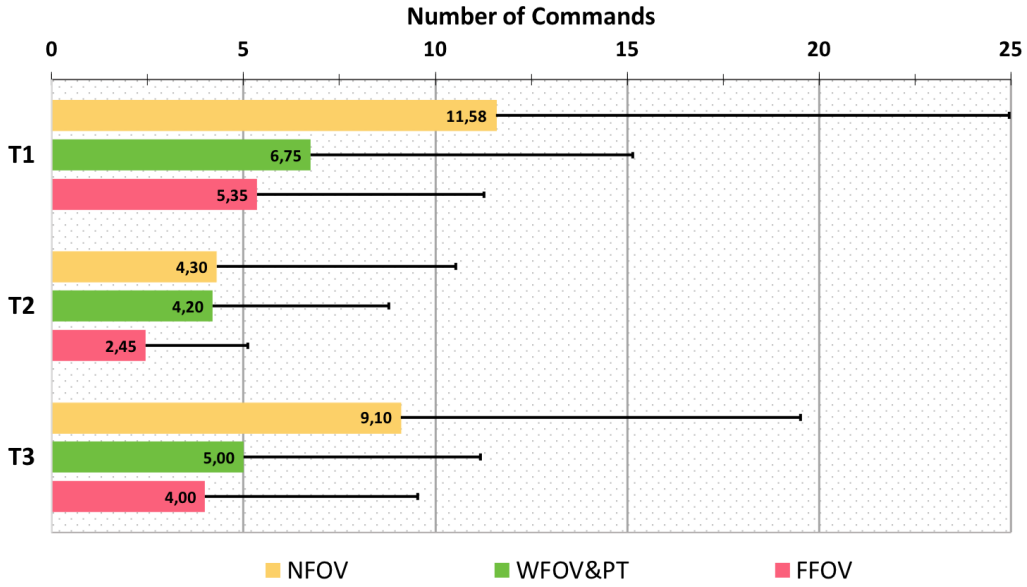


Figure 2.18: Second user study: results in terms of number of interactions required to complete the three tasks with each camera configuration. Bar lengths report average values (lower is better), whereas whiskers report standard deviation.

Data collected through the NASA-TLX and SART questionnaires are illustrated in Figure 2.19 and Figure 2.20, respectively.

Obtained results appear to describe an almost comparable situation. In fact, the *NFOV* was judged by participants as the most cognitive demanding configuration, followed by the *WFOV&PT* and then the *FFOV*. Moreover, as illustrated in Figure 2.20, the *WFOV&PT* was judged by participants as the configuration providing the highest SA. Therefore, it is possible to conclude that *NFOV* was evaluated as the most challenging

Table 2.7: Second user study: post-hoc analysis on number of interactions and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	NFOV vs. WFOV&PT	WFOV&PT vs. FFOV	NFOV vs. FFOV
T1	t[9] = 2.30, $p = 4.73 \times 10^{-2}$ (+)	t[9] = 0.71, $p = 0.4945$	t[9] = 4.37, $p = 1.80 \times 10^{-3}$ (++)
T2	t[9] = 0.09, $p = 0.9305$	t[9] = 3.57, $p = 6.02 \times 10^{-3}$ (++)	t[9] = 1.62, $p = 0.1407$
T3	t[9] = 3.01, $p = 1.46 \times 10^{-2}$ (+)	t[9] = 0.81, $p = 0.4409$	t[9] = 3.39, $p = 7.97 \times 10^{-3}$ (++)

configuration, as it worsened participants' awareness of the operating conditions.

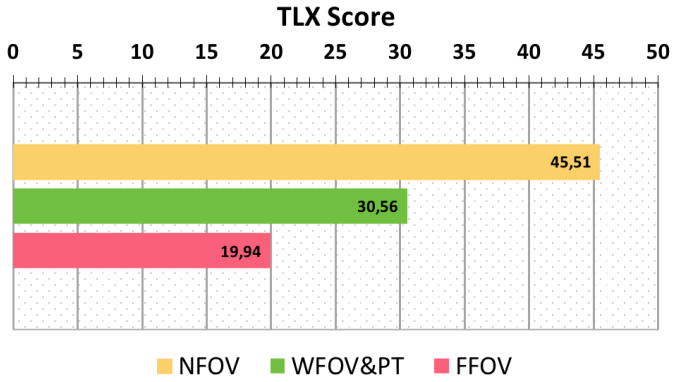


Figure 2.19: Second user study: results concerning MW measurements for the three camera configurations. Bar lengths report TLX score (lower is better).

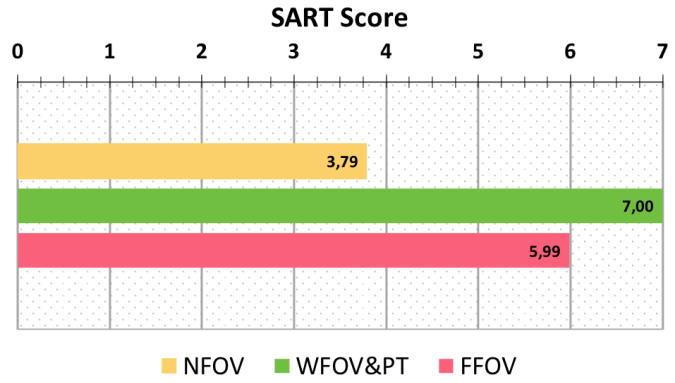


Figure 2.20: Second user study: results concerning SA measurements for the three camera configurations. Bar lengths report SART score (higher is better).

Similar considerations can be made for all the usability factors tackled by the USE and (adapted) SASSI questionnaires. In fact, as shown in Figure 2.21 and Figure 2.22, the

configurations leveraging a wider FOV (*WFOV&PT* and *FFOV*) were more positively evaluated by participants compared to the *NFOV* configuration.

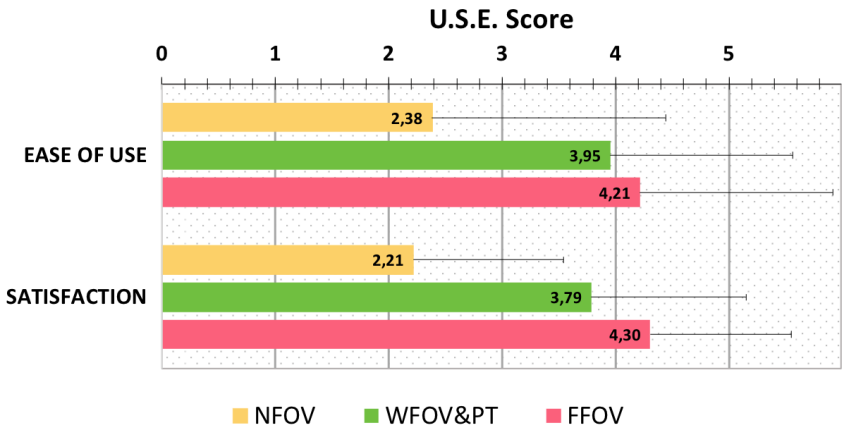


Figure 2.21: Second user study: results concerning the usability of the three camera configurations for the task as a whole based on USE questionnaire. Bar lengths report average values (higher is better), whereas whiskers report standard deviation.

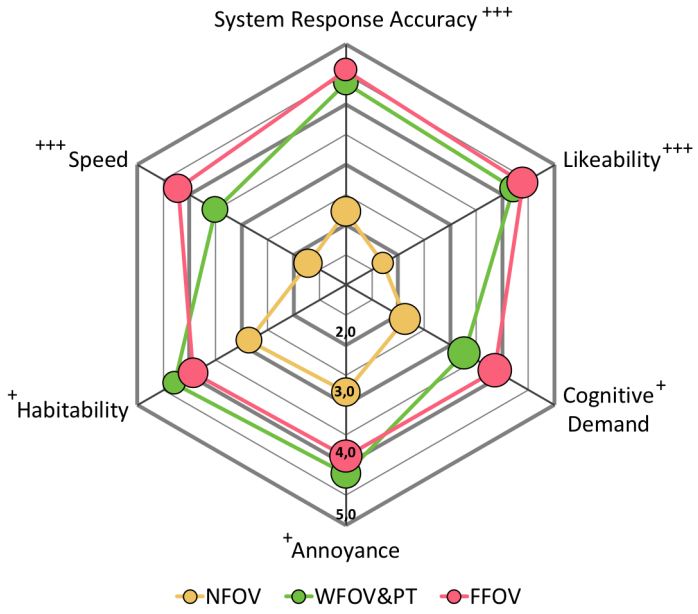


Figure 2.22: Second user study: results concerning the usability of the three camera configurations for the task as a whole based on adapted SASSI. Circle position reports average values (higher is better), circle dimension reports standard deviation, whereas + symbols report statistical significance determined with the ANOVA tests (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

In particular, as illustrated in Figure 2.21, the *WFOV&PT* and *FFOV* were judged the more easily usable (ANOVA: $p = 9.95 \times 10^{-3}$) and satisfactory (ANOVA: $p = 6.70 \times 10^{-5}$) configurations. In the same way, participants' evaluations gathered through the (adapted) SASSI methodology showed statistical significance differences for every usability factor (+ symbols report ANOVA tests results, i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

Post-hoc analysis highlighted no statistically significant differences between *WFOV&PT* and *FFOV* configurations for all the usability factors (Table 2.8 and Table 2.9). However, differences between the *NFOV* and both the *WFOV&PT* and *FFOV* configurations were found to be significant for most usability factors, except for the *annoyance* and *habitability* ones. Moreover, it can be noticed that the perceived *speed* factor was characterized by higher scores for wider FOVs, thus confirming findings obtained in [56] (where it is stated that with wider FOVs, navigation speed tends to be perceived as increased because of the scene compression). It is worth observing also that the CD factor corroborated results obtained by the NASA-TLX methodology for the MW assessment.

Table 2.8: Second user study: post-hoc analysis on USE results and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

	NFOV vs. WFOV&PT	**WFOV&PT vs. FFOV**	**NFOV vs. FFOV**
Ease of Use	t[9] = -3.44, $p = 7.34 \times 10^{-3}$ (++)	t[9] = -0.48, $p = 0.6448$	t[9] = -2.67, $p = 2.54 \times 10^{-2}$ (+)
Satisfaction	t[9] = -4.91, $p = 8.41 \times 10^{-4}$ (+++)	t[9] = -1.49, $p = 0.1692$	t[9] = -4.78, $p = 9.94 \times 10^{-4}$ (+++)

Table 2.9: Second user study: post-hoc analysis on SASSI results and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

	NFOV vs. WFOV&PT	**WFOV&PT vs. FFOV**	**NFOV vs. FFOV**
System Response Accuracy	t[9] = -7.61, $p = 3.30 \times 10^{-5}$ (+++)	t[9] = -0.60, $p = 0.5598$	t[9] = -6.13, $p = 1.73 \times 10^{-4}$ (+++)
Likeability	t[9] = -8.72, $p = 1.10 \times 10^{-5}$ (+++)	t[9] = -0.29, $p = 0.7751$	t[9] = -5.29, $p = 5.03 \times 10^{-4}$ (+++)
Cognitive Demand	t[9] = -2.50, $p = 3.41 \times 10^{-2}$ (+)	t[9] = -0.91, $p = 0.3884$	t[9] = -2.48, $p = 3.51 \times 10^{-2}$ (+)
Annoyance	t[9] = -2.48, $p = 3.52 \times 10^{-2}$ (+)	t[9] = 0.94, $p = 0.3726$	t[9] = -1.77, $p = 0.1102$
Habitability	t[9] = -5.07, $p = 6.71 \times 10^{-4}$ (+++)	t[9] = 0.65, $p = 0.5290$	t[9] = -1.75, $p = 0.1149$
Speed	t[9] = -4.29, $p = 2.01 \times 10^{-3}$ (++)	t[9] = -1.50, $p = 0.1678$	t[9] = -4.77, $p = 1.02 \times 10^{-3}$ (++)

Data about participants' preferences in using the three camera configurations, both for the whole experiment as well as for the individual tasks, are illustrated in Figure 2.23. By taking into account the overall rankings, results indicate a clear preference for the *FFOV* and *WFOV&PT* configurations over the *NFOV* one. This result holds also for task *T1*.

Considering the other two tasks, the *FFOV* appears to be the favorite configuration compared to *WFOV&PT* and *NFOV*. Based on the feedback gathered during the tests, preferences seemed to be mainly motivated by the fact that, as expected, wider FOVs allowed participants to frame larger portions of the environment wherein the robot was moving. By digging more in detail, the *FFOV* configuration was strongly preferred in the execution of task *T2* compared to task *T3*. This finding was reasonably due to the fact that, when obstacles had to be avoided, the wider FOV made it more suited to work with the semi-autonomous *point-and-click* teleoperation interface, since users could see more easily the point to click in which they wanted to make the robot move.

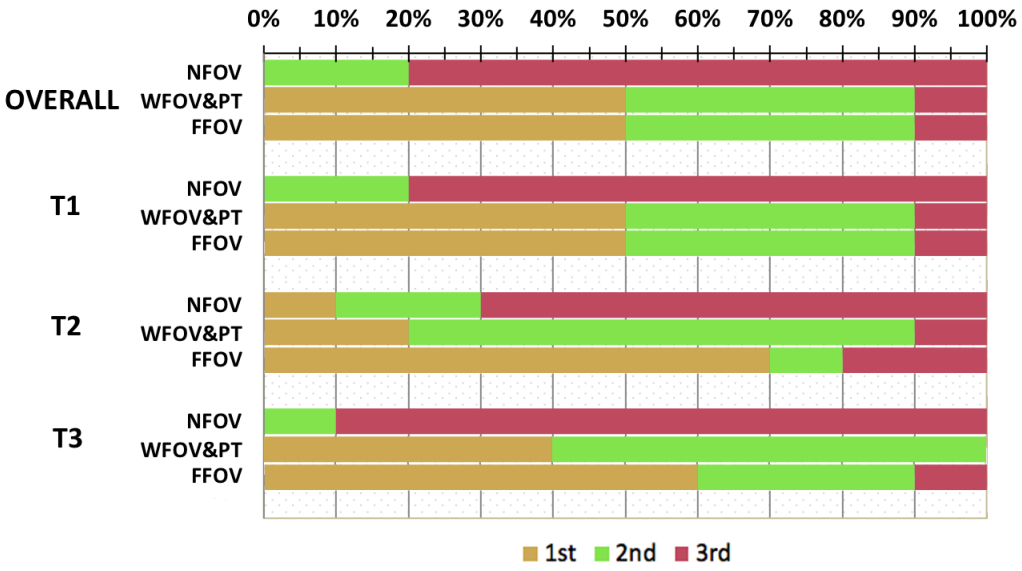


Figure 2.23: Second user study: number of times the three camera configurations were ranked 1st, 2nd and 3rd for the execution of the task as a whole (overall) and for the execution of individual tasks.

In conclusion, by combining the above results with number of interactions (Fig. 2.18) and completion times (Fig. 2.17), it can be observed that the *WFOV&PT* and *FFOV* configurations allowed participants to accomplish the tasks in less time by issuing fewer commands. The above observations were also confirmed by results regarding MW (Fig. 2.19), where the *FFOV* was evaluated as the less cognitive demanding configuration compared to the other two. It is also worth noting that this result may be also due to the fact that, in the *WFOV&PT* configuration, the pan-tilt capability of the camera required participants to issue a higher number of commands (to move the camera) than with the *FFOV* configuration. Concerning the participants' SA of the robot's surrounding environment (Fig. 2.20), the *WFOV&PT* was evaluated as the most effective configuration. This finding was reasonably due to the fact that the pan-tilt function of the camera allowed participants to better explore the environment without necessarily having to

move the robot. Lastly, the radial distortions in the *FFOV* configuration were evaluated by participants as altering the perception of the environment on the robot's sides.

2.2 Robotic Aerial Traffic Monitoring

The robotic aerial traffic monitoring domain (Fig. 2.24) was explored as a possible representative example of applications featuring a *supervisory control interface* and with the human user in the role of *supervisor* in a remote spatial proximity pattern. The next sections will present relevant works pertaining this research area as well as the AA framework developed in this domain to assist a UAV controller in UAV monitoring tasks by providing different LOAs. An overview of the UIs exploited in this study as well as the methodology adopted to perform the experimental tests will be described, together with experimental results.

Figure 2.24: Robotic aerial traffic monitoring concept.

2.2.1 Background

In recent years, the field of aerial service robotics applications has received significant attention in the scientific research community for the development of UAVs with some autonomous capabilities. Examples of UAVs applications are traffic monitoring, disaster response, surveillance, first aid, freight transport, etc. [83].

Today, however, these vehicles are often used in high uncertainty environments and dynamic contexts characterized by parameter disturbances and unpredictable failures. For this reason, a totally autonomous control system has not emerged yet [84]. Thus,

supervisory systems, including humans in the control loop, are needed to monitor UAV operations and assist controllers when critical situations occur [85, 86].

In these kinds of environments, supervisory systems exhibiting different LOAs and equipped with decision-making capabilities may be exploited to dynamically assign tasks either to human users or to the system itself by distinguishing situations in which the skills of operators are sufficient to perform a given task from situations where suggestions or system interventions may be requested [87, 88, 89]. The ability of the system to dynamically adapt the LOA based on the considered context is generally defined as "*adjustable or sliding autonomy*" [36].

The LOA required by the system can be determined in many ways, including the evaluation of the operators' performance and their cognitive load when performing UAVs monitoring tasks. Several studies have demonstrated the advantages of employing dynamic tasks allocation between humans and machines for handling the operator's cognitive load and keep him or her focused on control loops [90, 91].

In this regard, many works have studied the evaluation and classification of the UAV operator's cognitive load. In the literature, different techniques have been proposed, historically classified in three categories: subjective, physiological and performance-based [92]. Subjective measurements are used to assess the MW perceived by humans through rankings or rating scales. Physiological measurements are workload assessment techniques based on the physical response of the body. Objective or performance-based measurements are used to assess the ability of humans to carry out a given task.

Concerning subjective measures, in [93] and [94], the cognitive load perceived by a human operator was measured using the NASA TLX questionnaire [95] in gaze-writing and robotic manipulation tasks, respectively. The authors of [88] studied how self-assessed CD may have an influence on simulated supervision activities. However, although these measures represent a reliable method for assessing humans' cognitive load [78], they have the disadvantage of asking users annoying or repetitive interactions for filling out questionnaires or evaluation scales.

In parallel to these studies, other works focused on physiological measurements as techniques for assessing the cognitive load of the operator in real-time. For instance, in [96], the EEG power band ratios were used as an example of CD measurement in adaptive automation. Similarly, in [97], EEG channels, electrooculographic (EOG), electrocardiographic (ECG) and respiratory inputs were exploited as methods for CD evaluation in air traffic control tasks together with an artificial neural network (ANN) as a classification methodology. Also in [98], the electromyogram (EMG), ECG, skin conductance (SC), respiration and reaction time (RT) inputs were used to assess pilots' MW changes in flight simulation tasks and a BN was exploited as a classification methodology. In addition, Magnusson in [99] exploited pilots' heart rate (HR), heart rate variability (HRV) and eye movements in both flight simulations and real flight conditions. In [100], the relationship between EEG and RT was studied to categorize the degree of performance during a cognitive task, in order to anticipate human errors. Although these studies provided evidence that combining more than one physiological measures can

improve the accuracy of assessment and classification of workload [97, 101], these approaches proved to be very uncomfortable for real application scenarios because of the use of bulky and expensive equipment[102, 103]. Data about the suitability of alternative devices in physiological measurements are actually needed to adequately support next advances in the field. Some activities in this direction have already been performed. For example, in [104] the authors showed how a small device, namely, a consumer-level EEG headset developed by EMOTIV, can be successfully used to measure MW together with a SVM in a simple n-back memory task.

2.2.2 AA Framework

Based on the reviewed works discussed in Section 2.2.1, it can be observed that when human users are involved in UAV monitoring operations, their level of attention and consequently their MW in the execution of these tasks can be used to build and train AA systems able to assist them by inferring the appropriate LOA. In this context, the panorama of MW assessment measurements and of employed learning models is quite heterogeneous. For this reason, by taking into account advantages and drawbacks of the solutions discussed in Section 2.2.1, an AA system featuring three different MW evaluation techniques (i.e., subjective, objective and physiological) and two different learning models (i.e., BN and SVM) was developed.

The AA framework developed in this thesis is illustrated in Fig.2.25. It is composed of three main blocks: *UAVs Simulator* (left), *Bandwidth Simulator* (right) and *Control Tower* (down).

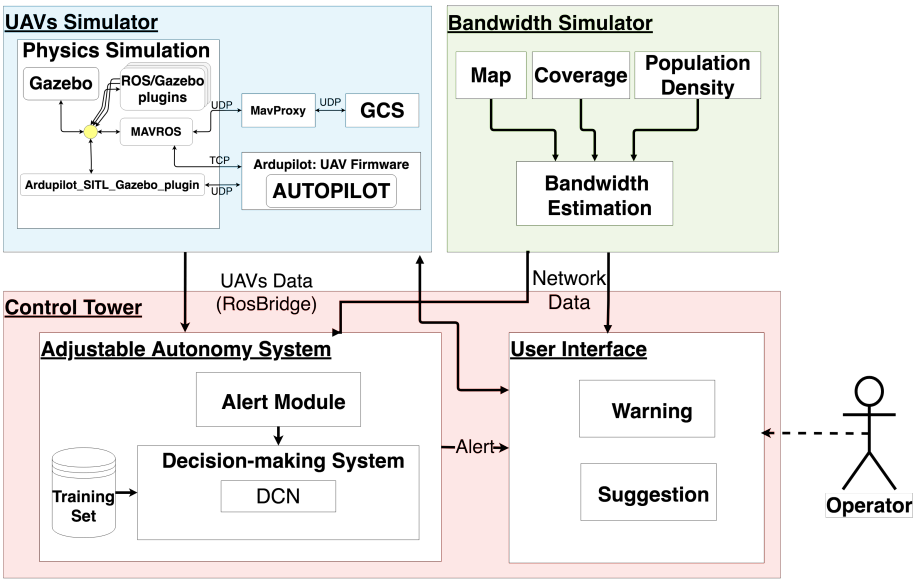


Figure 2.25: AA framework.

The *UAVs Simulator* is the module devoted to performing the UAVs flight simulation within a 3D urban environment. It is made up of the three different blocks described below.

- *Autopilot* is the module consisting of the software required by UAVs for flying stable both in real and simulated conditions. Specifically, the Software-In-The-Loop (SITL)[4] simulator was used for running the UAV PX4 Autopilot Flightcode[5] without requiring any specific hardware. In fact, the un-compiled autopilot code, which normally runs on the UAV's onboard computer, is compiled, simulated and executed by the SITL simulation software itself.
- *Physics Simulation* is the block devoted to reproducing the real world physics of UAVs' flight. In this specific case, the real-time physics engine Gazebo and the ROS framework were used for replicating in 3D the real models of UAVs, emulating their physic properties and constraints as well as their sensors (e.g., laser, cameras, etc.) in the 3D simulation environment (Fig. 2.26).

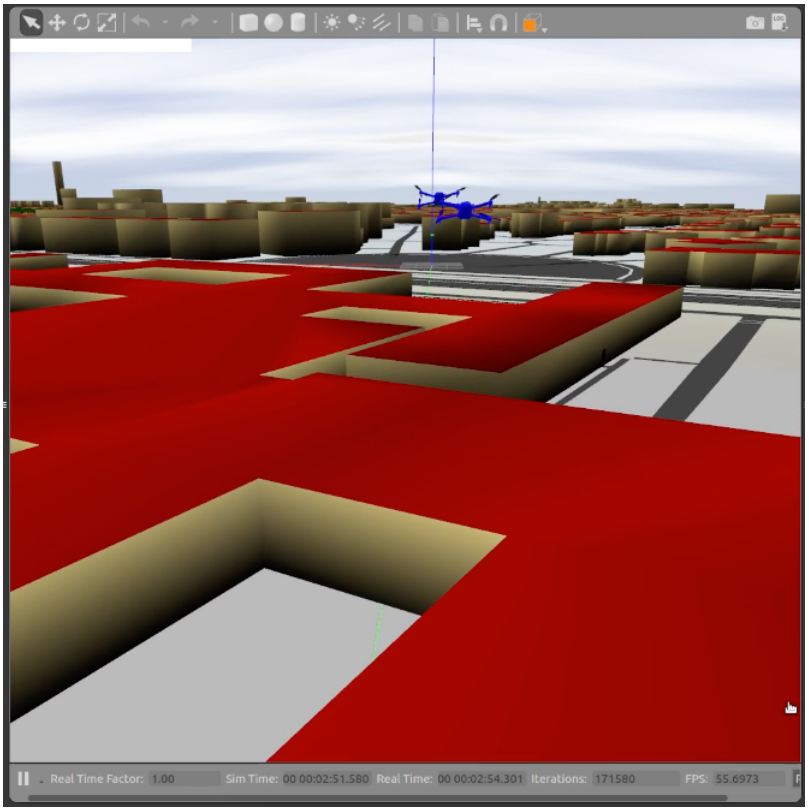


Figure 2.26: Physics simulation.

[4]http://ardupilot.org/dev/docs/sitl-simulator-software-in-the-loop.html

[5]https://px4.io

- *Ground Control Station (GCS)* is the module containing the software used to define the UAVs' starting positions (GPS coordinates), gather flight information in real-time, plan and perform UAVs' flight missions (Fig. 2.27). The communication between the GCS module and the PX4 Autopilot Flightcode is provided by the Micro Air Vehicle ROS (MAVROS) node with the MAVLink communication protocol. As illustrated in Figure 2.25, the MAVProxy module plays the role of the intermediary node between the GCS and UAVs.

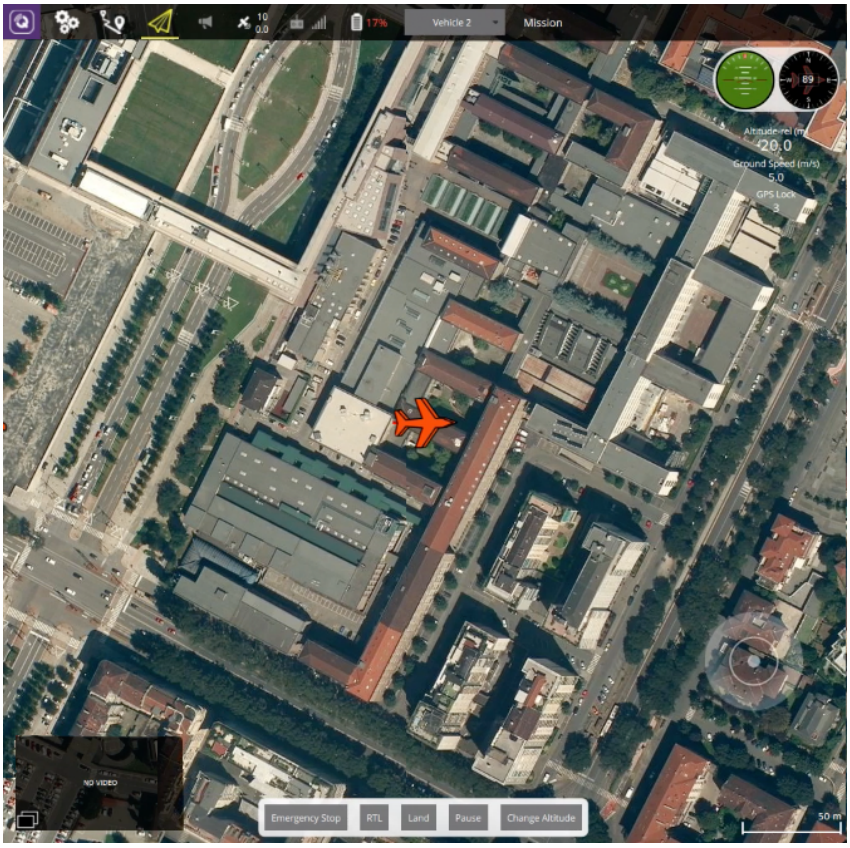


Figure 2.27: Ground control station.

As illustrated in Figure 2.25, this module is also responsible for providing UAVs' status data, like battery level, later abbreviated *b* and distance from obstacles (e.g., buildings), later abbreviated *o*, to the *Control Tower* module by means of the RosBridge Protocol[6]. More specifically, as listed in Table 2.10, this data is associated with a numeric value in the range [1;3] – where 1 is *"Low"*, 2 is *"Medium"* and 3 is *"High"* – and collected by the *Alert Module* (described below) to determine the status of each UAV.

[6]https://wiki.ros.org/rosbridge_suite

Table 2.10: UAVs' information association to variables.

Input Variables	Description	Variables/Numeric Values
o	UAV's distance from an obstacle	Low=[5-25]m; Medium=[25-50]m; High=[50-100]m
b	UAV's battery level	Low=[0-20]%; Medium=[21-60]%; High=[61-100]%

The *Bandwidth Simulator* is the module responsible for reproducing the Network Transmission Rate (NTR or n) of the simulated urban environment. This module was devised and developed by considering the fact that UAVs communicate or send information through the network. Hence, low NTR could lead to critical conditions for UAV controllers. In this thesis, the NTR was assumed to rely on two different variables, namely, the network coverage and the population density of the urban sites (parks, stadiums, schools, etc.). As illustrated in Figure 2.28, a grid was created on the map (depicting the considered urban environment) by storing in each cell the coverage and population density values for computing the NTR.

Figure 2.28: NTR grid map.

A linear polynomial function z (2.1) of the above values was used to calculate the NTR for each cell. Three different values, in the range [1;3] – where 1 is *"Low"* (red cells on the map), 2 is *"Medium"* (yellow cells on the map) and 3 is *"High"* (green cells on the map) – were defined to describe the network coverage and the population density levels

based on OpenSignal[7] and daily time slots data, respectively. As illustrated in Figure 2.25, these network data are sent to the *Control Tower* module in order to determine the NTR according to the position of each UAV on the map.

$$Bandwidth = \begin{cases} High & if \quad z < 0.5 \\ Medium & if \quad 0.5 \geq z < 1.5 \\ Low & if \quad z \geq 1.5 \end{cases} \tag{2.1}$$

The *Control Tower* module consists of the *Adjustable Autonomy System* (AAS) and the *User Interface* (UI). The AAS is made up of two components, i.e., the *Alert Module* and *Decision-making System*.

The *Alert Module* is responsible for computing the UAV's level of risk (later referred to as *Alert*) by exploiting the mathematical formula described in (2.2), where b and o represent the two inputs listed in Table 2.10 and n is the NTR. Moreover, y represents the UAV's level of risk which is determined as described in (2.3). It can be observed in (2.2) that when the value of one of the variables is *"Low"* (i.e., 1 in the numeric range [1;3]), the *Alert* assumes the *"Danger"* value. When the values of the variables increase, then the *Alert* decreases from *"Danger"* to *"Safe"* through the *"Warning"* level.

$$y = \frac{1}{b-1} * \frac{1}{o-1} * \frac{1}{n-1} \tag{2.2}$$

$$Alert = \begin{cases} Danger & if \quad b = 1 \quad \vee \quad o = 1 \quad \vee \quad n = 1 \\ Warning & if \quad 0.15 < y < 1.5 \\ Safe & if \quad y \geq 1.5 \end{cases} \tag{2.3}$$

The *Decision-making System* represents the core of the devised framework, devoted to infer the appropriate LOA by exploiting both the operator's MW and mission's outcomes based on the number of UAVs in critical situations (i.e., a *"Danger"* level of risk). Three different LOAs were proposed in this context, i.e., "*Warning*", "*Suggestion*" and "*Autonomous*". In the "*Warning*" LOA, the system warns the controller if critical situations occur; the "*Suggestion*" LOA suggests feasible actions to him or her; the "*Autonomous*" LOA monitors and performs actions autonomously without any human intervention.

[7] https://opensignal.com

2.2.3 Supervisory Control Interfaces

The UI developed in this thesis shows a 2D map of the considered urban environment for displaying UAVs' position and useful information for the human controller. A wide region of the human controller's screen is occupied by the 2D map of the environment in which UAVs are displayed in real-time. A colored marker on the map is used to indicate both the location of each UAV (GPS position) and its current *"Alert"* (Fig.2.29).

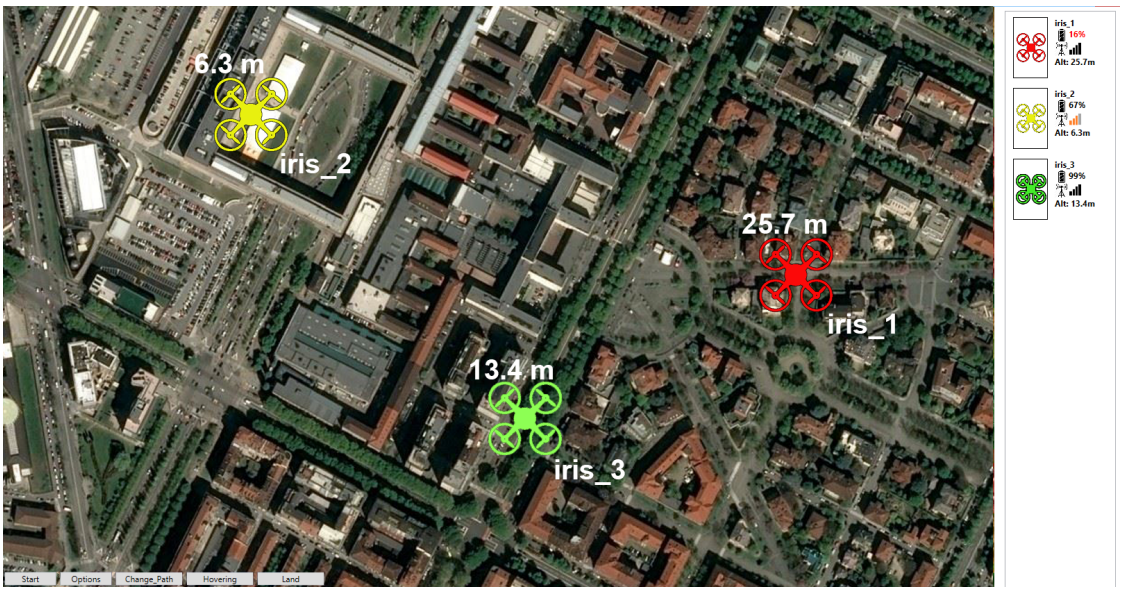


Figure 2.29: User interface.

Three different colors are employed to depict the UAV's level of risk: green (*"Safe"*), yellow (*"Warning"*) and red (*"Danger"*). UAV's marker color changes from green to red based on the linear interpolation formula described in (2.2). A detailed summary of data about each UAV is displayed on the panel in the right side of the interface (Fig.2.30).

For each drone, the UI reports its unique name, its battery level, the bandwidth coverage of the area around its position and its flying altitude. Different flight commands are integrated and displayed by the UI through five controls buttons, right below the map, for allowing human operators to take control of the selected UAV (Fig. 2.31).

Specifically, the "*Start*" button is used to start the UAVs simulation, whereas the "*Options*" button is used to either show/hide the NTR grid on the map (Fig. 2.28) or the UAVs' flight paths. The other three buttons, i.e., "*Hovering*", "*Land*" and "*Change_Path*" are used by the human controller to hover, land or change the UAV's flight mission by indicating the next waypoint with respect to the UAV's current location. The number or type of these commands dynamically change according to the current LOA determined by the AAS, defining in this way the "*Warning*", "*Suggestion*" and "*Autonomous*" interfaces.

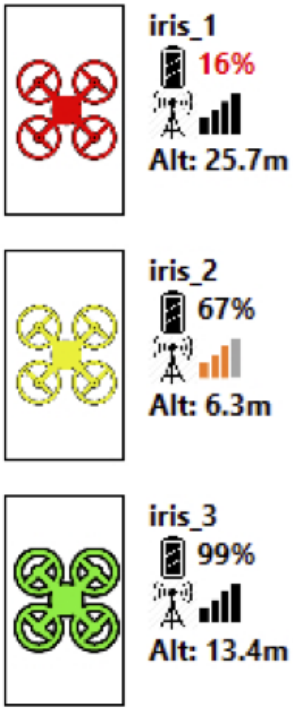


Figure 2.30: UAVs data summary.

Figure 2.31: Buttons for controlling selected UAV and displaying associated information.

In the "*Warning*" interface (Fig. 2.29), the human controller can take direct control of the UAV via the "*Land*", "*Hovering*", and "*Change_Path*" buttons.

On the contrary, in the "*Suggestion*" interface, the human controller can only select actions among those suggested by the AAS (according to Table 2.11) and displayed in the summary panel (Fig.2.32). The replanning operation provides an alternative flight path starting from the current location of the UAV until to its target destination via the Bing Map REST API[8] and a "route planning" request.

In the "*Autonomous*" interface, all the flight commands are disabled and the AA system implements one of the possible actions according to UAVs' *Alerts* and rule-based decisions. For instance, considering the drone's battery level, the AA system can perform one of the following actions:

- Return To Launch (RTL) if the residual battery charge is sufficient to let the UAV return to the starting point (an estimation is performed based on the battery level

[8]https://msdn.microsoft.com/it-it/library/ff701713.aspx

Table 2.11: System-suggested actions for each UAV.

Alert's Input Variables	Possible Values	Feasible Actions
o	$Medium \vee Low$	Hovering, Replanning, Land
b	$Medium \vee Low$	Land, Return To Launch (RTL)
n	$Medium \vee Low$	Replanning, Land

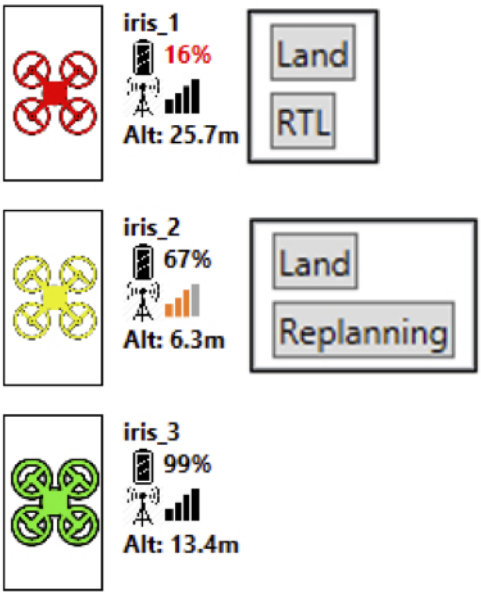


Figure 2.32: Examples of actions suggested by the system for each UAV.

degradation curve and the distance to be traveled);

- land in the nearest emergency landing area, defined as an obstacle-free landing zone, if the residual battery charge is sufficient to let the UAV fly until there;
- land "on-site" when the residual battery charge is too low.

2.2.4 Experimental Results

As anticipated, the goal of this study is to build an AA system exploiting decision-making capabilities able to assist the UAV operator by predicting MW changes or overload when the number of UAVs to be monitored increases significantly. To this aim, different MW assessment techniques were exploited to train different classification models to infer the appropriate LOA in drone-traffic-control tasks.

Details about CD assessment techniques and classification/learning models are provided in the paragraphs below.

MW assessment techniques In this domain, three different MW assessment techniques (one per each class discussed in Section 2.2.1) were explored. Specifically, the NASA-TLX questionnaire [95], was used as representative for the subjective measures class, the EEG signals for the physiological class and the human controller's outcomes in UAV monitoring tasks for the objective or performance-based class. Details of the aforementioned techniques are described below.

- *Subjective Measurement Technique*: as mentioned above, the NASA-TLX questionnaire [95] was exploited in this class to assess the operator's MW. In particular, the software reported in [105] was used. As illustrated in Figure 2.33, this software provides a score for the human operator's MW according to a weighted average of six subscales, i.e., *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort*, and *frustration*.

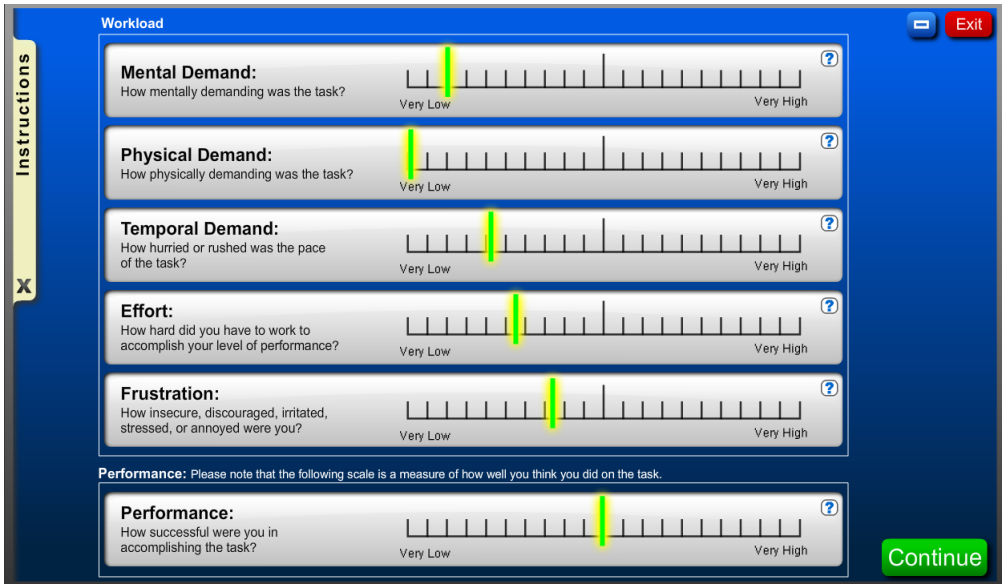


Figure 2.33: Screenshot of the NASA-TLX online software.

- *Physiological Measurement Technique*: in this class, the EEG signals were exploited to measure the operator's MW. In particular, a wireless Brain Computer Interface (BCI) device manufactured by Emotiv and named EMOTIV Epoc+®headset (Fig. 2.34), was used by building on the results reported in [104]. As illustrated in Figure 2.35, the headset is made up of 14 wireless EEG signals acquisition channels at 128 samples/s. The recorded EEG signals are sent to an USB dongle for transmitting the gathered information to the host workstation. In addition, Emotiv provides a subscription software, named Pure·EEG, devoted to getting the raw EEG data as well as the dense spatial resolution array containing data at each sampling interval.

Figure 2.34: EMOTIV Epoc+®headset.

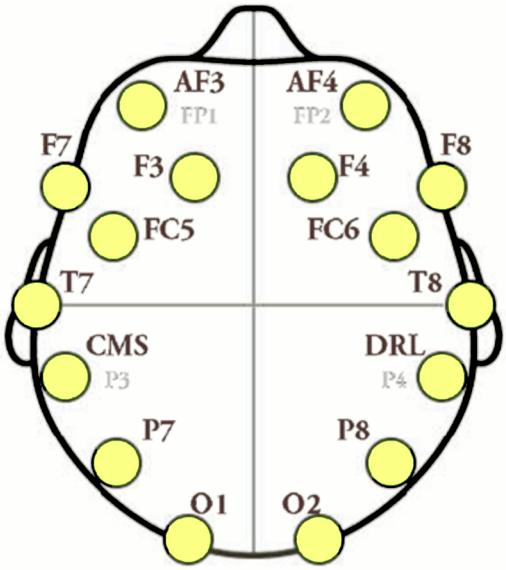


Figure 2.35: Positions of the 14 wireless EEG signals acquisition channels in the EMOTIV Epoc+®headset.

- *Objective or Performance-based Measurement Technique*: in this class, the UAVs controller's performance was used to assess the MW. To this aim, six different tasks were experimented in this thesis, later referred to as *MT1*, *MT2*, *MT3*, *MT4*, *MT5* and *MT6*. *MT1* consisted of a single UAV flying along an obstacle-free path, in order to not require operator's action to successfully complete the task. *MT2* and *MT3* were designed to assess the operator's performance in monitoring two UAVs flying in a medium NTR zone and at risk of colliding, respectively. In particular, collisions that may occur for each UAV, in *MT3*, were specifically designed to be distant over time for allowing the operator to be virtually able to handle

them while keeping the effort to complete the task relatively low. *MT4* consisted of three UAVs, one with a medium battery level and two of them at risk of colliding. *MT5* consisted of five UAVs, three of them at risk of colliding. This mission was intentionally designed to be very difficult to complete even though theoretically still manageable. Lastly, *MT6* consisted of six UAVs, each of them requiring the operator's interventions to successfully complete the task. This mission was intentionally designed to be hard to complete. Such tasks have been specifically devised to evaluate the operator's performance in the possible scenarios he or she could be involved into in air-traffic management. Furthermore, the outcome of each mission may be *"failed"* when at least one UAV crashed or *"successfully completed"* when all UAVs landed in the intended positions. It is worth noting that the number of UAVs (later abbreviated as #UAVs) in each mission was defined relying on a preliminary experiment. Specifically, users were invited to monitor from 1 to 6 UAVs characterized by a level of risk linearly proportional to the #UAVs. Obtained results showed that the #UAVs to monitor could be divided into three different ranges, i.e., "*Small*", "*Mid*", and "*Large*" consisting of 1, 2 and from 3 up UAVs respectively (Table 2.12).

Table 2.12: Labels associated with the #UAVs.

#UAVs	Label
1	Small
2	Mid
from 3 up	Large

Learning Models Two different learning models were used in this domain, i.e., a BN classifier and a SVM. Details are provided below.

- BN classifier [98, 106]: is a learning probabilistic model, in which all variables involved in the study and their relationships can be represented in a model in order to derive inferences from the observation of the variables. The structure of the model created in this context is illustrated in Figure 2.36. It can be observed that the LOA estimated by the AAS is a direct child of the mission outcomes node via the workload node. This is due to the fact that the probability of changes in the human controller's MW was assumed to be conditioned from changes in the #UAVs in the three considered *"Alert"* state. Consequently, the probability to successfully complete missions is affected by the operator's MW.

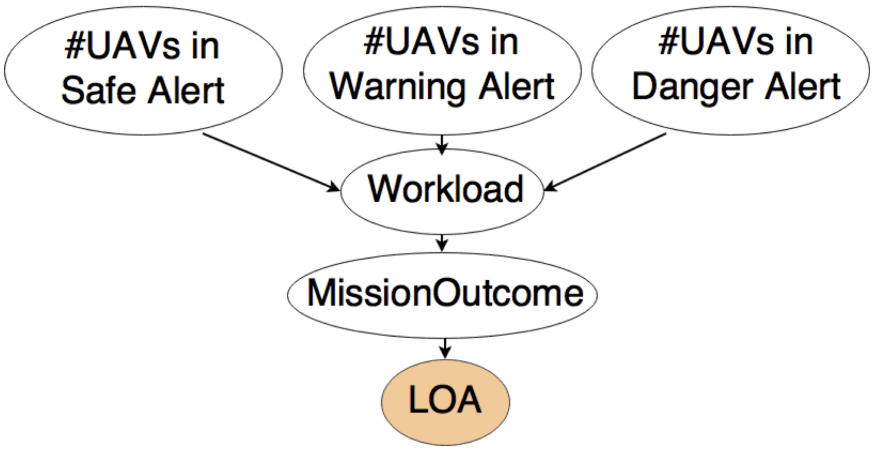


Figure 2.36: BN model inferring the LOA from UAV mission's outcomes, i.e., from operator's MW based on the number of UAVs grouped by *"Alert"* state.

- SVM [104, 107, 108]: is a learning classification algorithm able to deduce a model (representation of the data as points in space) from labeled training data. In this thesis, it is used for inferring the human controller's ability to successfully complete or not a mission from his or her MW measured by EEG signals. It is implemented with two different kernels: linear and Radial Basis Function (RBF). The former is generally used to find the best hyperplane separation in binary classification problems by tuning the *regularization parameter* C in order to regularize and control the bias-variance trade-off. The latter is generally used in problems that are not linearly separable and require to find also the best value of the γ *parameter* in order to define the variance of the RBF.

The above classification models were defined and exploited in this thesis with the aim of learning from data (i.e., operator's MW) collected through two user studies, how to infer the appropriate LOA in drone-traffic-control tasks. In particular, the first study combines subjective and performance-based MW measurements with the BN classifier. In the second study, physiological and performance-based MW measures were combined with the SVM learning model.

First User Study: Subjective and Performance-based MW Evaluation

Participants involved in the study (6 males and 2 females) aged between 24 and 27 years (M = 25.38 SD = 1.22), were selected among university students from Politecnico di Torino.

At the beginning of the experiment, participants were told that during the test, they would have to perform some supervision and to monitor tasks of a growing number of UAVs by acting like a real UAVs controller. To this end, a brief training was provided to participants in order to instruct them on the use of the UI and how to intervene

by means of the flight commands when critical conditions were warned by the UAVs through an alert.

The experiment consisted of six sessions (1 practice and 5 tests) of two trials, one in "*Warning*" mode and the other in "*Suggestion*" mode by using the related UI. A random order was used to select the above modalities in order to limit the effect of learning. Each trial lasted approximately 4 minutes.

During each trial, quantitative data about the outcome of each mission, the number of unmanaged UAVs, as well as the *"Alert"* of each UAV were recorded. At the end of each trial, participants were required to compile a NASA-TLX questionnaire [78] for each action performed on the UAVs.

After completing a session, participants were also asked to indicate which LOA of the system they preferred in carrying out the test. The execution of the whole experiment and the compilation of the questionnaires took about 2 hours per participant.

Results obtained in terms of percentage of participants able to successfully complete the missions as well as the average values of the operators' perceived MW are shown in Figure 2.37 and Figure 2.38 respectively.

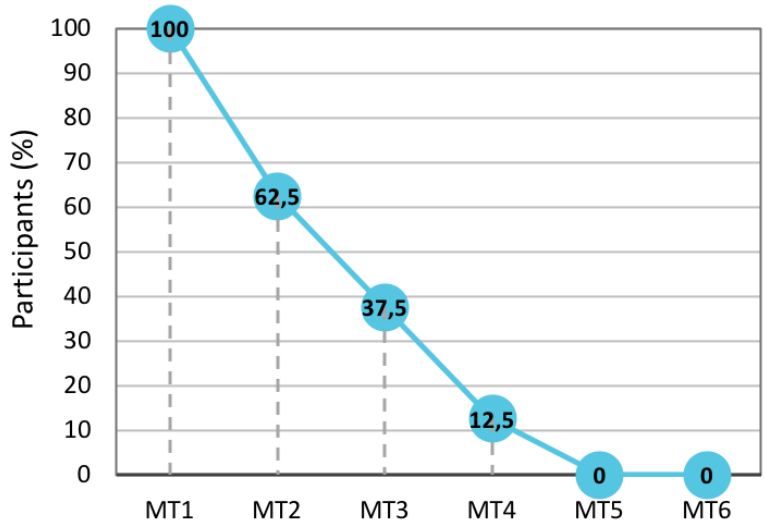


Figure 2.37: First user study: results in terms of percentage of participants able to successfully complete the missions.

It can be observed from Figure 2.37, that the percentage of participants able to complete the missions gradually decreases from *MT1* (1 UAV) to *MT5* (5 UAVs) by reaching the lowest value in *MT5* and steadying in *MT6* (6 UAVs). Concerning operators' self-assessed MW, the NASA-TLX average score gradually rises from *MT1* to *MT3* (2 UAVs), then slightly increases in *MT4* (3 UAVs) and rapidly surges from *MT4* to *MT6* (Fig. 2.38).

By moving from these findings and combining them as illustrated in Figure 2.39, it can be noticed, as expected, that *MT1* is the mission all participants were able to

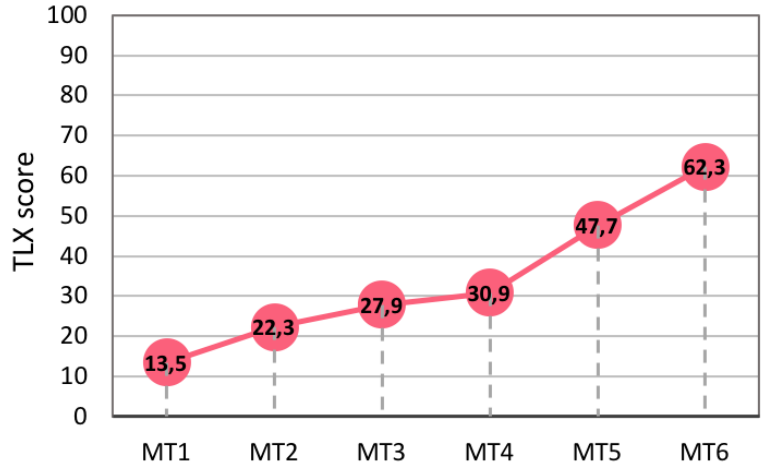


Figure 2.38: First user study: results in terms of NASA-TLX average score in the considered missions.

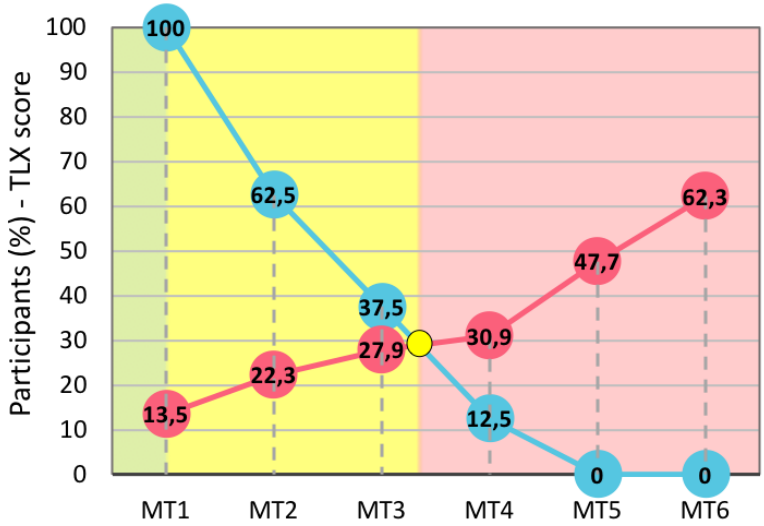


Figure 2.39: First user study: percentage of participants able to successfully complete the missions combined whit NASA-TLX average score in the considered missions. The intersection and turnaround point is depicted by a yellow circle.

complete by perceiving the lowest MW. On the contrary, *MT6* represents the mission that no participant was able to complete and the one showing the highest perceived MW. The yellow circle in the graph represents the turnaround point between two trend graphs, in which the higher TLX score the lower the percentage of participants who successfully complete the missions. By leveraging the above observations, operators' MW TLX score in managing 1, 2 and from 3 up UAVs may be labeled as "*Low*" (green

area in Fig. 2.39), "*Medium*" (yellow area in Fig. 2.39) and "*High*" (red area in Fig. 2.39) workload respectively (Table 2.13).

Table 2.13: First user study: labels associated with operators' MW TLX score.

MW TLX score	**Label**
TLX score ≤13.5	Low
13.6 ≤TLX score ≤30.0	Medium
TLX score > 30.0	High

Obtained results were then exploited to train the BN classifier to learn how to infer the appropriate LOA for the system. A cross-validation technique was then exploited to assess from the point of view of accuracy the BN classification performance and its ability to predict LOAs on unseen data. For this purpose, data collected were split into two different groups, namely, *training set* and *validation set*, as follows: 80% and 20% of the data, respectively. The former set was used for training the BN classifier, whereas the latter set was used for the classification accuracy validation.

The whole data set contains as many rows as the actions carried out by participants on UAVs by using the control buttons defined in Section 2.2.3. More in detail, each row consists of the #UAVs in the three *"Alert"* states, the operator's MW level, the outcome of the mission and the operator's preferred LOA in that situation. As a matter of example, a possible participant's test result, like the one shown in Table 2.14, could be considered.

Table 2.14: First user study: example of a test result with 3 UAVs.

UAV_1's Alert Status	**UAV_2's Alert Status**	**UAV_3's Alert Status**	**MW TLX score**	**Mission Outcome**	**LOA**
Safe	Safe	Warning	10.36	Success	Warning

During this test, the participant had to manage three UAVs, two of them in "*Safe*" alert and the other in the "*Warning*" one. The outcome of the test indicates that, in this case, the participant was able to successfully complete the mission by perceiving a MW equal to 10.36 TLX score and by preferring "*Warning*" as LOA for the AA system.

The corresponding row in the data set, built from the above test result, is shown in Table 2.15. It can be observed that the #UAVs in "*Safe*" alert was labeled as "*Mid*" (Table 2.12), whereas the #UAVs in "*Warning*" alert was labeled as "*Small*" (Table 2.12). None UAV exhibited a "*Danger*" alert, thus the #UAVs in this alert was set equal to *Null*. Furthermore, since the obtained MW TLX score < 13.5, the participant's MW was labeled as "*Low*" (Table 2.13).

Table 2.15: First user study: example of a row in the data set.

#UAVs in "Safe" Alert	#UAVs in "Warning" Alert	#UAVs in "Danger" Alert	MW level	Mission Outcome	LOA
Mid	Small	Null	Low	Success	Warning

The BN training phase was performed by means of the Netica Software[9], then the validation methodology was carried out by obtaining a classification LOA accuracy equal to 83.44%. Table 2.16 reports the confusion matrix for each LOA considered in this study.

Table 2.16: First user study: confusion matrix.

	True Warning	True Suggestion	True Autonomous	Class Precision
Predicted Warning	15	1	0	93.75%
Predicted Suggestion	1	30	7	78.95%
Predicted Autonomous	0	4	19	82.61%
Class Recall	93.75%	85.71%	73.08%	

Second User Study: Physiological and Performance-based MW Evaluation

In this study, EEG signals were exploited to build a prediction model of the UAV operators' MW in order to train the AA system to autonomously predict operators' performance (missions outcome) in monitoring operations. To this aim, a SVM classification model was exploited to learn the ability of operators to carry out assigned traffic-control tasks in different flying scenarios. Details of the data analysis and classification procedure entailing *data pre-processing*, *feature extraction* and *classification* steps are described below.

Data pre-processing The EEG consists of electric signals produced by the activation of thousands of neurons in the brain. These signals are collected by electrodes placed over a person's scalp. However, the presence of noise or artifacts, defined as signals with no cerebral origin, may affect the EEG data with some spurious signals. In particular, the artifacts can be grouped in two categories: related to physiological sources or consisting of mechanical artifacts. Examples of the former category are eye blinking, heart beating, and ocular movement, whereas examples of the latter are the movement

[9]https://www.norsys.com

of cables or electrodes during data collection [109]. Thus, a pre-processing stage is required to remove noise and all undesired signals. It consists of three different phases, namely, *filtering*, *offset removal* and *artifact removal*, whose details are provided below.

- Filtering: exploits a Finite Impulse Response (FIR) passband filter for removing signals with high frequencies and increasing signal-to-noise ratio, as the EEG signals frequencies lay within 0.5 and 45 Hz.
- Offset Removal: removes potential offset residues after the filtering phase.
- Artifact Removal: implements the Artifact Subspace Reconstruction (ASR) algorithm for removing artifacts [110].

Feature Extraction Given the preprocessed data, relevant features need to be extracted to train the considered classification model. For this purpose, the signals containing relevant events to be analyzed were split in different time ranges as follows: 15s after the start of the EEG recording and 15s before the first failure, divided in 5s windows. Data collected during the idle UAV's takeoff phase were neglected in order to guarantee that the related MW measurements were not used as baseline reference in the monitoring experiment. Data in the interval just before and after the first failure were not considered as they may be affected by biases due to the operator's frustration for failing the assigned task. For each window, differences features were calculated channel by channel, namely, Power Spectral Density, Mean, Variance, Skewness, Kurtosis, Curve length, Average non-linear energy and Number of peaks [104]. These features were then concatenated for making each window correspond to a row of features appearing in order of channel. Each row was then associated with a label that states whether the operator successfully completes the task or not for that particular mission.

Classification The aim of this step is to train the classification model considered in this study, i.e., the SVM, with the UAVs controllers' MW for predicting their performance in monitoring tasks. By digging more in detail, every single subject's MW, as well as the overall MW data gathered from all operators, were used to train the SVM, in order to understand whether a generalized model may be also employed. A procedure consisting of *feature scaling*, *hyperparameter optimization*, *results validation* and *learning model design* was proposed in order to assess the model considered from the point of view of accuracy. Procedure details are described below.

- Feature Scaling: the high variability of the features extracted from each subject as well as their different ranges represent an important issue in signal processing field, and in particular with the EEG data. Thus, an appropriate scaling method is required for normalizing all data into the same range. In particular, a *z-score* scaler was used as normalization method to subtract average values from all measured signals and then divide the difference by the population's standard deviation [111].

- Hyperparameter Optimization and Validation Methodology: since the purpose of the classification methodology is to achieve good accuracy on unseen data, an appropriate validation procedure is needed to measure the generalization error of the considered model. To this aim, a k-fold cross validation methodology was exploited in this thesis for both finding the best model with the optimal parameters and testing its performance on new unseen data. Specifically, in k-fold cross validation, the original sample is randomly partitioned into k equal sized subsamples. Then, a single subsample of the k ones is retained for testing the model, whereas the remaining $k-1$ subsamples are used in each iteration for training it. According to this methodology, three different data sets were created in this thesis, namely, *training set*, *validation set*, and *test set* as follows: 20% as *test set*, and the other 80% as *training* and *validation sets*. A 10-fold cross validation was then performed on *training* and *validation sets* as follows: samples were partitioned in ten folds, nine of them were used in each iteration for training the model, and the other one was used for assessing the results. These steps were then iterated until all folds were used one time as *validation set*. The training accuracy was then computed as the mean of all the obtained results in the different iterations. The parameters leading to the best model performance ("*Hyperparameters*") were then selected [112]. Lastly, the model was assessed using the *test set*.

- Learning Model Design: as detailed in Section 2.2.4, a SVM with two different kernels (i.e., linear and RBF) and different regularization parameters (C and γ respectively) was used in this study. The C parameter was tuned by using a search grid with powers of ten from 10^{-2} to 10^{2}, during the cross-validation phase. For the γ parameter, the powers of ten from 10^{-4} to 10 were used considering that larger values lead to better adjustment of the model to the training set, even though bringing possible variance problems or over-fitting. Smaller values may bring bias or under-fitting problems.

Experimental Setup and Results Participants involved in the study (8 males and 2 females) aged between 19 and 24 years (M = 21.10 SD = 2.08), were selected among university students of Politecnico di Torino.

As in the first study, at the beginning of the experiment, participants were informed that during the test, they would have to deal with a growing number of UAVs for carrying out some supervision and monitoring tasks. A brief training was then performed to instruct participants on the use of the UI and its functionalities, i.e., flight commands for controlling UAVs (Fig. 2.31). Afterwards, participants were invited to perform the six missions considered in this application domain (from *MT1* to *MT6*) in sequence through the *Warning* UI.

During each test (i.e., all tasks performed), physiological measurements gathered by the EEG signals through the EMOTIV Epoc+®headset were recorded. Each task took from 2 to 7 minutes depending on the operator's piloting choices.

Results obtained in terms of classification accuracy are reported in Table 2.17 specifying the hyperparameters used to train each single model. The first ten rows of the table show the obtained results with the model trained using single subject data. The last row represents the obtained results using all the collected data. The accuracy scores obtained with the 10-fold cross validation methodology are reported in Table 2.17 as "Accuracy (Validation set)", whereas those obtained with new unseen data are reported as "Accuracy (Test set)".

It is worth noting that, as illustrated in Table 2.17 with the $*$ symbol, some single subject data were discarded and not used to train the related individual model, as they presented corrupt information. More specifically, in these cases, participants were able to successfully complete one mission out of six, making it very difficult to train the model due to the skewness of the class. However, these data were used in the overall model.

Table 2.17: Second user study: results concerning the accuracy of the classification algorithm for the individual and overall models.

	SVM - linear kernel			**SVM - RBF kernel**			
Participant ID	C	*Accuracy (validation set)*	*Accuracy (test set)*	C	γ	*Accuracy (validation set)*	*Accuracy (test set)*
1	0.01	0.949	0.933	100	0.0001	0.949	0.933
2	100	0.923	0.973	100	0.0001	0.934	0.973
3	0.01	0.965	1	100	0.0001	0.965	1
4	0.01	0.851	0.965	10	0.0001	0.851	0.93
5	*	*	*	*	*	*	*
6	0.1	0.885	0.895	10	0.001	0.899	0.864
7	*	*	*	*	*	*	*
8	0.01	0.944	0.969	100	0.0001	0.936	0.969
9	0.01	0.986	0.927	10	0.001	0.897	0.864
10	0.01	0.995	1	10	0.001	0.995	1
Overall	0.1	0.852	0.839	10	0.001	0.872	0.856

From Table 2.17, it can be observed that the differences between the scores in the "Accuracy (Validation set)" and "Accuracy (Test set)" columns for the same row are not very pronounced. Therefore, it can be concluded that the considered model is not affected by problems of variance thus, performs well if tested with other participants under the same conditions.

Concerning the accuracy scores obtained in the test sets, it appears that the linear kernel always performs better or equal than the RBF kernel for individual models. On

the contrary, the RBF kernel performs better than linear kernel for the overall model. By digging more in detail, the SVM with the linear kernel is able to predict the operator's performance outcomes thus the level of his or her MW with an average accuracy equal to 83.9% and 95.8% when the model is trained on all collected and on a single user data, respectively. When the SVM - RBF kernel is employed, an accuracy equal to 85.6% and 94.1% is reached using the overall and single user data, respectively. This result may be reasonably due to the fact that individual models trained using single subject data represent simpler classification problems than those with all collected data.

Chapter 3

Interaction in Colocated Spatial Proximity Pattern

Part of the work described in this chapter has been previously published in [113, 114, 115, 116].

As depicted in Chapter 1, HRI represents a very broad research area containing various interaction techniques which differ depending on whether they are used for industrial or service robots and whether the human and the robot are in close proximity to each other or not (thus defining the colocated and remote spatial proximity patterns). As already said in Section 1.4, the focus of this thesis is on HRI interfaces used in service robotics applications lying in the space between remote and colocated scenarios. Therefore, Chapter 2 has explored HRI in the remote spatial proximity pattern, whereas this chapter will investigate HRI in the colocated one.

In the colocated proximity pattern, robots interact closely with human users in their everyday environment. As discussed in Section 1.3, HRI is affected by problems arising from direct interaction and users' expectations in terms of usability and user acceptability. The robot is expected to have self-awareness (about what it can do) and to be aware of humans' presence (about what the human user can do), to exhibit semi- or autonomous behaviors (e.g., avoiding hazards or maintaining users' safety) and to be able both to communicate effectively with humans and to adapt its behavior/actions according to the inputs coming from users.

The purpose of this chapter is, on the one hand, to explore and investigate the related work and background defining the state of the art of the two application domains considered in this thesis as representative examples of colocated HRI, i.e., robotic gaming and (socially) assistive robotics (Section 1.4). On the other hand, the goal is to leverage the acquired knowledge to develop and implement frameworks dealing with concrete use cases in order to identify and develop suitable HRI paradigms for addressing problems arising from the considered colocated pattern.

3.1 Robotic Gaming

The robotic gaming domain was chosen as a representative example of applications in colocated spatial proximity pattern with the aim, on the one hand, of investigating users acceptability related to the introduction of emotional features and autonomous behaviors in a service toy robot. On the other hand, with the aim of studying whether a MR Phygital Play platform can be exploited to set up robotics gaming scenarios capable to engage users and limit sedentary and solitary behaviors by combining real and digital contents.

3.1.1 Background

In every age and culture, gaming was always considered the basic form of daily entertainment. Nevertheless, due to technological advances, the types of games have undergone crucially changes over the years. Hence, today, video gaming is the most popular type of recreational entertainment [117].

A video game may be defined as an electronic game executed by a digital system (e.g., hand-held or an arcade machine) providing feedback on a screen [118]. However, the use of a screen implies carrying out sedentary and static activities performed while sitting in front of it, without requiring any physical exercise. Therefore, the excessive use of video games was found to be correlated with some disorders resulting from a sedentary lifestyle, such as cardiovascular diseases, obesity, metabolic syndrome, psychological problems and stress [119, 120]. Even playing a multi-player video game does not promote in-presence collaboration or social activities among players since it occurs through virtual meeting places [121].

Hence, the entertainment industry is currently focusing on the design and development of new game concepts involving both real and digital elements to foster active participation and encourage socialization among players. Thus, on the one hand, traditional games are being enhanced with digital and social content; on the other hand, video games are trying to reintroduce the physical dimension [122]. Within this context, one of the emerging trends consists of including service robots into the gaming domain by developing so called "*robotic gaming*".

In [123], robotic gaming was defined as a game involving a number of autonomous agents (at least one robot and one person) interacting with each other as *peers* within an unpredictable and a variable playing environment with some rules to satisfy for leading the user to have fun. The authors of [123] also defined a novel robotic gaming paradigm centered on HRI called "*Physically Interactive Robotic Games*" (PIRGs) and developed a set of guidelines to design it. Figure 3.1 illustrates the most relevant ones.

As discussed above, when implementing PIRGs involving autonomous agents as players, much attention has to be devoted to robot's autonomy, intended as "*an internal and integral component of reasoning on interactions*". That is, "*if the determination of the agent's behavior is local and without input from other agents, the agent is autonomous*"

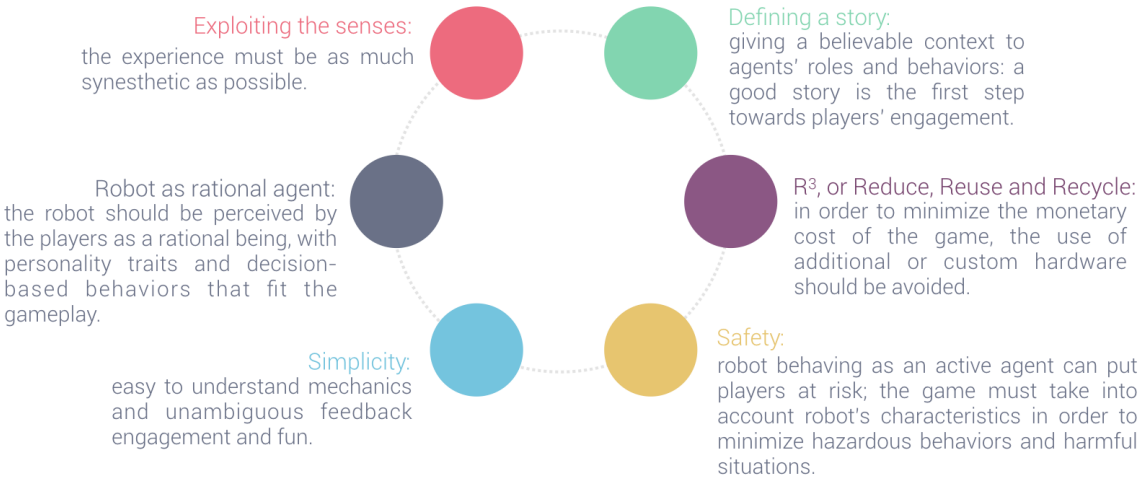


Figure 3.1: PIRG guidelines defined in [123].

[124]. The concept of autonomous behavior is strictly related to the one regarding Artificial Intelligence (AI), broadly speaking. Today, two prominent frameworks are used to guide AI implementation, i.e., the classical deliberative model, also known as *Sense-Plan-Act* (SPA), and the modern reactive model, which is referred to as *subsumption architecture* by Brooks in [125]. The main difference between these models consists of that the SPA follows a closed loop iterative approach, whereas the *subsumption architecture* privileges resilience and robustness of a subset of features by organizing autonomy tasks hierarchically. Therefore, a reference frame for guiding the design of a game involving autonomy needs to be identified. In [126], several key aspects concerning robot's autonomous behaviors for a user-centered design perspective are considered, as illustrated in Figure 3.2.

Another fundamental aspect relevant to the implementation of PIRGs is represented by the ability of robots to interact with humans in a way resembling human-like interaction. For this reason, the encoding of emotions represents an essential factor to be taken into account for improving HRI and fostering users' acceptability of robots. A good way to increase engagement and promote long-term social presence is to consider the so-called *affective loop* [127], i.e., the inseparable relationship between body and mind in embodied affective systems that gives the illusion of life. To deal with affective interactions it is required to identify which embodiments and modalities are to be used for modeling and expressing emotions. Concerning the embodiment, the lack of facial expressions and body movements in a robotic system make more difficult the transfer of emotional information since human beings are instinctively more sensitive to emotions conveyed by an anthropomorphic system [128, 129]. The authors of [130] have tackled the above-mentioned problem by observing that whether emotional behaviors respect some patterns of occurrence and distinct paradigms, a human being is able to identify such actions and distinguish them from the overall behavior of the robot.

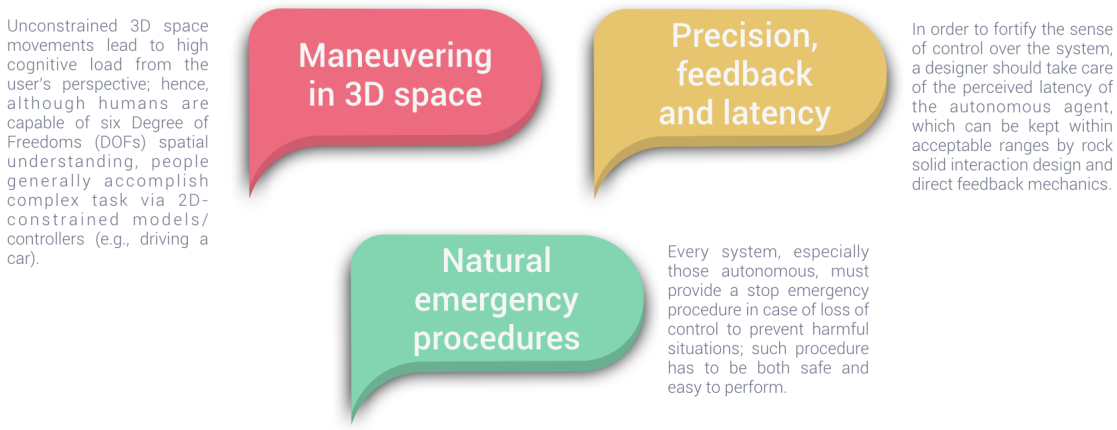


Figure 3.2: Key aspects concerning robot's autonomous behaviors according to [126].

Furthermore, in order to validate the illusion of life, sound and voice should be exploited not to evoke a feeling as human-like as possible, but, rather, to make it consistent with the "appearance" of the robot.

In parallel to the principles defined in PIRGs, a more technical guideline was proposed in [131] with the aim of providing an organized structure to developed games. According to the authors of [131], the design of games consists of analyzing and defining the Mechanical, Dynamic and Aesthetic (MDA) factors, as illustrated in Figure 3.3.

Figure 3.3: MDA factors as defined in [131].

In the literature, the above design principles and guidelines have been exploited in several PIRGs, such as [122, 132, 133, 134]. Although different types of service robots were used, one of the most common and prevalent approaches is represented by the use

of drones, as highlighted in [135].

By leveraging the above consideration and focusing on drone-based games (although discussion holds also to other robots), most of them assume that the fun for the player arises from mastering and controlling these vehicles. Despite the popularity and success, humans operate these robots through remote control and, hence, HRI is very poor. For this reason, a number of drone-based games centered not only on the player but also on the HRI have been developed.

For instance, in [123] the *Jedi Trainer* game was designed to evaluate the PIRG guidelines. The game is a First Person Shooter (FPS) consisting of a player deflecting the incoming fire from an autonomous drone by defending himself or herself through the use of a "lightsaber". Similarly, in the competitive game called *LaserGame*, a drone is able to shoot the human player (later abbreviated HP) through a virtual laser by looking for him or her autonomously. The human user can defend himself or herself by shooting back to the drone. Moreover, many drones can be grouped together for creating a swarm by ensuring an acceptable LOA for each drone but at the same time ignoring the R^3 of PIRGs principle since additional hardware for localization is required. A further example is represented by *Drone Laser Game* (*DLG*)[1], where a drone battle is engaged between a player-controlled drone and an autonomous one, with each drone aiming to shoot down the other one. Indeed, in this game the robot is perceived as a *rational agent*. However, the game fails in *exploiting the senses*, since the player has a static behavior. Moreover, Mechanics targeted to defense and a proper *story* are missing. In *LaserDrone*[2], the already existing and appreciated paintball game, with established design Mechanics, is considered and transformed in a drone-based game. Drones are used as additional fighting units, which are remote-controlled by players. Despite the high level of engagement, autonomy aspects are totally absent.

In parallel to the studies described above, other works have explored the previously mentioned combination of real and virtual elements with the aim of improving player's engagement when interacting with the robot within its environment.

In this context, AR and MR technologies have been exploited to generate new gaming environments by overlapping physical and digital objects in the same playing area that interact with each other in real-time [136]. These scenarios have been generally referred to as "*Phygital Play*" [122]. A list of requirements that are considered to be crucial for creating phygital games is illustrated in Figure 3.4.

A number of games based on the above gaming environments have already been presented. For instance, in [137], small mobile Do-It-Yourself (DIY) robots endowed with low-cost light sensors for tracking purposes were used together with AR environments exhibiting obstacles, virtual paths, and battles between robots through the use of a projector. The authors of [138] have developed tabletop projection robotic games by

[1] http://airlab.ws.dei.polimi.it/index.php/Drone_Laser_Game

[2] https://it.ulule.com/laserdrone/

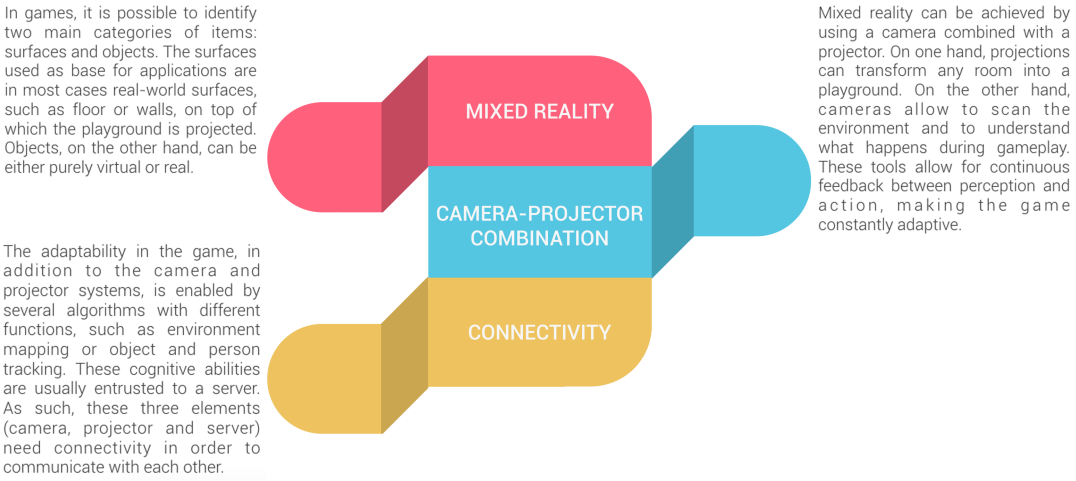


Figure 3.4: Key aspects in *Phygital Play* scenarios according to [122].

exploiting an AR platform and Lego Mindstorms robots. In [139], a low-cost, extensible platform based on the commercial iRobot architecture and supporting projection, intelligent objects, spatial sound, and gesture recognition is proposed to demonstrate a wide range of robotic gaming possibilities, with particular attention to physical interaction with the robot in immersive environments. In [140], a mobile MR system consisting of a robot and a portable projector was presented to allow children producing their stories in a gaming area combining virtual and digital elements. A further example is provided in [133], where a configurable cloud-based robotic gaming platform exploiting MR technologies has been illustrated. The platform allows the creation of different interactive gaming experiences through the use of a projector, one or more autonomous toy robots and different localization technologies.

In summary, the entertainment industry is focusing, on the one hand, to design and develop new games combining real and digital elements together, and on the other hand, to reintroduce the physical dimension in video games. The reason behind this trend could be attributed to the intention of fostering active participation and encouraging socialization among players as well as to overcome arising disorders resulting from a sedentary lifestyle.

By leveraging the above considerations, in this application domain, two games were developed: a FPS game, called *Protoman Revenge* (later abbreviated to PR), involving a drone as physical element and a floor-projected MR game, called *RoboQuest* combining real elements (a toy robot and a set of tangible interfaces) and digital elements (displayed on the floor). The former game was devised for evaluating the impact on user experience of an autonomous robot exhibiting emotional features, whereas the latter one for investigating how to favor an engaging interaction between players and real/virtual game elements.

The following two sections will present the design phase of the aforementioned

games, their game logic, and some implementation details.

3.1.2 Protoman Revenge Game

The story of this game (satisfying the first principle of PIRG guidelines illustrated in Figure 3.1) took inspiration by the Megaman VII video game, where a human-controlled and a computer-controlled fighter are engaged in a duel and both use a laser beam shooter (placed at the extremity of the arm) as weapon. The human fighter is also endowed with a defense ability through a handheld shield. Recreating such fight within a PIRG scenario by introducing a drone means that the HP can shoot laser beams and protect himself/herself with a shield, whereas the antagonist drone can move in the 3D space and shoot (Fig. 3.5).

Figure 3.5: Human player and robot player in *Protoman Revenge*.

Game Design

By moving from drone-based games described in Section 3.1.1, it can be observed that none of the discussed PIRGs harmonized all the guidelines listed in the design requirements. In this thesis, with the aim of bringing an original contribution, the *Protoman Revenge* game was designed for enhancing the user experience, being entangled to HRI and including all the design principles suggested by state-of-the-art in the field. The conceptual design process behind *Protoman Revenge* is depicted in Figure 3.6.

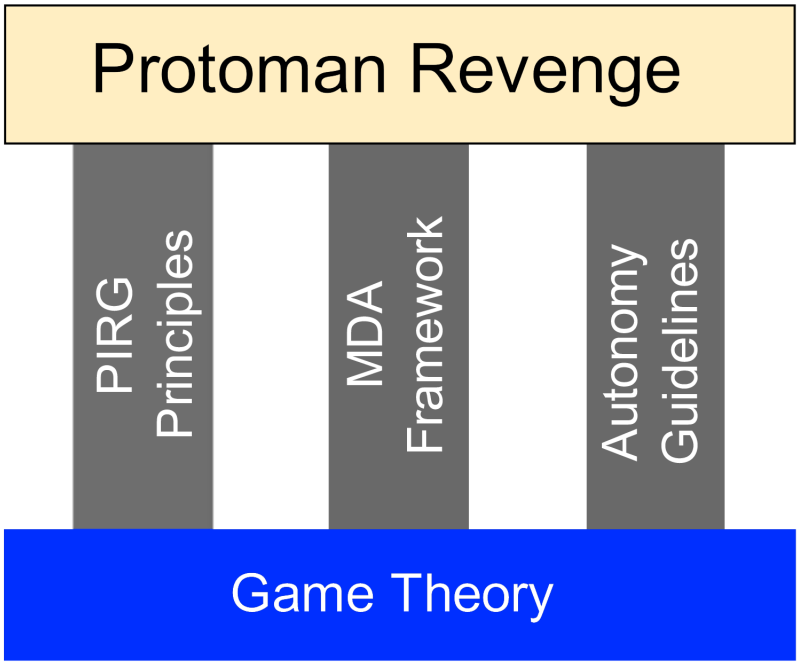


Figure 3.6: Game design frameworks behind *Protoman Revenge*.

As illustrated in Figure 3.6, game theory principles [141] represented the basis of the design process. According to these principles, a game is defined as a mathematical object describing the interaction among players, whereas a player is a rational agent trying to optimize a given utility function or payoff. A set of strategies and actions are also defined for being chosen by each player. According to game theory, the *Protoman Revenge* game can be described in the following terms:

- *non-cooperative*, since the goal of each player is to achieve individually the maximum possible payoff;
- *zero sum*, i.e., the total benefit of all the players in the game is equal to zero (meaning that a player benefits for the success in his or her strategy only at an equal expense of the other player);
- *simultaneous*, that is, players move simultaneously with no prior knowledge of the opponent's move (in contrast with sequential games, where players move alternate, like in chess);
- *perfect information*, that means all players are perfectly informed of all the events that occurred previously;
- *continuous*, namely, players can choose a strategy from a continuous strategy set.

Based on the MDA framework [131], the game can be described with the elements listed below.

- *Mechanics*: a "Shoot-Pursuit-Evasion" model is adopted, consisting of attacking with the laser beam, protecting with the shield, moving to evade or avoid attacks and getting ready for another shooting attempt. Micro Mechanics are developed for the laser shooter to balance the three steps of the core Mechanics by envisaging cool-down and reloading timers. The PR game based on Mechanics states that the player who obtains three shoots correctly first wins the game. These Mechanics represent easily understandable concepts, thus satisfying to PIRG principle of *simplicity* (and fun).
- *Dynamics*: defines the movements and actions actuated by the drone; they consist of six main routines, which are defined according to the Mechanics above:
  - tracking, follow the HP by satisfying safety requirements;
  - shooting, launch an invisible laser towards the HP;
  - escaping, perform actions and movements to avoid as much as possible attacks from the HP;
  - reacting, execute maneuvers to express unhappiness, frustration or anger for a given game event;
  - detecting, recognize the HP and his/her equipment from the surrounding environment;
  - searching, perform maneuvers devised to seek the HP autonomously even in the case that he/she exits robot's FOV.

  For each routine, the drone is expected to show different behaviors by choosing among a set of pre-programmed trajectories. As a matter of example, when it attacks the HP by shooting the laser beam but the shot hits the shield, the drone will select a maneuver between the set of possible trajectories. Specifically, if the shot hits the HP for a given number of attempts, the drone will execute a victory maneuver, whereas if it is hit multiple times, it will execute a defeat maneuver. These human-like behaviors of the drone have the aim of making the HP perceive the drone as a rational agent (thus satisfying the PIRG guidelines). All the trajectories executed in the game are accompanied by sounds for helping the player to identify the movement associated with the desired emotion and enhance sensorial feedback.
- *Aesthetics*: the aesthetics features selected from the set described in [131] are listed below:

- sensation (game as sense-pleasure), the game stimulates the player's senses in a considerable way: in this case, the HP experiences something unfamiliar, i.e., a game leveraging an autonomous drone;
- fantasy (game as make-believe), the HP experiences something that can never become in real life: in this case, an imaginary world in which the HP imagines the laser beam, the shots, the bombs or missiles (described later) even though they are not visible;
- challenge (game as obstacle course), deriving fun from overcoming arbitrary obstacles: in this case, game's playability is boosted, since the HP tries to correctly shoot the drone down by protecting himself or herself with the shield from drone attacks.

By leveraging the aforementioned game design, Figure 3.7 depicts how the PR game developed in this application domain is positioned with respect to the drone-based games discussed in Section 3.1.1, in terms of autonomy and human-drone interaction.

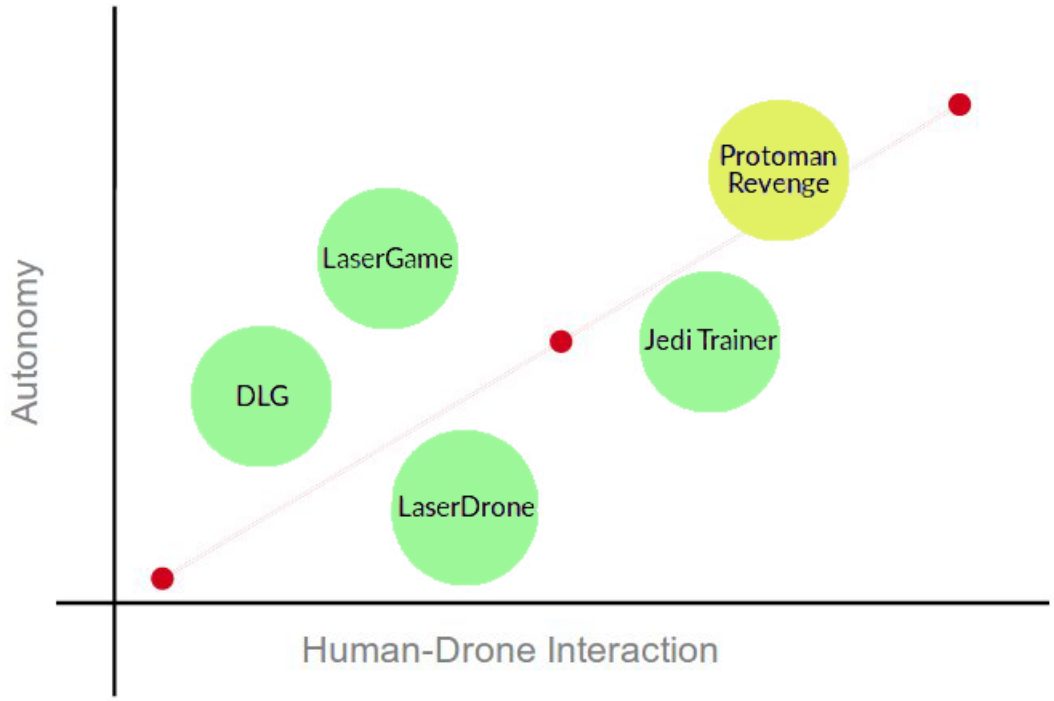


Figure 3.7: Visualization of the line of tendency in robotic games using drones.

Technology

As said in the section above, when dealing with PIRGs, the hardware and software choices shall be consistent with the design guidelines, and, particularly, with the R^3

principle. The robot selected for this game was chosen according to some criteria, i.e., commercial availability, existing Software Development Kits (SDKs) and community support, ability to fly (indoor/outdoor), cost, featuring sensors and game design functionalities those sensors enable. In particular, the Parrot® AR.Drone 2.0 (Fig. 3.8) was selected, in order to satisfy the above criteria but also considering the trend to use this drone[3] in several games, created both for commercial and research purposes, like, for instance, *Astro Drone*[4], *Drone Escape*[5] and *TargetHunter*[6]. It is worth noting that all of them exploit the remote control paradigm thus lacking autonomous behaviors.

Figure 3.8: Parrot® AR.Drone 2.0.

The AR.Drone 2.0 is a Commercial Off-The-Shelf (COTS) quad-copter endowed with a six DOF IMU and a three-axis magnetometer, an ultrasound telemeter and an altimetric pressure sensor. Two different cameras are mounted on it, namely, a primary 90° FOV, 720p, 30fps front-facing camera and a secondary 64° FOV, QVGA, 30fps down-facing camera. A cover for indoor flight is also provided in order to make the drone suitable for home-based settings. A smartphone app connected to a Wi-Fi hotspot (created by the drone itself) is generally used to control the drone. Nonetheless, developers are allowed to create third-party applications (not necessarily mobile) through the use of the official SDK. Moreover, due to the popularity of the drone, many other toolkits and APIs are available.

The PR game was designed to be deployed within the ROS-based cloud robotics

[3] http://ardrone.parrot.com/other-apps/games

[4] http://www.esa.int/gsp/ACT/ai/projects/astrodrone.html

[5] https://www.youtube.com/watch?v=CICFcZdaTNw

[6] https://www.youtube.com/watch?v=QdMfaQt0fTU

platform developed by TIM. Hence, the *ardrone_autonomy*[7] library was used, a wrapper over the official SDK built on top of ROS. A Bluetooth headset was also used for providing audio feedback to the player and improving the intelligibility of sound effects.

Implementation

This section reports on activities performed to implement the designed PR game. In particular, three different implementation steps were executed, namely, *object detection*, *localization & motion control*, and *emotional encoding*. In Figure 3.9, a simplified decomposition of the drone's AI game logic, fitting the *subsumption architecture* described in [125, 142] and implementing the six routines discussed above, is illustrated.

The *object detection* is the step devoted to identify and distinguish the HP and props from the surrounding gameplay area. Two different approaches were tackled for creating and detecting the laser shooter. In particular, an approach exploiting Infrared Radiation (IR) laser transmitter and receiver was first considered, but soon discarded because of the need of additional hardware (thus violating the R^3 principle of PIRG). The second and selected approach relied on the idea of exploiting drone's front-facing camera and Computer-Vision (CV) algorithms to detect and track the HP and his or her laser shooter and shield. The drone communicates via Wi-Fi with the machine (PC) hosting the detection process. As illustrated in Figure 3.10, several props were fabricated using colored cardboard material. The cylindrical prop represents the laser beam shooter and, as for the shield, is expected to be handed by the HP. Very bright and distinguishable colors were used to make handcrafted elements in order to allow object detection by using basic color segmentation algorithms (sophisticated CV methods were out of the scope of this thesis). Similarly, as illustrated in Figure 3.5 the HP was provided with a fluo orange t-shirt.

The images captured by the drone's camera in RGB are converted to HSV and then segmented using fixed thresholding values. Afterwards, a noise reduction step is performed by using erosion and dilation. In this case, the robustness of segmentation may be influenced by the environmental light causing ideal HSV threshold values to fluctuate. For this reason and with the aim of mitigating this behavior, a gameplay area where light is as uniform and time invariant as possible should be chosen. Some additional processing steps are executed. In details, since handcrafted elements can be considered bigger compared to other possible disturbances, a filtering step is performed on the area of pixel blobs found in the camera image, which are further screened using a circularity feature threshold calculated on the contours. Lastly, in order to improve reliability of the detection process at runtime, a recovery strategy is implemented in case of missed detection, by combining adaptive thresholding with search maneuvers (consisting of

[7] https://github.com/AutonomyLab/ardrone_autonomy

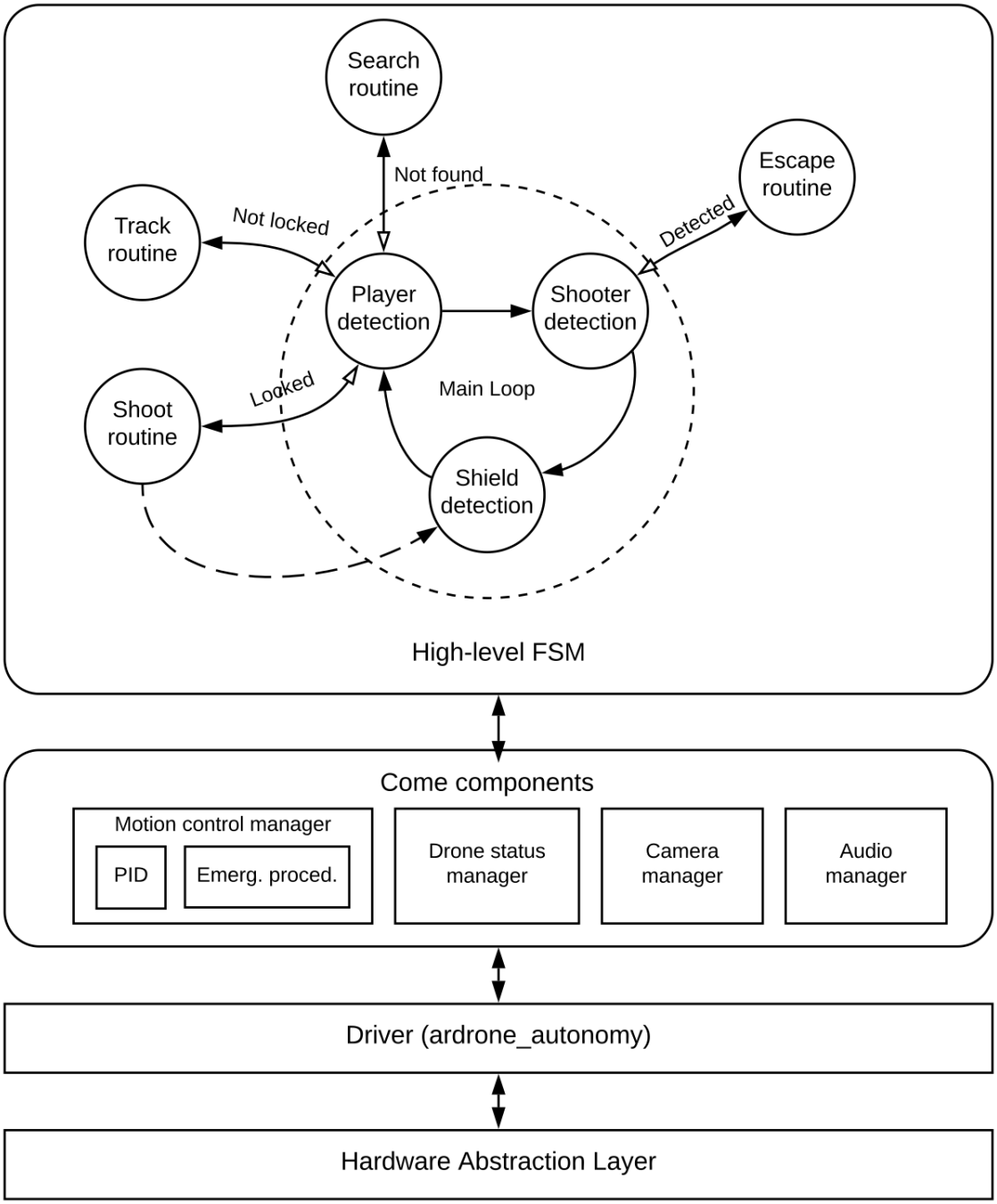


Figure 3.9: Drone's game logic.

ascending and descending clock-wise spiral movements).

Figure 3.10: Handcrafted player's equipment.

Concerning the *localization & motion control* step, some common approaches relying on global- or map-based localization algorithms (like SLAM navigation) were discarded. As mentioned above, the use of additional hardware (required by these algorithms) would have violated the R^3 principle of PIRG by also requiring highly demanding visual processing procedures. In the proposed game, the AI was implemented to rely on relative localization algorithm in order to let the drone move in the 3D space by preserving player's safety. As illustrated in Figure 3.11, a simple approach based on the portion of the player that is visible in the drone's camera FOV was chosen. Specifically, the position of the HP's orange t-shirt in the camera's FOV is used to feed a Proportional–Integral–Derivative (PID)-based motion control mechanism. The output of the PID is connected to the drone's driver software, which is programmed to manage only velocity commands. If the player exits the drone's camera FOV, the AI actuates the search maneuvers described in game design section to recover the tracking.

Lastly, the *emotional encoding* is the step devoted to encoding in a believable form for the drone the emotions defined for inducing the HP to perceive the opponent robot as a living and *rational agent*. A combination of voice and sound effects over special maneuvers was used to this purpose. In particular, maneuvers were defined to be as different as possible from each other in order to be distinguished from normal flight behavior. Thus, they were devised as cyclic movements and implemented through the parametrization of known curves for bringing the drone back to the position where the maneuver was initiated. Some video files showing the maneuvers are available for download[8]. Details about maneuvers and related emotions are described below.

- *Happiness*, a semi-ascending-spiral movement trajectory is followed, with the drone executing cyclic rotations clockwise and counter-clockwise while going

[8]https://goo.gl/fytDEq

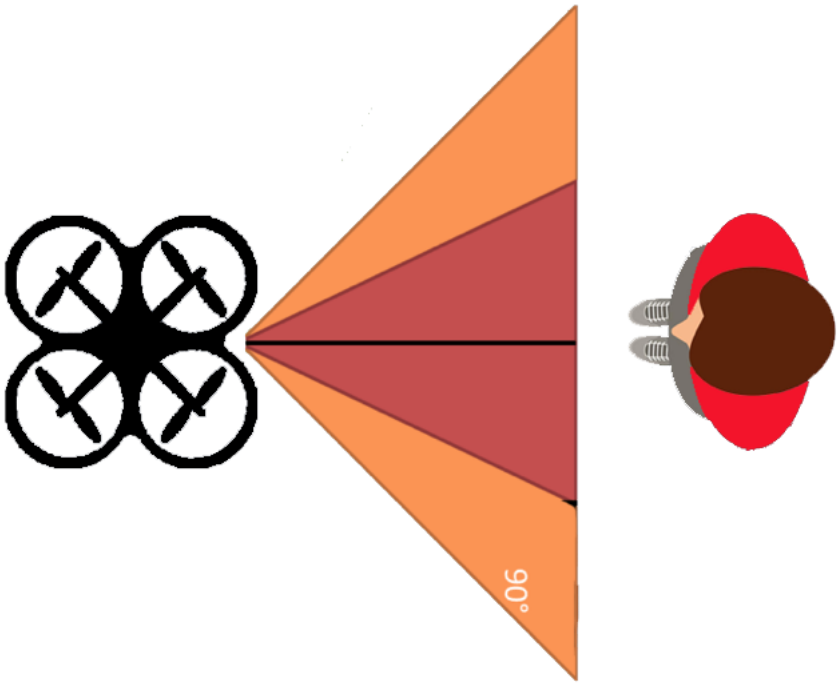


Figure 3.11: Drone's camera FOV for localization purpose.

up and down; this maneuver is associated with a satisfaction sound and is meant to be executed when the drone successfully shots the HP.

- *Frustration*, a Lemniscate of Bernoully movement trajectory is followed, whose shape is similar to the ∞ symbol. This maneuver was selected since it resembles an aircraft trying to regain stability after difficulties; it is associated with a frustration sound.
- *Anger*, the drone executes small amplitude clockwise and counter-clockwise rotations with a lot of repetitions accompanied by a fiery sound.

Game Experience

A snapshot is reported in Figure 3.12 and a video with gameplay is made available for download[9].

The game begins with the drone on the ground and the player ready in front of it. As soon as the game starts, the drone takes off. At this stage, the drone needs a "recognition" time for detecting the player and adapting its control parameters to the (body shape of the) specific individual who wants to play the game.

[9] https://goo.gl/ty2eWG

Figure 3.12: Gameplay: player is attacking the drone with the laser shooter.

After having completed the recognition phase, the drone will inform the HP that it is ready to fight by means of a voice feedback. It is worth noting that, differently than many other drone-based games, this game is not endowed with a display interface (for example, on a smartphone). Rather, body input and audio output are employed, which are expected to lead to more emphatic relations with autonomous systems [143] (which nonetheless could be pushed further by means of visual feedback, as proved, e.g., in [144], [145] and [146]).

The drone flies around the player, keeping itself at a fixed distance of approximately one meter. Shield detection is performed by checking the amount of area in the camera FOV that is covered by the green board. Evaluation is executed a certain number of frames after the sound/the maneuver associated with the laser shot has been played/executed, in order to give the HP enough time for moving the shield and protecting himself/herself. Score points are assigned accordingly. Score is constantly checked by the system and announced at each change using audio feedback.

Drone is equipped with two shooting modalities, i.e., *missile* and *bomb*. The former arises as a penalty for the lack of movements or continual defensive planning by the HP. As soon as the drone is able to keep the HP at the center of the camera's FOV by locking on him or her for a given amount of time, the missile is sent off. A sound warning is used to notify the HP that the missile was dropped and that he or she has to use the shield to protect from the missile. The latter is released periodically after a given time interval as long as the HP is detected in the camera FOV. A sound warning about the bomb is issued by the drone for informing the HP that will have to protect himself or herself with the shield (otherwise, he/she will be hit).

During these attacking routines, the drone attempts to track and follow the HP at

all times, by using as a reference the colors associated with the (t-shirt of the) player, the laser shooter or the shield.

The player has the opportunity to shoot the drone back using his/her laser shooter. Similarly to drone's missile mechanism, the player must keep the shooter in the center of the drone's camera FOV for a given amount of time by using the virtual aiming frame illustrated in Figure 3.13. The pink circle represents the shooter position (top of the prop handed by the player), which must be maintained in the yellow area for triggering the shot.

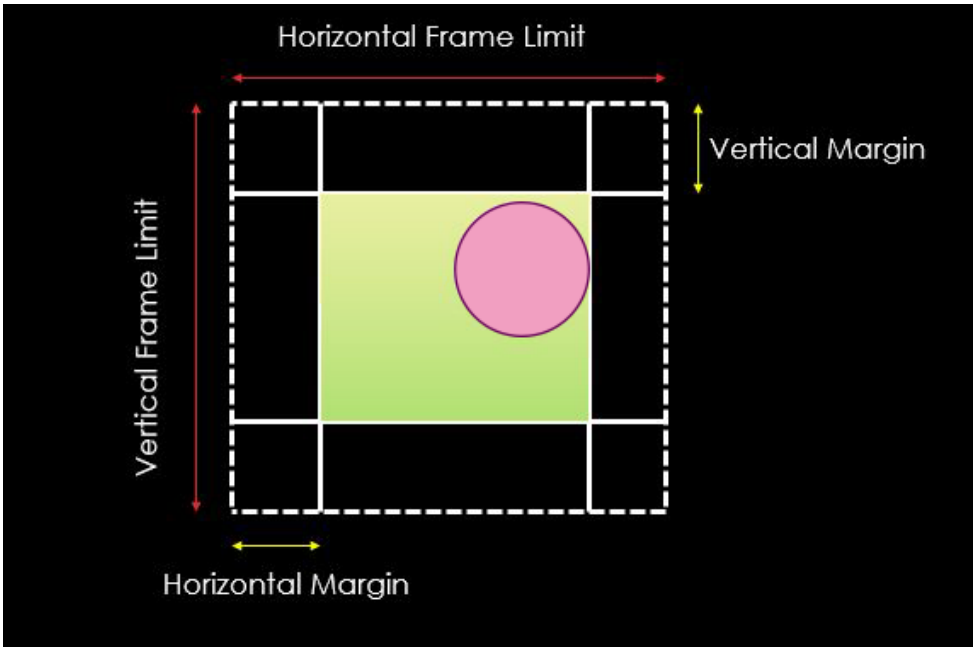


Figure 3.13: Virtual aiming frame.

After a successful hit, the HP's shooter will need to be reloaded. A reloading time was introduced to compel the HP to create defense strategies for the time he or she cannot use the attacking tool. The shooter will alert the player when it is ready again.

At all times, the player can protect himself/herself with the shield. However, in order to avoid stagnation-loop strategies, the shield was set with an expiration time. That means, if the player protects himself/herself for a given amount of time behind the shield, the timer will expire and the drone will immediately attack the defenseless player.

Each score points assigned corresponds to an increase in the game's difficulty by changing the parameters listed below:

- the margins of the aiming frame (Fig. 3.13);
- the interval for releasing bombs;
- the expiration timer for the player's shield.

Experimental Results

In this section, the experimental observations that were carried out to assess the impact on user experience of an autonomous robot exhibiting emotional features are presented. Specifically, a user study involving 16 participants (13 males and 3 females) aged between 22 and 28 years (M = 25.06 SD = 1.73), recruited among university students, was performed.

At the beginning of the experiment, participants were informed that they would have to play a robotic game leveraging a drone, in the three different versions described below.

- A "full" version of the game entailing all the stages and features described above, with the drone exhibiting Autonomous behavior and Emotional features (later referred to as *AE*).
- A version preserving drone's Autonomous behavior and with Non Emotional features (later referred to as *ANE*).
- A version replacing drone's autonomous behavior with a Directly Controlled drone while preserving Emotional features (later referred to as *DCE*). Specifically, the drone is mastered by another player by means of a remote controller and a GUI showing gameplay data and the status of the drone.

For each version, a training phase was performed. Afterwards, participants were requested to play the game in the following order: *ANE*, *AE* and *DCE*. Several videos showing the differences among the three versions are available for download[10].

At the end of the test, participants were requested to fill in a post-test questionnaire by expressing their agreement with 18 statements (of which 15 extracted and adapted from [147], [148], [149] and [150]) on a 5-point Likert scale. Sentences used in this study were grouped in four categories, each of which refers to a specific evaluated aspect, as illustrated in Table 3.1. For consistency reasons, scores for Q2, Q3, and Q7 were inverted so that higher scores reflect positive opinions.

Concerning the *likeability*, a good appreciation of the game overall was expected, with higher ratings for the *AE* version. However, lower scores were expected for the *ANE* version compared to the *AE* and *DCE* ones in terms of *emotional features*'s influence. Concerning the impact on user's experience of drone's *autonomous behaviors*, it was expected to observe a better or at least comparable experience with the *AE* and *ANE* versions compared to the *DCE* one. Lastly, with respect to the *game structure*, the aim was to validate the design phase implemented in this study for creating a game complying with the guidelines discussed above and that is safe, simple and fun.

[10] https://goo.gl/ou8hAq

Table 3.1: Statements in the post-test questionnaire.

Evaluated Aspect	Statement
Likeability	
Q1	I enjoyed the game
Q2	I felt bored
Q3	I found the game difficult
Q4	I would play the game again
Emotional Features	
Q5	I felt like the drone was reacting to my actions
Q6	I felt the drone had a personality
Q7	I felt confused during most of the game
Q8	While I was interacting with the drone I felt as if it was communicating with me
Q9	I could distinguish emotions in the drone
Q10	While I was interacting with the drone I felt involved with it
Autonomous Behaviors	
Q11	I felt in control at all times
Q12	I always knew what the drone was doing
Q13	I perceived the drone as an intelligent opponent
Q14	I felt the drone was moving naturally
Q15	I perceived the drone as competent
Game Structure	
Q16	I was fast at reaching the game's goals
Q17	I think the sounds helped in understanding the game
Q18	I felt safe

Survey data were then analyzed using one-way repeated measures ANOVA test (significance level of 0.05) and a two-tailed paired t-test (significance level of 0.05) in post-hoc analysis, in order to detect any overall differences between the three versions of the game and highlight exactly where these differences were actually occurring.

Results related to individual questions (average values and standard deviation) are reported in Table 3.2, highlighting statistical significance determined with the ANOVA tests. In Figure 3.14, results are aggregated per category for the three game versions considered (+ symbols report ANOVA tests results, i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

Concerning aggregated results, it is immediately evident that, the *AE* and *DCE* versions performed better than the *ANE* one, for all the categories considered. This evidence is also confirmed by observing results related to single questions.

Considering the *emotional features* category, it appears that the introduction of human-like behaviors (*AE* and *DCE* versions) enhanced user experience. These findings were also confirmed by t-tests analysis results reported in Table 3.3, where statistically significant differences were found between the *ANE* and both *AE* and *DCE* versions. Furthermore, it is worth noting that all the sentences belonging to this category proved to be statistically significant (Table 3.2).

Results obtained in terms of participants' evaluations on drone's *autonomous behavior*, appear to show a slight preference for the *AE* version compared to the *DCE* one. However, no statistical significance was reached between the two above versions in the post-hoc analysis (Table 3.3). The fact that drone's *autonomous behavior* did not appear to influence user experience in a noticeable way indicates that the implemented

Table 3.2: Feedback collected via the post-test questionnaire for the three versions of the game, average scores (and standard deviation); statistical significance determined with ANOVA is highlighted (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

Evaluated Aspect	ANE	AE	DCE	*p-value*
Likeability				
Q1	4.44 (0.63)	**4.81** (0.40)	4.50 (0.52)	1.56×10^{-2} (+)
Q2	4.94 (0.25)	4.94 (0.25)	5.00 (0.00)	0.6211
Q3	3.13 (0.81)	3.50 (0.97)	**3.69** (0.87)	1.44×10^{-2} (+)
Q4	4.81 (0.40)	4.94 (0.25)	4.94 (0.25)	0.1349
Emotional Features				
Q5	4.06 (0.85)	4.38 (0.62)	**4.63** (0.72)	3.48×10^{-2} (+)
Q6	3.38 (1.02)	4.00 (0.82)	**4.25** (1.06)	2.14×10^{-3} (++)
Q7	3.81 (0.91)	4.38 (0.89)	**4.63** (0.62)	9.21×10^{-5} (+++)
Q8	3.75 (0.68)	**4.31** (0.60)	4.13 (0.72)	1.30×10^{-3} (++)
Q9	2.56 (1.09)	**3.56** (1.03)	**3.56** (1.21)	1.14×10^{-3} (++)
Q10	4.06 (0.85)	**4.56** (0.51)	4.50 (0.89)	3.66×10^{-2} (+)
Autonomous Behaviors				
Q11	3.63 (0.72)	4.31 (0.70)	**4.48** (0.72)	4.23×10^{-5} (+++)
Q12	3.50 (1.03)	**4.00** (1.15)	3.94 (1.12)	1.22×10^{-2} (+)
Q13	3.94 (0.68)	4.06 (0.57)	4.13 (0.96)	0.6415
Q14	4.06 (0.68)	4.12 (0.72)	3.88 (1.15)	0.5533
Q15	4.31 (0.60)	4.31 (0.70)	4.12 (0.72)	0.4196
Game Structure				
Q16	3.31 (1.08)	4.19 (0.75)	**4.56** (0.63)	5.56×10^{-7} (+++)
Q17	4.31 (1.35)	4.38 (1.20)	4.31 (1.25)	0.7896
Q18	4.44 (0.89)	4.56 (0.51)	4.63 (0.50)	0.4302

Table 3.3: Post-hoc analysis on results collected per question category and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

	ANE vs. AE	AE vs. DCE	ANE vs. DCE
Likeability	t[15] = -3.05, $p = 8.10 \times 10^{-3}$ (++)	t[15] = 0.44, $p = 0.6692$	t[15] = -2.78, $p = 1.39 \times 10^{-2}$ (+)
Emotional Features	t[15] = -3.94, $p = 1.30 \times 10^{-3}$ (++)	t[15] = -0.46, $p = 0.6501$	t[15] = -5.77, $p = 3.69 \times 10^{-5}$ (+++)
Autonomous Behaviors	t[15] = -3.91, $p = 1.41 \times 10^{-3}$ (++)	t[15] = 0.47, $p = 0.6461$	t[15] = -1.63, $p = 0.1233$
Game Structure	t[15] = -2.96, $p = 9.75 \times 10^{-3}$ (++)	t[15] = -2.09, $p = 0.0544$	t[15] = -4.21, $p = 7.52 \times 10^{-4}$ (+++)

AI was able to engage the players at least as much as the remote-controlled version did, confirming the maturity of robotic games including autonomous agents (drones). Participants' judgments regarding the *likeability* and *game structure* categories confirmed

Figure 3.14: Feedback collected per question category. Bar lengths report average values (higher is better), whiskers report standard deviation, whereas + symbols report ANOVA tests results (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

the previous considerations. Specifically, results indicated that, as expected, the *AE* version achieved higher scores than the *ANE* one in terms of *likeability*, although reaching comparable ratings with the *DCE* one. Moreover, the structure of the game was also positively judged by participants, reaching higher scores in the *AE* and *DCE* versions, especially in terms of participants' speed in achieving the game's goals (Q16).

3.1.3 RobotQuest Game

RobotQuest is a Role-Play Game (RPG), where robot and players are companions. They collaborate in order to complete a mission consisting of making the robot cross a battlefield by defeating a number of enemies. Various alternatives were previously explored to define the story of this game (as suggested by the first principle in PIRG guidelines in Section 3.1.1), i.e., a hide & seek game, a robotic minesweeper, an obstacle course, to name a few, by considering features to be included and limitations to live with. *RobotQuest* represents the evolution of preliminary solutions considered.

Game Design

RoboQuest was designed by both sticking to the PIRG principles and adhering to the requirements of phygital games described in Section 3.1.1. By leveraging the guidelines mentioned above and with the aim of fostering players' active collaboration, the game was designed and developed for being played in a MR gaming environment through the use of a robot able to move autonomously in the room-scale play area augmented with projected content (Fig. 3.15). More specifically, physical and real elements are represented by the robot and a set of TUIs, whereas digital elements are projected on the floor. Tangible elements were introduced both to enable proximity interaction with the robot but also to enable collaboration among players during the experience. In fact, human players are enabled to move, deposit and even exchange these tangible props with other players, according to specific game objectives.

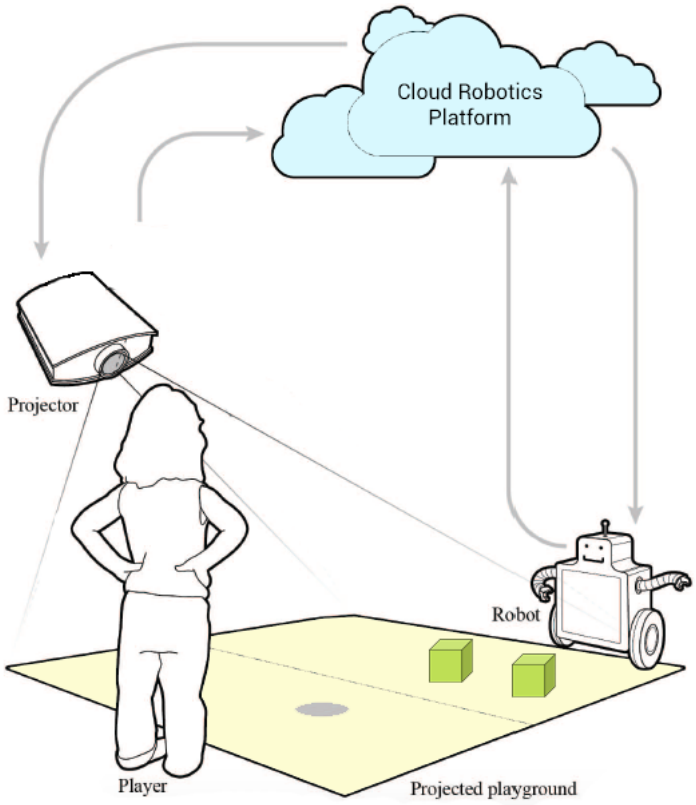


Figure 3.15: *RobotQuest* gaming scenario.

Technology

The *RoboQuest* game was built upon the MR-enabled Phygital Play platform described in [133], which demonstrated already to be flexible enough for implementing a

range of different games [132, 134]. However, some changes were made with respect to the framework defined in [133]. In particular, as described in the game design phase, a set of TUIs were added, by leveraging (hand-held) proximity beacons. In particular, the Estimote's beacons were used, which are endowed with Bluetooth Low Energy (BLE) communication capabilities (Fig. 3.16). Custom covers were 3D-printed for the beacons for showing possible meanings they could assume during the game (Fig. 3.17).

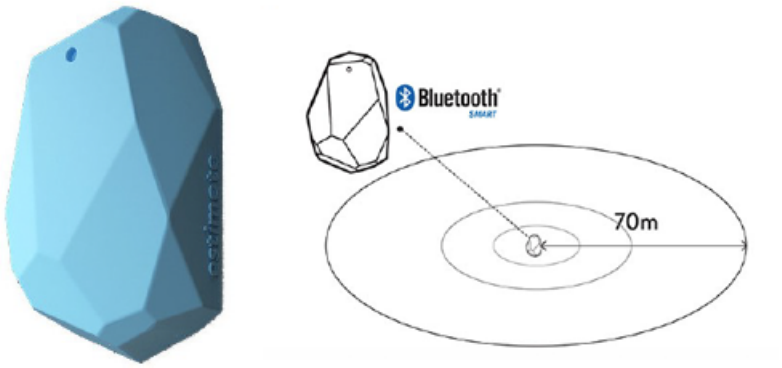


Figure 3.16: Estimote beacon (left), and range of action (right).

Figure 3.17: 3D-printed beacon covers.

Furthermore, compared to the original platform where the robot's tracking was implemented using an external depth-aware camera (according to the second Phygital requirement in Fig. 3.4), in this game, the camera was moved from the external (instrumented) environment to the robot. That is, the robot was endowed with autonomous

navigation capabilities based on the interpretation of projected content. This choice allowed blending the robot's movements with projected gaming contents, thus ending up with a system much less sensitive to occlusions that may occur with multiple players moving in the play area. The shift from an exogenous to an endogenous tracking also allowed to reduce costs associated with the setup based on an external camera.

By leveraging the PIRG principles and in particular the R^3 guideline, a consumer-grade robot was chosen. Specifically, the Parrot's Jumping Sumo was selected (Fig. 3.18). It is a "mini-drone", available on the market since 2015, equipped with two motors for ground movements (up to 7 km/h), plus a third motor to control a spring system used to make jumps up to 80 cm. The motion of the robot is internally controlled by an inertial sensor. The robot is also endowed with a low-resolution (640×480 pixels), low-frame rate (15 fps) RGB camera facing forward. Wi-Fi connectivity is used by the robot for transmitting the camera feed and receiving control commands. The Jumping Sumo was devised to be controlled by human users at short distance by using a hand-held personal device.

Figure 3.18: Mini-drone Jumping Sumo by Parrot.

Since the Jumping Sumo does not come equipped with BLE connectivity (required by the beacons), a board for rapid prototyping (namely, a Raspberry Pi) was mounted on the top of the mini-drone, and equipped with two USB plugs. The board was powered with an external battery (front-mounted, outside the camera's FOV), and connected to two BLE dongles which were used to localize the beacons through Received Signal Strength Indicator (RSSI) filtering.

By sticking to the Phygital requirements (which are aligned to cloud robotics principles), the ROS system was used for connecting the robot to the back-end. The back-end contains services responsible for interpreting image data gathered by the onboard camera (which are processed using OpenCV[11]) and sending control commands back to the robot in order to convince the players that it was actually behaving as an autonomous agent. Communication with the robot was achieved using the ARDrone 3.0 libraries, whereas the game logic and the projection of the digital contents were implemented with the Unity game engine, like in [133].

Implementation

This section describes the architecture developed in this study. As illustrated in Figure 3.19, it is made up of three different main modules, namely, *Robot*, *Cloud-based back-end* and *Game engine*.

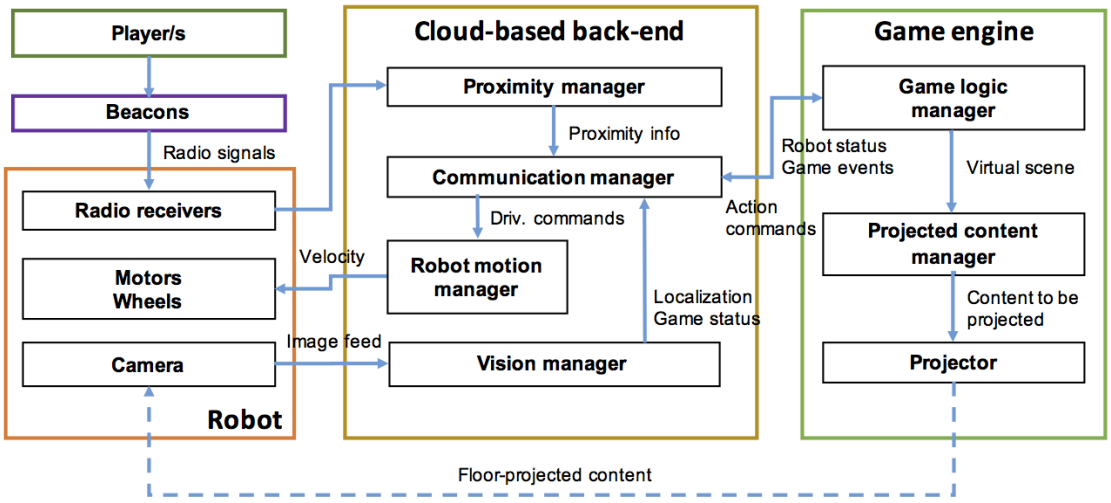


Figure 3.19: Architecture of the robotic-gaming platform supporting the devised *RoboQuest* game concept.

The *Robot* module contains the sensors described in the previous section. The *Cloud-based back-end* module consists of the four components described below.

- Proximity manager: it is the block devoted to collect information about all the beacons detected by the robot with their identifier (ID) and RSSI. This proximity information is then transmitted to the Communication manager module.
- Vision manager: it receives the video stream from the robot's camera as input, processes it by analyzing colors, shapes and brightness of the digital contents

[11] https://opencv.org

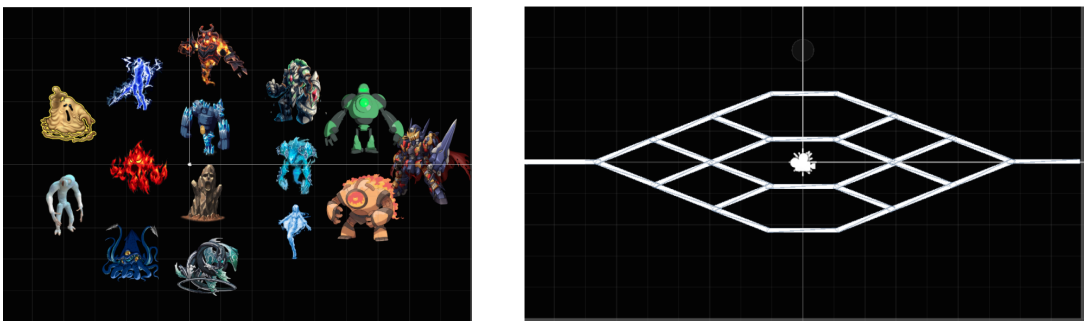
projected on the floor and generates robot's localization data. This data is sent to the Communication manager which is in charge of transmitting it to the Robot motion manager.

- Robot motion manager: it is the module devoted to make the robot move in the game area. The robot's motion can be determined from different data, namely, the output produced by the Visual information manager, the dynamics of the game and the presence of nearby beacons. On the basis of this information and according to internal priorities or parameters, this module produces and sends to the robot an output containing the robot's linear or angular speeds.
- Communication manager: it represents the core of the system, since it is responsible for managing the exchange of information. It gathers information from the Proximity, Vision, and Game logic modules, processes this data and sends back to the Game logic manager the robot's current state as well as the ID of beacons detected by blending this information with the logic of the game.

Lastly, the *Game engine* module includes the Game logic manager and Projected content manager. The Projected content manager provides all the digital elements that visually interact with both the user and the robot according to the logic of the game.

Game Experience

As said before, *RobotQuest* is a RPG, where robot and players are companions. They collaborate in order to complete a mission consisting of making the robot cross a battlefield by defeating a number of enemies (Fig. 3.20a). To this aim, as illustrated in Figure 3.20b, the robot follows a path projected on the floor.



(a) (b)

Figure 3.20: Robot's enemies (a) and path with possible routes (b).

At the beginning of the game, only a portion of the path is shown on the battlefield, but it will grow (during the game) depending on human players' actions. Once the robot is positioned on the path, it starts following it until an intersection is encountered (Fig.

3.21). At this stage, two possible enemies, as well as their names, skills (speed, defenses, attack) and energy levels, are projected on the floor (Fig. 3.22), whereas the robot waits for an input from the human players. Possible enemies are, for instance, the *Icy Golem*, the *Mud Monster*, the *Flame Fury*, etc. Each monster can belong to one of the following categories: ice, fire, ground, water, and electricity.

The human players are required to instruct the robot about the monster to fight with and how to face it. For this reason, players are provided with different support tools (the beacons). Possible tools are the icing ray, the fireball, the electric discharge, etc, as illustrated in Figure 3.17.

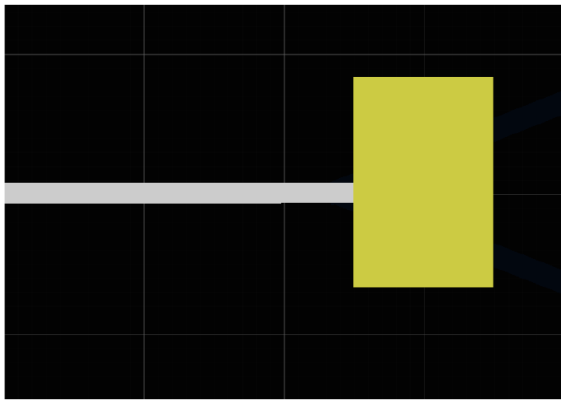


Figure 3.21: Shape and color of an intersection point.

At each round (intersection point on the battlefield), the human players select the monster to fight with by placing the beacons in the two red circles projected on the floor, just under the right and left BLE receivers mounted on the robot (Fig. 3.23).

Each tool is effective against a particular enemy category: for instance, water can be used against fire, fire against ice, ice against ground, ground against electricity, electricity against water, etc. Not all the tools are directly available, but they can be created through so-called "combos". A combo tool is obtained by using more than one beacons at a time, selected through the collaboration of all the human players. Thus, for instance, water can be created by using both fire and ice tools together.

Once support tool/s is/are placed near the robot, the path is modified and grows towards the chosen monster. The battle starts, and the robot simulates shoots towards the enemy and recoils by performing repeated movements back and forth (Fig. 3.24). Projected animations serve as a feedback to the human players for showing how the battle proceeds. The outcome of the battle is determined by the monster chosen and the support tools selected.

In case of a bad choice, the robot is defeated and the player loses the game. Otherwise, the robot may suffer some damages but it can proceed to the next round. At this stage, the game logic will draw the path towards the next intersection point characterized by other two monsters and the robot will start moving again, repeating the game

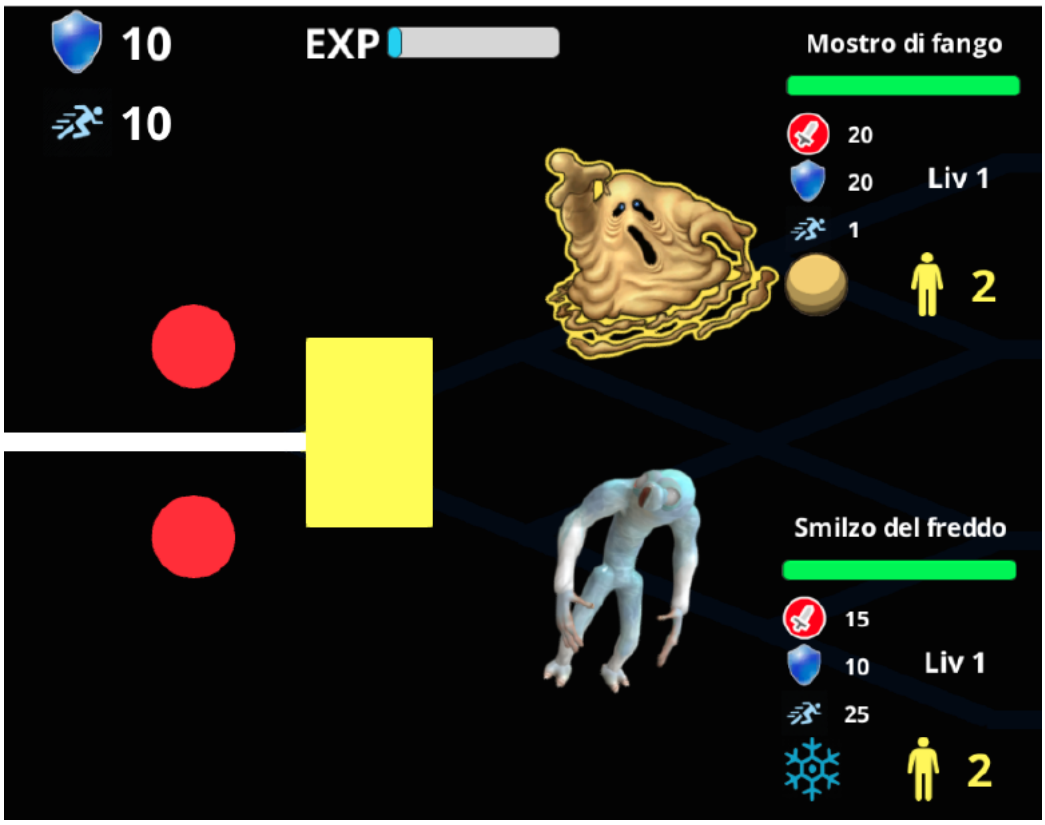


Figure 3.22: Possible enemies according to the logic of the game.

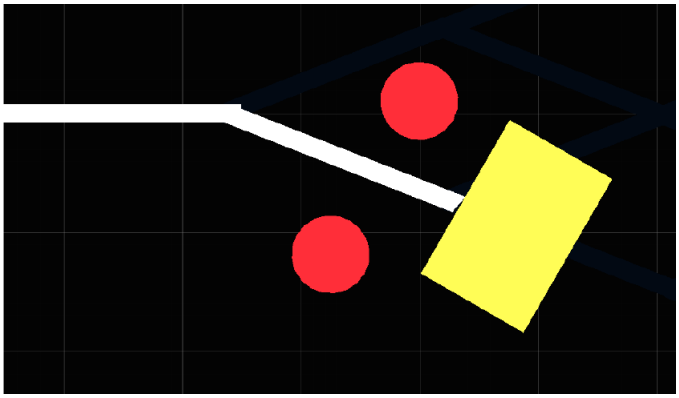


Figure 3.23: Circular areas depicting places where players have to put the beacons.

mechanics above (Fig. 3.25a, 3.25b).

At each round, robot's experience grows. As the game proceeds, challenges become harder. To win the game, a final monster has to be defeated. The difficulty of this final battle depends on the choices made in previous rounds. At the and of the game, after winning the last battle, players can decide whether to start a new and more challenging

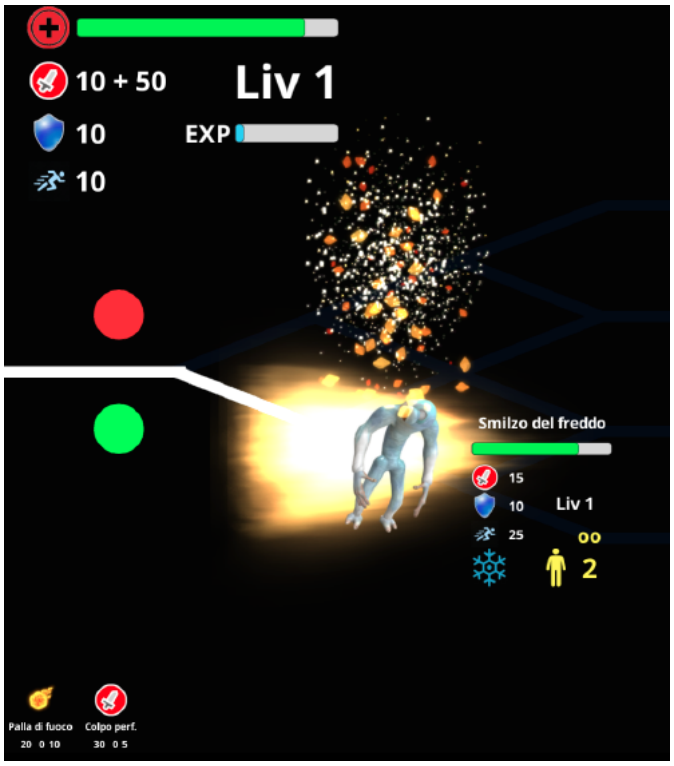
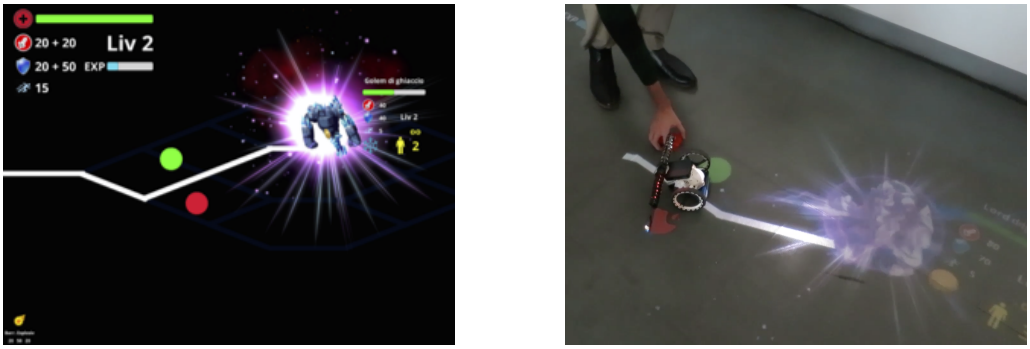


Figure 3.24: Projected animations showing how the battle proceeds.



(a) (b)

Figure 3.25: Gameplay: a) projected path, enemies and battles and b) player placing beacons to choose enemy and use robot's tools.

mission or to terminate the game. A video with gameplay is available for download[12].

In summary, with this robotic game, it was possible to show that an MR Phygital Play platform can be used to set up cloud robotics-based gaming scenarios capable

[12]RobotQuest gameplay: https://goo.gl/3Li4Eu

to achieve most of the objectives identified in this domain. In particular, a seamless interaction between real and virtual game elements is achieved and the RPG logic combined with proximity interaction can be regarded as a means to limit players' sedentary and solitary behaviors. Despite the achievements, the setup suffers from several issues, mainly related to technological limits of the employed components. For instance, data transmission on the Jumping Sumo's Wi-Fi network may suffer from delays up to 500 ms, due to the interferences and poor processing power available onboard. Moreover, the frame rate of the onboard camera could dramatically drop to 1-2 fps in case of low signal, preventing any form of real-time control based on image processing. Other issues were experienced with BLE RSSI filtering, which proved to be poorly accurate in terms of distance estimation (and, hence, of localization and identification). Indeed, many of the above issues could be solved, e.g., by designing and building an *ad hoc* robot and possibly moving part of the processing on it. Nonetheless, this choice would clash with the principle of reusing as much as possible existing hardware and to leverage cloud-powered resources.

3.2 Assistive Robotics

The assistive robotics domain was selected as a possible representative example of applications in colocated spatial proximity pattern in order to study human users' perceptions, reactions, and evaluations of a robotic system (and its interaction paradigms) able to give aid or support to them. To this aim, two use cases involving two different robots able to assist human users by means of different interaction paradigms were considered. The former case consists of a mobile robotic assistant (exhibiting a pyramidal shape) able to guide office visitors towards the desired destination by accompanying them along the path. The latter case consists of a socially interactive robotic assistant (with a humanoid shape) able to assist students to locate a room in an unknown university by providing them with vocal, arm pointing gestures-based directions and exhibiting human-like social behaviors (e.g., gaze, face tracking).

3.2.1 Mobile Robotic Assistant

The mobile robotic assistant use case was selected and explored in order to investigate the natural interaction paradigms between a human user and a robotic assistant exhibiting semi-autonomous behaviors, communicating useful information to the user and with the ability to adjust its actions according to the inputs coming from users. In the following sections, significant works in the considered domain are reviewed. Then, details concerning the developed framework as well as the UIs created will be also illustrated.

Background

In recent years, service robotics and related applications have received much attention due to the useful contribution to the humans' daily life. Robotic assistants will soon serve many assistive roles in the society. Thus, understanding how these robots can interact and communicate with human users is of crucial importance.

In direct interaction with a robotic assistant, an important aspect to take into account is the type of information exchanged between the human user and the robot. The way this information is combined defines the type of interface, which can be either unimodal or multimodal. The former allows users to interact with robots with a single modality at a time, e.g., through a mouse, a keyboard, a joystick or auditory and visual techniques. The latter lets users to interact with robots by combining two or more modes in a complementary or redundant way. Inputs are obtained from different sources and merged based on contextual and temporal constraints for allowing their interpretation. Both approaches have their pros and cons [151]. Unimodal interfaces may not represent the most natural way of communication for human users, and may not provide the same possibility for every user to interact with robots. Notwithstanding, they generally exhibit shorter reaction times. Multimodal techniques, thanks to the ability of the underlying technologies to recognize the natural forms of human communication principles, may provide more flexible interactions between humans and robots [152]. Therefore, NUIs are generally built on such techniques favoring an intuitive and effective collaboration among the entities involved, even though the simultaneity of different modalities can produce ambiguities due to inaccuracies, noise or other factors [153]. A robotic assistant endowed with a natural interaction system would let human users ask for what they need in any way they choose.

Many studies in the HRI domain have investigated the design and evaluation of NUIs exploiting, among others, voice commands, body and hand gestures, tactile feedback, touch input, eye and gaze tracking, etc. For instance, in [154], Skubic et al. explored the way in which a human-robot spatial dialogue (in which the environment is described or referred via spatial relationships) combined with a multimodal interface can lead to a powerful and more natural interaction between a human user and a robotic assistant that regulates its LOA appropriately. Similarly, in [155], authors showed how the integration of multimodal interaction techniques such as a user's visual perception, speech recognition, pointing gestures and head orientation in a mobile assistance robotics platform may improve HRI in terms of ease of use. In [156], a robotic assistant called HERMES equipped with visual, tactile, auditory, kinesthetic and vocal synthesis as well as body movements was designed and realized for natural communication and interaction with humans. The results of a six-month user study showed how the robot's ability to communicate in a multimodal manner as well as the understanding of the situation proved to be the key elements for a user-friendly interaction.

In parallel to these studies, other works focused on improving the effectiveness of multimodal interfaces with AR techniques for interacting with colocated robots. For

instance, in [41], an AR robotic agent, that is an agent equipped with a physical robotic body and an AR virtual avatar appearing on it was developed. Results showed that the use of AR can offer compelling and engaging HRI, as well as easily adaptable and customizable interfaces. In [157], an AR system for interacting with an autonomous mobile robot was developed. The human user is can guide the robot by showing it the target destination on the floor through a special device called "Magic Hand" combined with the voice command "go there". An AR wearable display was used to show in the human's FOV the robot's sensor data and path planning information. Similarly, in [158], two modes for proximal HRI have been proposed to ensure safety and productivity in the collaboration between human users and robots. An AR display is mounted on the human user's head to show the robot's intentions. The human's EEG signals are used to monitor the execution of the task and to adapt the robot's working policy accordingly. This exchange of information allows humans and robots to perform collaborative activities in real-time.

Other works focused on demonstrating that also simulators can be successfully employed as a valid tool for evaluating the efficiency of interaction [159], the users' MW [160], the degree of SA [161] and the level of shared understanding between human users and robots [162, 163] in performing HRI tasks. As a matter of example, the authors of [164] used a 2D robotic simulator called *Stage* to study how to operate multi-robot systems. Although obtained results shown that a 2D simulator may be effective for managing robotic teams in a simplified environment, 3D robotic simulators (such as Webots[13], Gazebo or RoboLogix[14], to name a few) are required for more complex scenarios. Thus, in [165], a 3D virtual environment was exploited for designing and testing a new social communication paradigm for the interaction between a human user and a service robot called *Cero*. Simulation was exploited by authors in order to assess users' reactions and understand how to shape the robot's aspect for improving the model of communication. Similarly, in [166], a service robot called *Virbot* was simulated in a 3D virtual house in order to demonstrate the efficacy of simulation in the experimented domain. In the above cases, the use of simulators endowed with tools "easy to use" for implementing the required logic (like those used for interactive applications or the creation of 3D games) is generally preferred to the use of professional simulators able to reproduce all the physical elements in a faithful way and manage complex robot's dynamics. An example of this approach is reported in [167], where the Modular Open Robots Simulation Engine (MORSE) tool based on the open-source Blender Game Engine (BGE)[15] was presented.

[13] https://www.cyberbotics.com/overview

[14] https://www.robologix.com/.

[15] https://www.blender.org

Proposed Framework

By building on the works discussed in Section 3.2.1, this paragraph presents the activities that have been carried out to develop the simulation framework used for supporting the design and the assessment of natural human-robot interfaces in a robotic office scenario. The simulation framework proposed in this application domain is a three-layered architecture, made up of the *Input*, *Middleware* and *Application* layers illustrated in Figure 3.26.

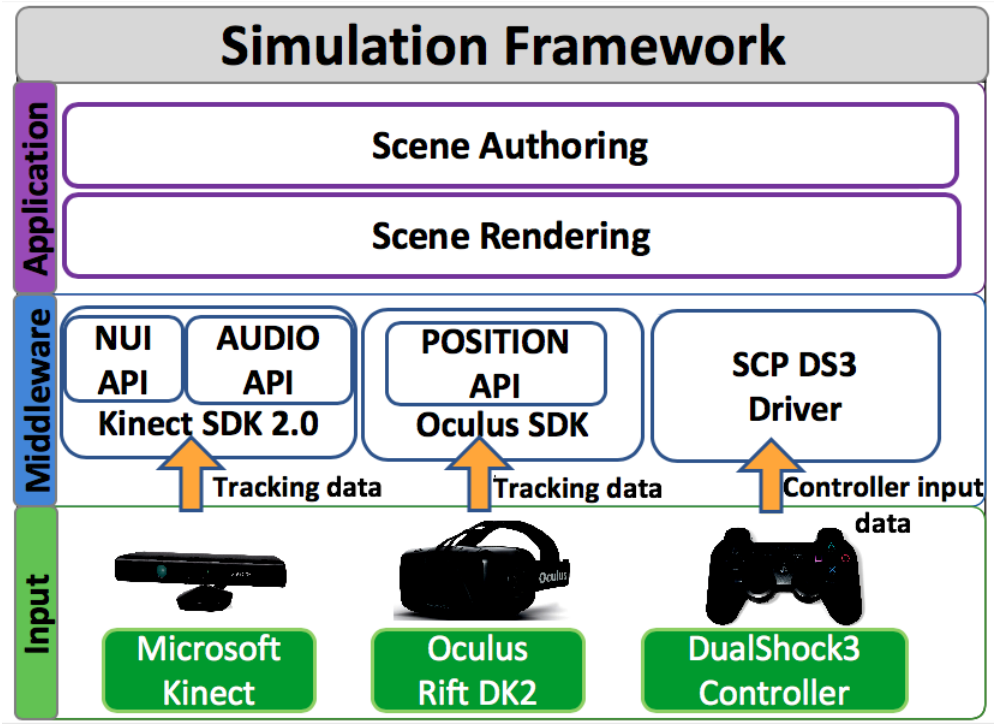


Figure 3.26: Logical architecture of the simulation framework.

The *Input* layer includes the devices exploited by the system for both gathering interaction commands issued by the human user and (in some cases) presenting system outputs. In particular, the DualShock3 controller[16] is employed to collect motion inputs from the human user for letting his or her virtual avatar explore the simulated office environment. The Oculus Rift DK2[17] is used to adjust the avatar's FOV, according to the human user's head rotation information and as a way to select the objects in the scene to interact with (in one of the experimented interfaces). Lastly, the Microsoft Kinect V2[18] is

[16] https://www.playstation.com/en-us/explore/accessories/dualshock-3-ps3/

[17] https://www.oculus.com/en-us/dk2/

[18] https://www.xbox.com/kinect/

exploited for gathering human user's voice commands and (in one of the experimented interfaces) for tracking his or her arm in order to select objects in the scene through the use of pointing gestures.

The *Middleware* layer contains the Application Programming Interfaces (APIs) used to elaborate raw data from the devices in the *Input* layer and to convert them in meaningful information for the top layer. More in detail, the Microsoft Kinect for Windows SDK 2.0[19] was used to implement the speech processing, whereas the Zigfu Development Kit[20] for Unity[21] was used to track the user's body movements. The SCP DS3 management software drivers were used to handle the DualShock3 controller inputs, whereas the Oculus SDK APIs were exploited for obtaining information from motion sensors, determining user's head position in the real world and synchronizing virtual camera's view in the simulation environment.

The *Application* layer completes the framework stack. It consists of two modules, namely, the *Scene Authoring* and the *Scene Rendering*. The former is the module responsible for creating the virtual environment and defining the simulation logic. The latter is the module devoted to managing objects' interactions and behaviors for the real-time execution of the simulation. In the context of this domain, these two modules were implemented through two elements of the development platform Unity, i.e., the editor and the game engine. Details concerning the virtual environment and the simulation logic are described in the paragraphs below.

Virtual Environment This paragraph illustrates the steps pursued for the generation of the virtual office environment selected as a use case for this application domain. In particular, the virtual environment consists of a robotic-enabled office scenario depicting the TIM JOL CRAB headquarters created by means of two different steps, i.e., *modeling* and *importing*.

Modeling represents the first step accomplished in any virtual reality development process devoted to the creation of the 3D models of all the objects in the scene. In this case, each object (e.g., walls, doors, furniture, etc.) was made up of different parts and modeled separately in SketchUp[22] for being animated appropriately, and later finished in Blender for configuring individual visual attributes. An overview of the 3D environment during the modeling phase is illustrated in Figure 3.27. A similar procedure was also pursued for the generation of the virtual robot employed in this domain. Specifically, a 3D model of the *Virgil* robot (described in Section 2.1.2) was created in different parts, i.e., chassis, wheels, tablet, camera, and other sensors. Several virtual characters

[19] https://www.microsoft.com/en-us/download/details.aspx?id=44561

[20] http://zigfu.com/en/zdk/overview/

[21] https://unity3d.com/

[22] http://www.sketchup.com

were introduced into the environment for improving realism as well as for adding obstacles in order to test robot's obstacle avoidance and autonomous navigation capabilities. Lastly, a humanoid avatar (representing the human user in the 3D office) was added to the environment for allowing the user to receive feedback on his or her arm movements.

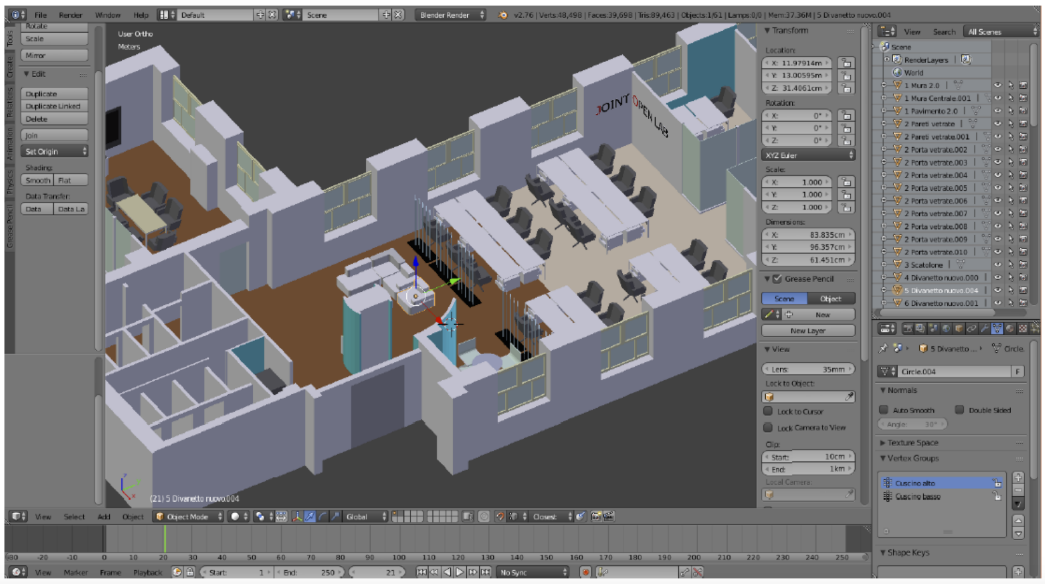


Figure 3.27: 3D model of the office environment in Blender.

Importing is the second step responsible for importing in Unity (as game objects) the 3D models generated in the "modeling" phase and for the configuration of their collision properties in order to make them visible by obstacle avoidance algorithms. *Virgil* model configuration in Unity is illustrated in Figure 3.28.

Simulation Logic This paragraph provides details concerning the operations required to endow 3D objects with the necessary intelligence in Unity. C# scripts were used and associated with the objects in the scene in order to deal with responses to user inputs, guarantee that internal events of the simulation are triggered at the appropriate time and define robot's autonomous behaviors. In particular, scripts linked to the scene were used to recognize user's speech inputs and convert them into commands, as well as to guide the avatar's motion by processing user's body and head tracking data. These scene scripts were also used to manage information shown on a TV screen placed at the entrance of the 3D office for providing the human user with information about interaction modalities available.

Scripts associated with the robot are mainly used for implementing the robot's obstacle avoidance capability, tracking the characters' pose and gestures when moving in the environment, updating either content displayed on the robot's tablet screen or the

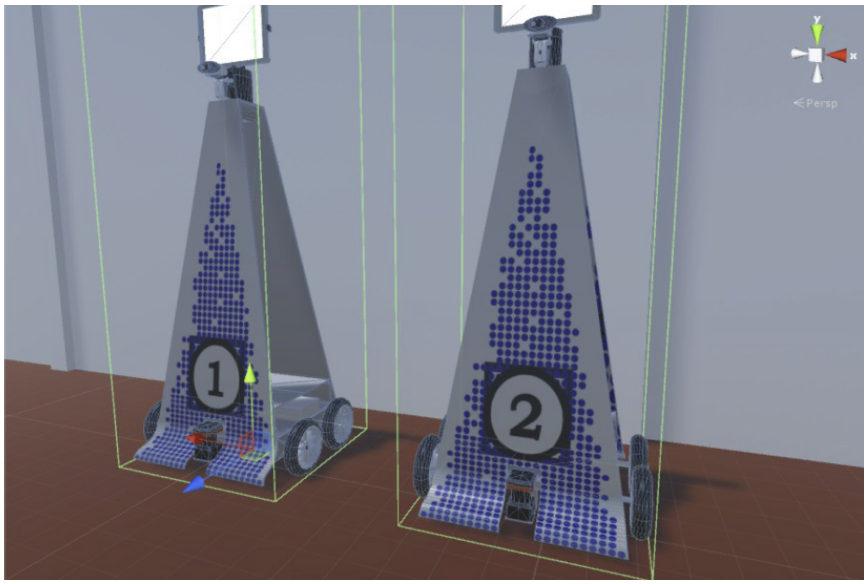


Figure 3.28: 3D model of the *Virgil* robot and configuration of its collision properties in Unity.

AR hints (shown to the user) about the robot's status, keeping the tablet screen pointing towards the avatar of the user, etc. It is worth noting that navigation algorithms developed in this context represent an approximation of those actually used in the real robot (described in Section 2.1.4), since a precise reconstruction of *Virgil*'s behavior in Unity was out of the scope of this study. The same consideration holds also for the tracking capability, which was implemented by using the basic ray-casting technique. Robot's scripts were also used to define the robot's behavior in performing the three possible tasks (later referred to also as function) it could be involved into, i.e., *robotic guide*, *follow me* and *free destination*.

The *robotic guide* function was devised for helping office visitors to reach a given location (e.g., a conference room, an employee's office, etc.) in an unknown environment by asking the robot to serve as a guide. For this reason, robots and their docking station are placed at the entrance of the office environment waiting for visitors and the user starts the simulation with his or her virtual avatar placed close to the entrance door of the considered office. Then, the user selects the robot he or she wants to use as a guide and pronounces the name of the specific destination he or she wants to reach. The robot conducts the user towards the destination by trying to keep his or her avatar in its FOV (at a given distance) and waiting for him or her when the user's avatar is too far. When the target location is reached, the robot comes back to the waiting area.

The *follow me* task depicts the situation in which a human user wants to move the robot to a given position by simply making the robot follow him or her. In details, the user issues the speech command "follow me" to the robot for enabling it to track him or her by moving in the environment.

Lastly, the *free destination* function, as in the *follow me* task, is used in those situations in which the human user wants to move the robot in a specific position. In this case, differently than in the *follow me* function, the target position is explicitly specified through its 3D coordinate in the virtual space. However, depending on the adopted interaction modality, 3D coordinates can be defined by tracking either user's hand gestures or gaze direction.

User Interfaces

In this application domain, two user interfaces were experimented, which differ in the way the selection interactions (required to select the robot or specify destination coordinates) are gathered and in the way information useful for controlling the robot (e.g., commands available, current status, etc.) are presented to the user. One of the interfaces assumes that the human user wears a see-through AR device. For this reason, in the following, it will be referred to as the *AR interface*. In the other interface, information is shown on the robot's tablet, rather than as AR hints. Hence, it will be referred to as *non-AR*, or *NAR interface*. In the *AR interface*, the visualization of the information takes place through the use of AR visual hints, whereas user's gaze/head tracking is used both for the selection of the robot and the definition of a target destination in the 3D environment via the wearable device. It is worth observing that, in this configuration, a head-mounted VR device is exploited to simulate a see-through AR scenario. In the *NAR interface*, the visualization of the information relies on the tablet mounted on the top of the robot, whereas the user's arm gestures are used to define a target destination in the 3D environment by tracking the pointing direction relative to the robot's current position. In both the interfaces, all the other commands are issued through the use of voice.

The next paragraphs provide a detailed description of the two interfaces. A video showing user interaction with the three robot's functions using the two interfaces is also available[23].

AR Interface This configuration assumes that the human user can select the robot to work with by framing it in his or her FOV and a see-through head-mounted AR device endowed with head tracking functionalities. The robot reacts by enabling the "selection mode" in which displays the available commands as AR hints on top of it (Fig. 3.29a) and waits for the human user to select the function to carry out (or to deactivate itself). A yellow light on the robot is used to indicate the activation of the "selection mode", whereas a green light is used for indicating that the robot is actually executing a given function.

When the *robotic guide* function is selected, the robot navigates the environment

[23] https://www.dropbox.com/s/oc8nhe970iqp6v3/video.mp4?dl=0

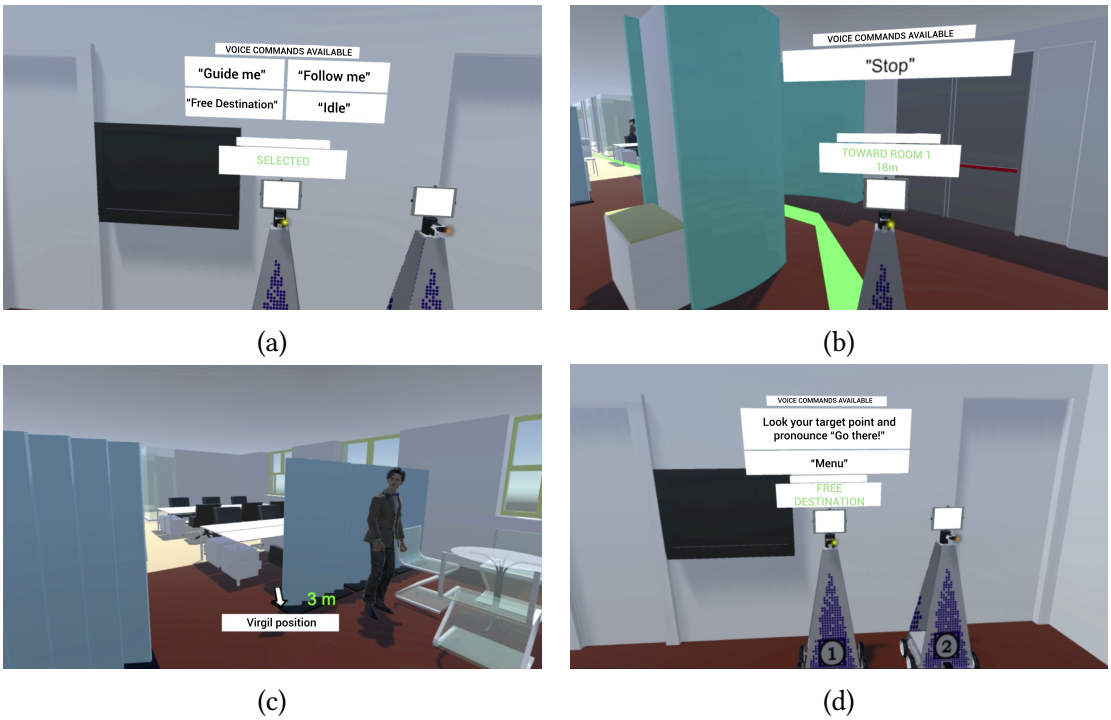
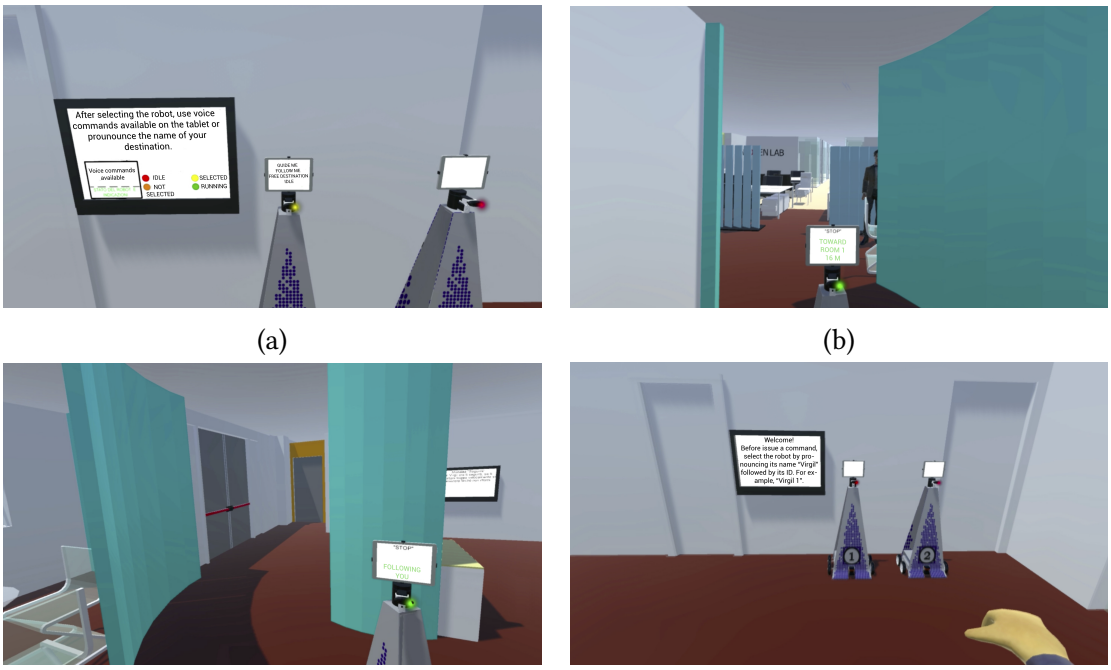


Figure 3.29: Aspect of the AR interface for (a) function selection, and (b)–(d) during operation in each of the three tasks considered.

autonomously towards the target destination. The planned path is displayed on the floor through an AR-based polyline. Useful commands for interacting with the robot (e.g., to stop it) together with distance from destination and current status are also displayed in AR (Fig. 3.29b). When the *follow me* function is activated, the distance and location of the robot relative to the avatar's position are displayed as AR visual hints, independently of where the user's gaze is actually directed (Fig. 3.29c). Lastly, when the *free destination* function is selected, the user is asked to frame in his or her FOV a location in the 3D office environment and issue the vocal command "go there" to make the robot navigate towards such location (Fig. 3.29d).

NAR Interface In this configuration, the selection of the robot to work with occurs by pronouncing the keyword "Virgil", followed by the robot's ID printed on the chassis. This activation method is motivated by the assumption that, in real life, more than one robots could be simultaneously available and ready for assisting users. Hence, different commands from other users to "their" robot could generate "interferences". Therefore, specifying the robot's ID allows to maintain and preserve the user-robot association. Once the robot has been selected and activated, commands and status are shown on the tablet screen. Given the small size of the display, additional information is reported on the TV screen (Fig. 3.30a).



(a) (b)

(c) (d)

Figure 3.30: Aspect of the NAR interface (a) for function selection, and (b)–(d) during operation in each of the three tasks considered.

When the *robotic guide* function is triggered and the robot starts moving, distance to destination and commands available are displayed on the tablet screen, which is always kept oriented towards the user (Fig. 3.30b). When the *follow me* function is activated, no information about robot's location are provided to the user (Fig. 3.30c). Lastly, when the *free destination* function is selected, the user specifies target destination's coordinates by indicating a given point in the 3D space with his or her arm (Fig. 3.30d).

Experimental Results

In this section, experimental observations that were carried out to assess the proposed AR and NAR interfaces are presented. Specifically, a user study involving 36 participants selected from the students of Politecnico di Torino (28 males and 8 females) aged between 19 and 35 years (M = 25.03 SD = 3.38) was performed. According to declarations collected, 23% had previous experience with NUIs (mainly with Leap Motion Controller[24], Microsoft Kinect as well as with Microsoft Cortana[25] voice assistants and

[24]http://www.leapmotion.com/

[25]http://windows.microsoft.com/en-us/windows-10/getstarted-what-is-cortana

Apple Siri[26]), 24% had experimented with AR or VR applications (by using Google Glass[27], Google Cardboard[28] or Samsung Gear[29]) and 54% of them had used already 3D applications and games. Configuration of technologies used for experimental tests is reported in Fig. 3.31.

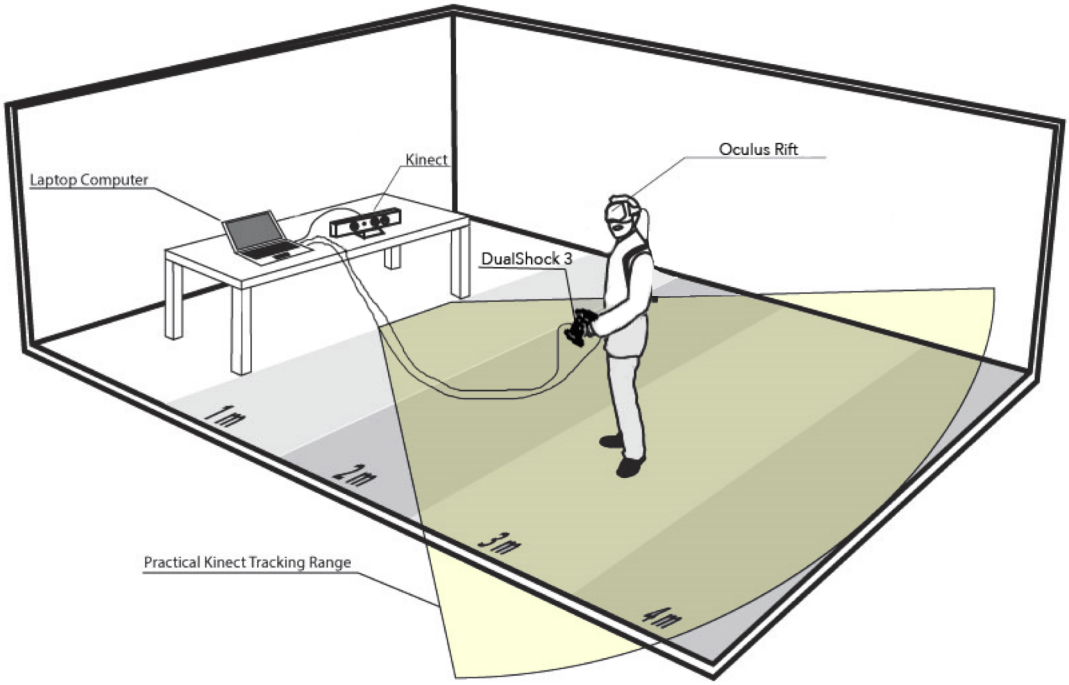


Figure 3.31: Configuration of technologies used to manage user's interaction with the simulation framework and carry out the experiments.

Study participants were divided into two groups, namely, AR vs. NAR, with an equal distribution on the two interfaces. Afterwards, participants of both groups were informed that they would have to perform the three tasks defined in this study by interacting with an assistant robot within a virtual office environment. Then, participants in the AR group were told that supporting information for interacting with the robot would be provided in AR. Participants in the NAR group were informed that supporting information would be reported on the tablet mounted on the top of the robot or on the TV screen placed at the entrance of the virtual environment. To compensate for possible learning effects, a random order was used for choosing the sequence of the functions

[26] http://www.apple.com/ios/siri/

[27] http://www.google.it/glass/start/

[28] http://www.google.com/get/cardboard/

[29] http://www.samsung.com/us/mobile/wearable-tech

to be experienced.

During the experiment, the time required to complete each task was recorded. After having performed all the tasks, each participant was asked to compile a usability questionnaire split in two parts.

The first part was created by considering the After-Scenario Questionnaire (ASQ) developed in [168]. ASQ is a three-item questionnaire (i.e., ease of use in completing the task, the amount of time required and support information provided) with the aim to evaluate participants' satisfaction after the completion of the experiment on a 5-point Likert scale. This part was the same for both participants in the AR group and for those in the NAR one.

The second part was designed by considering the SASSI methodology [80] already used in the studies described in Section 2.1.6 and adapting it to let participants judge the proposed interaction means by expressing their agreement on a 5-point Likert scale. As in the studies described in Section 2.1.6, scores for CD and *annoyance* usability factors were inverted (thus, higher scores have to be interpreted as being more positive).

Data gathered from the study were then analyzed using a two-tailed independent two-sample t-test (significance level of 0.05) in order to detect any significant differences among the participant evaluations of the considered interfaces.

Objective evaluations, in terms of time required to complete the tasks, are illustrated in Figure 3.32. It is clear from the chart, that participants who experimented the NAR interface took more time to complete each function compared to those who experimented the AR one. This evidence was also confirmed by the statistical significance analysis results illustrated in Table 3.4, where differences between the AR and NAR interfaces were found to be significant for all the tasks.

Table 3.4: Statistical significance analysis on completion time results performed with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	AR vs. NAR
Robotic Guide	$t[34] = 3.87$, $p = 4.65 \times 10^{-4}$ (+++)
Follow Me	$t[34] = 3.43$, $p = 1.60 \times 10^{-3}$ (++)
Free Destination	$t[34] = 3.42$, $p = 1.65 \times 10^{-3}$ (++)

Subjective evaluations concerning overall participants' satisfaction collected through the ASQ are illustrated in Figure 3.33. According to the chart, the AR interface performed better than the NAR one for all the usability factors. However, only the differences between the two interfaces in terms of ease of use were found to be statistically significant, as reported in Table 3.5.

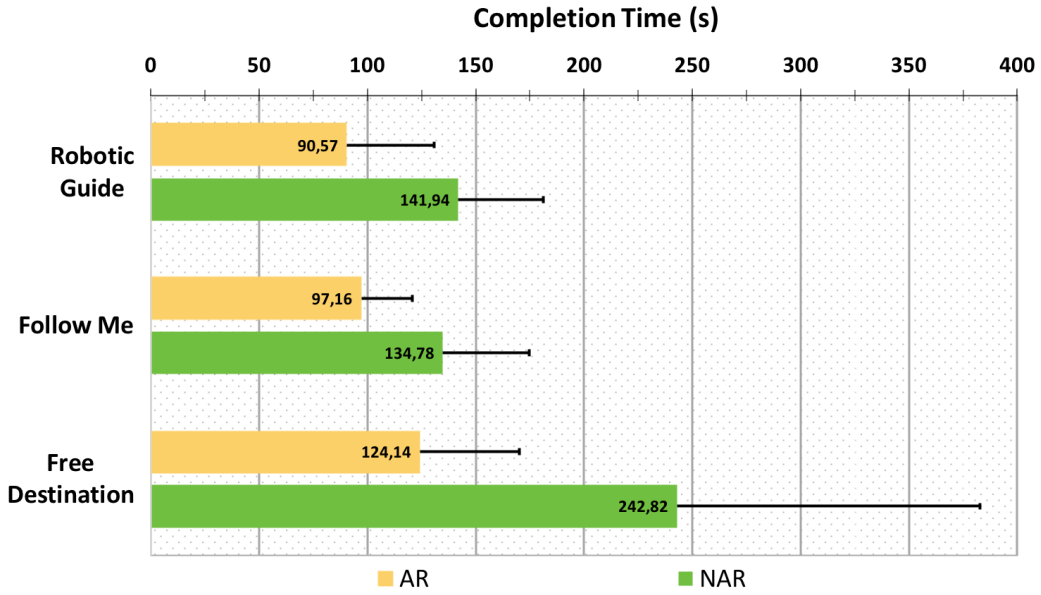


Figure 3.32: Results in terms of time required to complete the three functions using the two considered interfaces. Bar lengths report average values (lower is better)whereas whiskers report standard deviation.

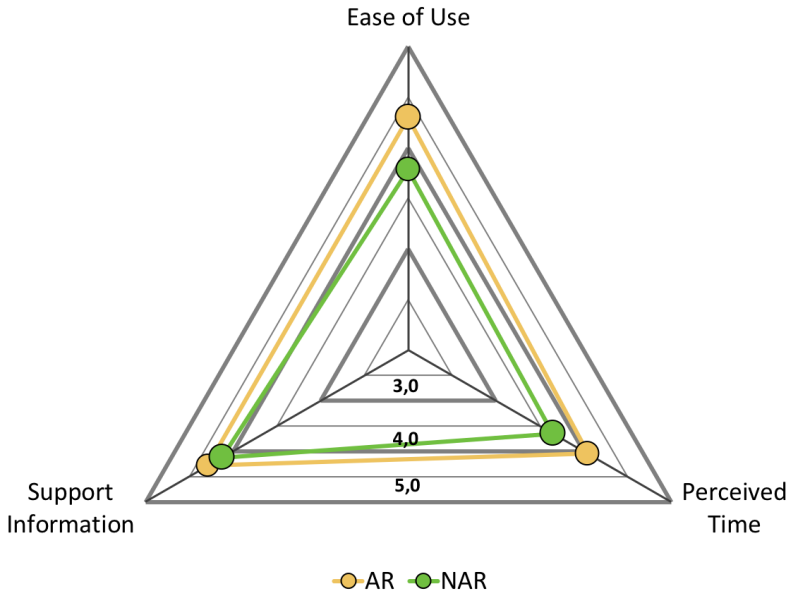


Figure 3.33: Results in terms of overall satisfaction (ASQ) for the two considered interfaces. Circle position reports average values (higher is better), whereas circle dimension reports standard deviation.

Table 3.5: Statistical significance analysis on overall satisfaction results performed with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	AR vs. NAR
Ease of Use	t[34] = 2.38, $p = 2.32 \times 10^{-2}$ (+)
Perceived Time	t[34] = 1.80, $p = 0.0794$
Support Information	t[34] = 0.73, $p = 0.4729$

Results obtained considering the single questions of the ASQ, i.e., ease of use, perceived time and support information provided in completing each function appear to describe an almost comparable situation, as illustrated in Figures 3.34, 3.35 and 3.36, respectively. In fact, the AR interface performed better than the NAR one both in terms of ease of use (Fig. 3.34) and perceived task duration (Fig. 3.35) in all three functions. As illustrated in Table 3.6, t-test analysis corroborated a statistical significance concerning ease of use only for the *follow me* and *robotic guide* tasks. However, despite differences between the AR and NAR interfaces were pronounced no statistically significant differences were found in terms of perceived task duration (Table 3.7).

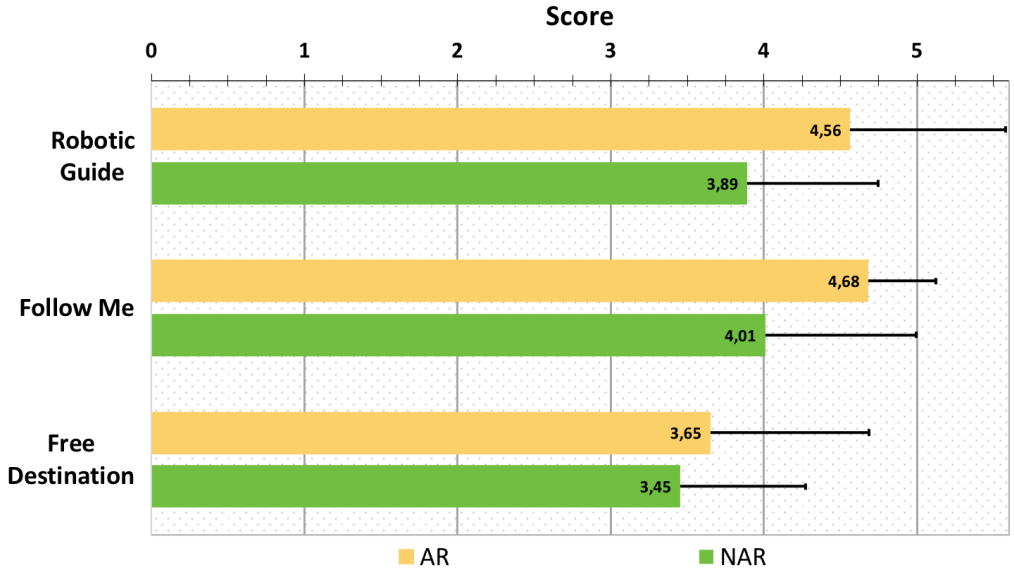


Figure 3.34: Results in terms of ease of use for the three functions using the two interfaces. Bar lengths reports average values (higher is better), whereas whiskers report standard deviation.

Table 3.6: Analysis on ease of use for the three functions and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	AR vs. NAR
Robotic Guide	$t[34] = 2.16, p = 3.82 \times 10^{-2}$ (+)
Follow Me	$t[34] = 2.65, p = 1.20 \times 10^{-2}$ (+)
Free Destination	$t[34] = 0.64, p = 0.5287$

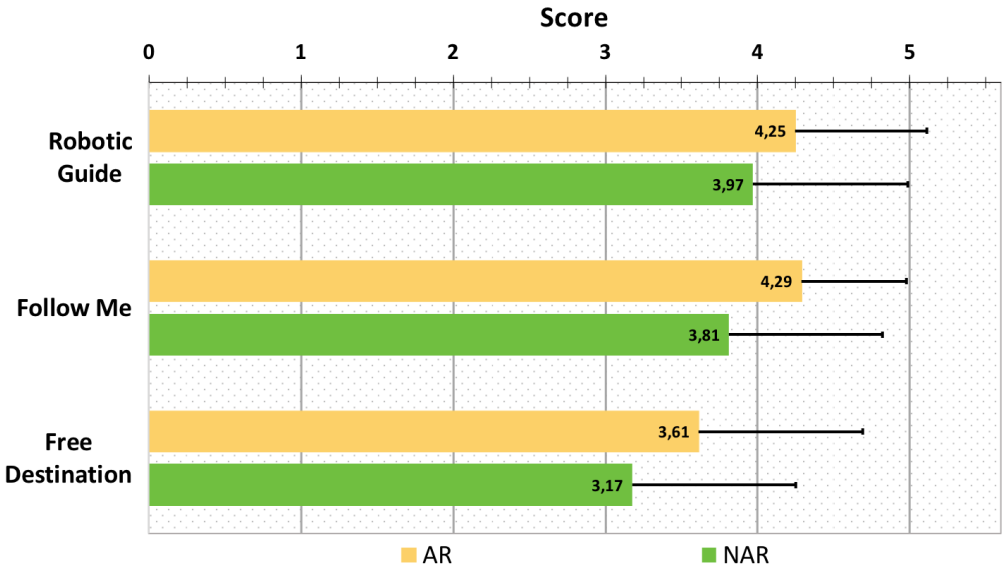


Figure 3.35: Results in terms of perceived time requested in the three functions using the two interfaces. Bar lengths reports average values (higher is better), whereas whiskers report standard deviation.

Table 3.7: Analysis on perceived time results and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	AR vs. NAR
Robotic Guide	$t[34] = 0.88, p = 0.3836$
Follow Me	$t[34] = 1.65, p = 0.1083$
Free Destination	$t[34] = 1.22, p = 0.2327$

Concerning the support information provided by the two interfaces, it appears from Figure 3.36, that the AR interface was judged more positively only in the *robotic guide*

and *follow me* functions, whereas the NAR interface was preferred than the AR one in the *free destination* function.

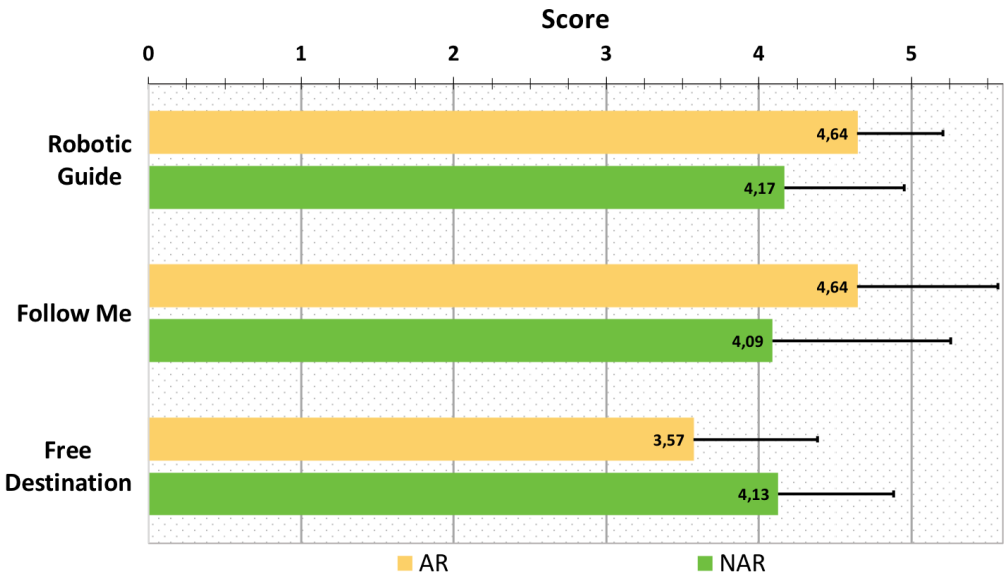


Figure 3.36: Results in terms of support information provided in the three functions using the two interfaces. Bar lengths reports average values (higher is better), whereas whiskers report standard deviation.

These findings were also confirmed by t-test analysis results shown in Table 3.8, where statistically significant differences were found both in the *robotic guide* and *free destination* functions.

Table 3.8: Analysis on support information results and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	AR vs. NAR
Robotic Guide	$t[34] = 2.09,\ p = 4.37 \times 10^{-2}$ (+)
Follow Me	$t[34] = 1.58,\ p = 0.1233$
Free Destination	$t[34] = 2.12,\ p = 4.12 \times 10^{-2}$ (+)

Based on the comments collected during the experiment, this preference for the NAR interface compared to the AR one in performing the *free destination* function seems to be due to the fact that with the AR interface the participant had to use the gaze both to obtain the information necessary to perform the task (looking at the robot) and to define the target destination he or she wanted the robot to reach. Therefore, a common scenario consisted of participant first framing the target point (commanding the robot

to "go there"), and immediately after looked at the robot (for checking to have issued the correct command) thus setting another point in the 3D space to be reached. In the NAR interface, participants could define the target destination using arm pointing gestures while kept looking at the robot to read the support information shown on the tablet.

Data about participant evaluations collected through the adapted SASSI methodology are illustrated in Figure 3.37.

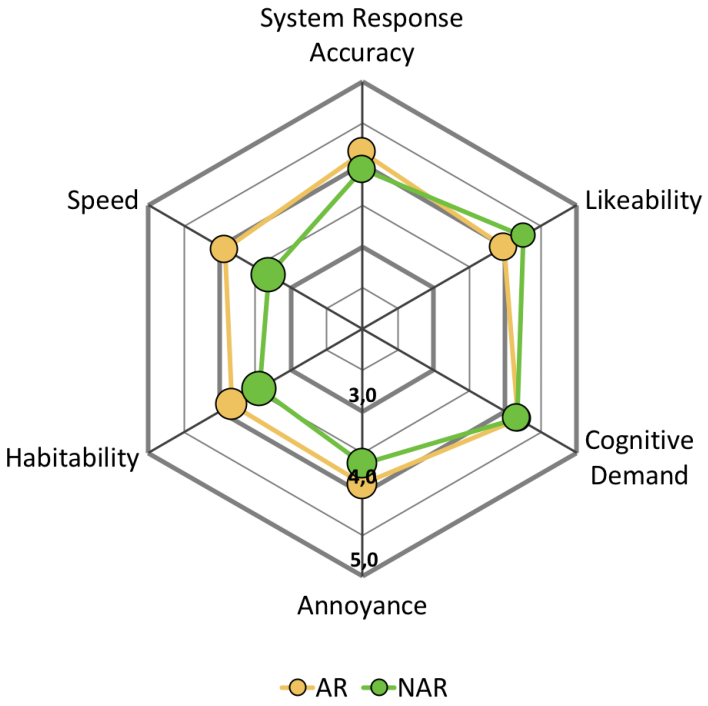


Figure 3.37: Results obtained by applying the adapted SASSI methodology. Circle position reports average values (higher is better), whereas circle dimension is used to show standard deviation.

At first sight, it can be observed that the AR interface performed better or equal than the NAR one, except for the *likeability* factor. In fact, the interaction with the robot was judged more efficient and robust when the AR interface was experienced (*system response accuracy*), whereas the NAR interface was evaluated more repetitive and boring (*annoyance*) than the AR one. In terms of *habitability*, it appears that with the AR interface it was easier for participants to keep track of where they were in the interaction flow compared to the NAR one. Regarding the *speed* factor, participants rated the robot equipped with the AR interface to respond faster to their inputs than with the NAR one. However, t-test analysis results showed a statistical significance only for the *speed* factor (Table 3.9), thus confirming the improved performance in terms of completion time obtained with the AR interface (Fig. 3.32).

Table 3.9: Analysis on SASSI results and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

	AR vs. NAR
System Response Accuracy	t[34] = 0.93, p = 0.3608
Likeability	t[34] = 1.40, p = 0.1711
Cognitive Demand	t[34] = 0.07, p = 0.9426
Annoyance	t[34] = 0.94, p = 0.3538
Habitability	t[34] = 1.21, p = 0.2348
Speed	t[34] = 2.13, $p = 4.12 \times 10^{-2}$ (+)

Based on feedback collected during the experiment, this result seems to be due to the fact that, with the NAR interface, participants had to use their voice for repeatedly activating the robot, whereas with the AR interface they simply had to frame it in the FOV. Furthermore, in the NAR interface, support information was available only on the tablet screen mounted on top of the robot, whereas in the AR interface they were always available and easy to obtain by looking at the robot even at distance.

3.2.2 Socially Interactive Robotic Assistant

The socially interactive robotic assistant use case was explored in order to study users acceptability and perception of robots featuring or not a human-like appearance and endowed or not with social behaviors. In particular, the receptionist scenario, where a robot assists people in finding the places of interest by giving them directions, was selected as a specific use case since it can be considered a good benchmark for social assistive robotics applications, given its implications on HRI. In the following sections, relevant works in the field of reception, direction-giving, and wayfinding applications will be reviewed. Then, receptionist systems considered in this domain, as well as the hypotheses formulated and the methodology adopted to perform the experimental tests, will be introduced.

Background

Robots are rapidly advancing towards becoming autonomous and skilled entities in a wide range of environments, and it is likely that more and more people will soon interact with robots in their daily lives. In this way, it essential that robots are designed to be easy to use and interact with, reducing the need for people and environments to adapt to them.

Emerging research in HRI suggests that people have a strong natural tendency to treat robots as social entities, anthropomorphizing, zoomorphizing, and in general by attributing social characteristics and roles to them. Thus, robots exhibiting more

human-like appearances as well as robots endowed with social behaviors may reasonably play a fundamental role with respect to other robots, since these features may contribute at making them more acceptable [2].

Although many robots with human features are already available, research activities are needed to adapt their behavior to the specific tasks they are expected to perform in the selected context, in order to guarantee consistency with end-users' needs and make interaction as natural as possible [169].

In the scenario considered in this application domain, a robotic receptionist can be intended both as a *Socially Interactive Robot* (SIR), a definition coined by Fong et al. to describe robots endowed with social interaction skills as main features [170], and as *Socially Assisted Robot* (SAR), since it exploits social interaction for providing assistance to human users [171].

A number of works have already explored direction-giving and wayfinding systems and in particular, the robotic receptionist domain, by developing different solutions using, e.g., deictic (in the following often intended as arm pointing) gestures, route tracing on a map, audio feedback, unembodied systems without social interaction skills (like audio-maps), socially interactive embodied (physical or virtual) systems, etc. However, an approach that can be regarded as the ultimate solution to perform receptionist tasks has not been identified yet.

In the next paragraphs, key research studies pertaining functionalities and embodiments of reception, wayfinding, and direction-giving solutions are reviewed.

Maps Nowadays, the most familiar navigation means for wayfinding tasks is represented by maps, since the type of information and its meaning are easily understandable concepts for most people [172]. Specifically, when used for navigation activities, maps generally adopt conventions, such as people localization and orientation. For instance, the "*You Are Here*" (YAH) maps exploiting a symbol to show the user's position with respect to the surrounding environment were proposed by Levin in [173], whereas the "track-up" or "forward-up" orientation, in which the upper part of the map is aligned with the forward direction of movement, was presented in [174].

Various methods have been experimented both by the industry and academics to enable the navigation of an environment through map-based direction giving systems. For instance, in [175], a digital map based on GPS (showing the user's current position, dynamically updated while moving in space, and a static route to the destination) was compared with a paper map (showing the user's initial position and the destination of interest) in wayfinding activities. The comparison also included direct experience, where the knowledge of the environment is acquired by physically walking in it. In this work, authors' aim was to study the way in which spatial navigation information is acquired by the users. Results showed differences in terms of speed: users with the map were slower than users with direct experience. Moreover, due to the lack of information about the surrounding environment, users with the GPS-based map spent more time reaching the destination and made more direction errors than the other two groups.

The small size of the map did not allow users seeing the complete route, although at the same time, the GPS-based system allowed them to follow it by simply observing their current position on the screen. In the paper map, a wider portion of the environment was actually shown by displaying together the beginning and the end of the route. However, the need for users to adjust their travel direction with the orientation of the map negatively affected performance. In [176], the navigation aids provided by route maps (including inter-turn mileages, landmarks, etc.), by voice directions, as well as by the combination of the two, were investigated in terms of effectiveness. Authors found that users performed better when listening auditory directional signals compared to using route maps, both in terms of direction errors and time needed to reach the destination. This result showed that, even though auditory directional signals needed the exact knowledge of users' current position to be effective, the way in which spatial information was acquired affected users' performance.

A similar and more recent study comparing different methods for providing route instructions ranging from spoken/written directions to 2D/3D map visualizations is reported in [177]. Results showed that participants preferred the 2D map solution with respect to the other methods, both because they perceived to be able to locate and reach the destination faster, but also because 2D map represents the navigation means they were more familiar with. These findings highlight some disadvantages of the above solutions. More precisely, speech and text instructions do not demand visual attention from users, but they lack contextual information and require accurate positional information to work. 3D maps provide information about the surrounding environment, but have significant requirements from the technological point of view. 2D maps incorporate contextual information and compensate possibly missing positional information by requiring fewer resources.

Another interesting study is reported in [178]: here, a digital map was compared to a paper map. Results showed beneficial effects for both the proposed solutions, since they satisfy different needs. In particular, the paper map, due to its presentation perspective and size, provides the user with a better overview of the surrounding environment and, thus, a better understanding. Digital map lets the users interact with, query and zoom it, and also provides the users with the possibility to spot the current location on it. In [179], the authors' interest was focused on the presentation of spatial information by comparing two different graphical representations, i.e., a true aerial picture and a generalized and abstract map. Obtained results showed that the minimalistic map allowed users to perform better compared to the real map which exhibited too cluttered features not easily discriminable by users. In [180], Fewings stated that whether static, interactive 2D or 3D, maps for indoor environments should always provide features such as size, lettering, user's position visualization, color and landmarks, etc. in order to maximize comprehensibility.

In parallel to the above works, other studies focused on free-standing units called "*kiosks*" commonly found in public areas. These systems, generally equipped with interactive displays, allow users to receive intuitive information about routes to travel in

order to reach a given destination based on their current location. For instance, in [181], a kiosk system called *Touch 'n' Speak* provided users with information about restaurants located in a specific area through a touch-screen map. Food or price selection could be performed by using some filters with voice commands and/or touch inputs. Even though this solution uses multimodal inputs, the speech recognition technique does not accept arbitrary inputs, since it relies on a small vocabulary. In [182], the authors presented *Calisto*, a system able to allow users to drag&drop interesting contents on their smartphone after plugging it to the touch-screen kiosk. Users could communicate with the system through both voice and touch commands. In [183], a multimodal kiosk called *MUSE* designed for shopping malls was introduced. Users could request store locations in two different ways: by connecting their mobile devices to the kiosk through a QR code displayed on the screen or via a touch menu. The mobile device allowed users to issue voice or text commands to the kiosk, whereas the menu was used to search for stores' information through a list (showing shops' names grouped by category or in alphabetical order). Directions were displayed on an animated map. In [184], a 3D touch-screen kiosk called *i-Showcase* was employed in a shopping mall in order to let users find the correct route to travel from their current location to the intended store. For this purpose, users could receive information via a search bar, a category menu or shortcut menu. Results showed a high success rate in completing the task especially for those who were familiar with the kiosk experience. Moreover, the kiosk was judged by users to be satisfactory, easy to use and useful to get the right way to go. The three-dimensional features were not considered as particularly helpful. In [185], a touch-screen kiosk was used in a health center to facilitate both staff and patients to find a particular place. Users could search the route through an interface exhibiting a set of icons (representing places) arranged based on spatial criteria. After selecting the intended target location, the 2D map displayed an animated path combined with photographs, text, and (optionally) voice directions. Experimental results showed a high users' success rate in completing the task without further assistance, with a strong appreciation for the audio feedback.

Works considered in the above review were selected as representative examples of the developments made by the academics. From a commercial point of view, it can be observed that systems relying on maps devised for wayfinding tasks are usually based on a simplified version of architectural blueprints [186]. Some examples are HERE WeGo [187], Cartogram [188], Google Maps [189] and Mapwize [190]. Similarly, commercial wayfinding kiosks are often endowed with touch-screen displays showing an interactive YAH map and animated paths drawn on it. The map can be consulted by the users, e.g., to search for places possibly grouped by category or listed alphabetically [191, 192].

Virtual Receptionist In recent years, the possibility to use virtual agents for assisting human users in their daily life activities has been extensively studied. Historically, two main features have been taken into particular account in designing these agents: the choice to equip agents with an embodiment [193, 194] and the adoption of NUIs in

human-agent interaction (including facial expressions, speech- and gesture-based communications, gaze tracking, etc.). Agents featuring these characteristics are generally referred to as *Embodied Conversational Agents* (ECAs) [195].

Several works have explored the use of ECAs in direction-giving and reception applications. As a matter of example, in [196], a virtual guide showed on a screen is used in a theater to assist visitors in wayfinding tasks. When the user requests the position of a given place, the virtual guide gives him or her indications on how to reach it by using speech and in-the-air arm pointing gestures. Then, an animated path towards the destination is displayed on a very minimalistic 2D map placed beside (to the right of) the virtual guide. The main limitation of this solution is represented by the different reference systems adopted by the ECA to indicate the route with arm gestures and by the map to display the path. In fact, arm gestures are mirrored with respect to visitor's perspective, since they are executed in the agent's (speaker's) perspective. Similarly, the path showed on the map is illustrated in the visitor's (listener's) perspective, thus, mirrored with respect to agent's perspective. This mismatch between the two reference systems disoriented the users.

In [197, 198, 199] a similar configuration is presented: a kiosk endowed with a conversational virtual robot (displayed on a screen) named *MACK* (later evolved into *NUMACK* [200]) is used to provide route information to human users. The "bridge" between the real and the virtual worlds is represented by a small paper map placed on a table in front of the screen where the virtual robot is displayed. *MACK* can help users with orientation tasks by resorting to three different interaction modes: voice, hand/arm gestures with head movements and eye gaze, and an LCD projector highlighting a region of the map in order to allow the agent to refer to it. Users can interact with the agent through a pen on the map or using voice. According to the authors, the kiosk configuration made it difficult for users to understand when to look at the agent and when to look at the map, even though the system succeeded in engaging and entertaining users. Specifically, authors claimed that when users paid attention to the agent's actions, they did not look at the map and vice-versa. They also reported that an interesting evolution of the proposed solution may focus on investigating how the users are involved by ECA's gestures, the attention they dedicate to a map as well as their preferences or performance with different configurations.

Another research direction was oriented towards the level of engagement between a virtual receptionist and a human user, as well as towards the social aspects of the interaction. For instance, in [201], the authors introduced *Marve*, a virtual receptionist exploiting face detection capabilities to recognize users of a computer laboratory and greet them with voice. Although the agent was not able to provide indications, *Marve* could deliver messages, talk about movies or weather, tell jokes and understand humans' speech. Experimental results showed that, despite *Marve* was perceived as a true social being, the level of expertise required to interact with it represented the main drawback. Most of the users were indeed computer scientists working in the laboratory, with the exception of the building guardian who interacted with *Marve* more often and

judged it more engaging than any other user. The authors concluded that virtual agents' social features and the interaction between humans and robot might be judged and perceived in a diverse way by different users.

Robotic Receptionist The receptionist role has been extensively studied in the field of service robotics. In particular, focusing on static, humanoid robot receptionist systems, multiple implementations have been developed, which can be grouped in two main categories according to the level of embodiment: robots with a physical body and a virtual face displayed on a screen, and fully physical robots.

Examples of the former category were provided in [202, 203, 204], where a robotic receptionist named *Valerie*, to give directions within a university using the voice and exhibit its personality through a number of pre-defined facial expressions was introduced. Voice commands and text typed on a keyboard could be used by users for interacting with the robot. In this work, the authors' aim was to study the robot's social attraction and visitors' engagement when involved in interaction tasks with the robot. Experimental results showed that even though users were engaged with and attracted by the robot, interaction modalities had to be improved since they were limited to keyboard inputs in some circumstances. In [205], another robotic receptionist system called *AVARI* was presented. The robot was used in a university scenario for answering questions about professors' office location and email address by using a knowledge repository and voice interactions. The drawback of this solution was the timing of the interaction, since the robot's answers were not synchronized with users' questions. As studied in [206], the timing of the interactions may affect turn taking and engagement as well as social strategies and task performance. Other studies were oriented towards intercultural phenomena in order to analyze how a robot's gender, politeness and language may improve users' level of attention and acceptance [207, 208]. Experimental results showed that design choices should be culture-aware: the social presence and acceptance of the robotic receptionists may be improved by the use of local language, polite behavior, a female appearance and voice.

Similar studies have been conducted also for fully physical robots. A number of works explored the role of robotic receptionists by focusing on different factors ranging from social interaction to more technical features. For instance, in [209], a speech-oriented humanoid robot called *ASKA* was employed as a receptionist in a university. The robot is able to understand users' questions about the route to travel for reaching a destination and give directions by using voice and arm pointing gestures. Since in this work authors' aim was devoted to develop a robust HRI dialog system in a real-world scenario, no consideration was made regarding *ASKA*'s effectiveness as a direction-giving system. Similarly, in [210] authors created a robotic receptionist system for identifying office workers and learn their names, by focusing on and exploring the structure of human-robot dialogue. The voice was used by the robot for providing information about the office where an employee could be found and giving directions to the office's location. Again, no evaluation was carried out on the effectiveness of the

guidance provided. In [211, 212], the focus was on the effectiveness of arm pointing gestures to identify a given target. Specifically, in [211] a robot equipped with a statistical model was developed to evaluate ambiguity in arm pointing gestures performed by humans when used to indicate places or objects. Authors proved that during the detection phase the mutual distance between different targets as well as the distance between the human user and the indicated target may generate uncertainty. In [212], the use of arm pointing gestures in deictic interactions to indicate a region in space was investigated for identifying the modality that is more effective for carrying out the considered task.

As mentioned above, other studies investigated social phenomena rather than technical aspects. As a matter of example, in [213], a Wizard-of-Oz receptionist robot was used to investigate human-robot interaction with the aim of identifying possible behavioral patterns. Another robot receptionist named *SAYA* and equipped with the ability to nod its head (robot's body cannot move) was developed in [214]. Authors demonstrated that robots exhibiting this human-like feature may improve human-robot interaction in terms of understanding, human-likeness, and familiarity.

Other studies specifically studied direction-giving tasks. For instance, in [215], the *Robovie* robot was used to give directions to human users by combining utterances and gestures. Dialogue was articulated by introducing pauses timed on how humans speak and listen. The obtained results proved that, although users understood the given path by receiving directions through the use of the voice, deictic gestures considerably enriched the utterances. In addition, users preferred the listener's pause-based interaction mode rather than the speaker's pause-based one. Voice and gestures were also used in [216], where the *NAO* robot was employed to give indications to office workers by tracking them through face detection. Experimental results showed that directions received by leveraging this approach might be complicated and hard to remember when used for long paths. Similarly, in [217], voice and deictic gestures as well as pointing the destination on a physical map were used by the *iCub* receptionist robot to provide route information. Like in other works, the considered approach has not been evaluated from quantitative or qualitative points of view.

A Comparison of Direction-giving Solutions Since the panorama of direction-giving solutions is quite heterogeneous, information about the suitability and performance of a particular implementation compared to other solutions is required.

By moving from the works reviewed above, it can be noticed that most of them did not focus on the evaluation of the provided directions from the point of view of effectiveness. Nevertheless, some activities in this direction have been conducted already.

For instance, the authors of [218, 219] performed a comparative analysis between two receptionist systems exhibiting different embodiments, namely, a humanoid robot with a mechanical look named *KOBIANA* and an on-screen virtual ECA named *Ana*, both using vocal directions for providing guidance to human users about the position of two different rooms (red and blue). The authors' purpose was to study how different

voices (robotic vs. human-like) and visual appearances, as well as different embodiments (virtual vs. physical), could affect humans' evaluation of the considered receptionists as social partners. Subjective observations indicated that participants who had received indications by *Ana* judged it a better receptionist than *KOBIANA* because of the human-like voice and appearance. However, objective observations showed that the number of participants who got lost with *Ana*'s directions was much higher than those who received route information from *KOBIANA*. Unfortunately, according to the authors, obtained results were biased by the fixed order adopted by the users to experiment the two receptionist systems: thus, no definitive conclusion on the impact of receptionists' aspect could actually be drawn.

Similarly, in [220], authors compared two different embodiments for a receptionist system from the point of view of effectiveness. However, a more conventional (unembodied) wayfinding means, i.e., a map, was also included in the comparison. More in detail, a humanoid robot (named *KHR2-HV*), a virtual ECA (named *NUMPACK*) and a GPS-based map were considered. The two embodied receptionist systems (with different sizes) were configured to give directions in three different ways: voice only, voice and arm pointing gestures with speaker's perspective, and voice and arm pointing gestures with listener's perspective. The GPS-based system was configured to either play back a voice guide without showing any map or show the map and provide audio directions. Authors' goal was to conduct a study on what they called the "gesture factor" (no gesture/no map vs. listener's perspective/map vs. speaker's perspective/map) and the "agent factor" (ECA vs. robot vs. map) in wayfinding tasks. During the experiments, subjective and objective observations were collected for assessing the effect of the embodiments and their social perception, as well as the effectiveness of navigation aids and their impact on users' performance. Objective measurements consisted of a route drawing task (in which users were requested to draw on the map the route they saw/heard) and a retelling task (in which users had to refer the directions received from the system). Objective results showed no difference in users' performance based on the type of embodiment. However, a strong impact on both perception and performance was measured when using listener's perspective gestures than speaker's perspective and no gestures. Subjective results indicated that the two embodied systems (robot and ECA) positively affected users' evaluation in terms of social perception. Furthermore, the physical robot was preferred and evaluated as more co-present and understandable than the others solutions when listener's perspective gestures were employed. The virtual robot was judged as more enjoyable and familiar than both the map and the physical robot when the no map/no gesture configurations were used, confirming findings obtained in [218, 219].

By summarizing findings from the above studies, it can be observed that physical humanoid robots are judged as better receptionists when used to give directions by means of gestures, whereas virtual embodied receptionist systems are preferred when voice directions are employed. Moreover, it can be noticed that when gestures are employed the listener's perspective lead to better performance compared to the speaker's

one (and voice directions alone).

Despite the relevance of these empirical cues, a comprehensive exploration of the extended design space for receptionist systems has not been conducted yet. For instance, only a few works explored the benefits possibly ensured by the integration of a map, by studying the resulting configurations (virtual or physical embodiment, combination of voice and map, as well as arm pointing gestures, etc.).

Receptionist Systems

By moving from the works discussed in Section 3.2.2, three direction-giving solutions, which differ in the type of the embodiment and interaction interfaces, were considered and developed: a socially interactive physical humanoid robot capable of uttering directions and showing them on a map or leveraging arm pointing gestures for giving directions; an ECA featuring the same social behaviors and interaction interface of the physical robot but exhibiting a virtual embodiment; lastly, the most common approach used today as a direction-giving system, i.e., an interactive audio-map, which represents an unembodied system without social interaction interface.

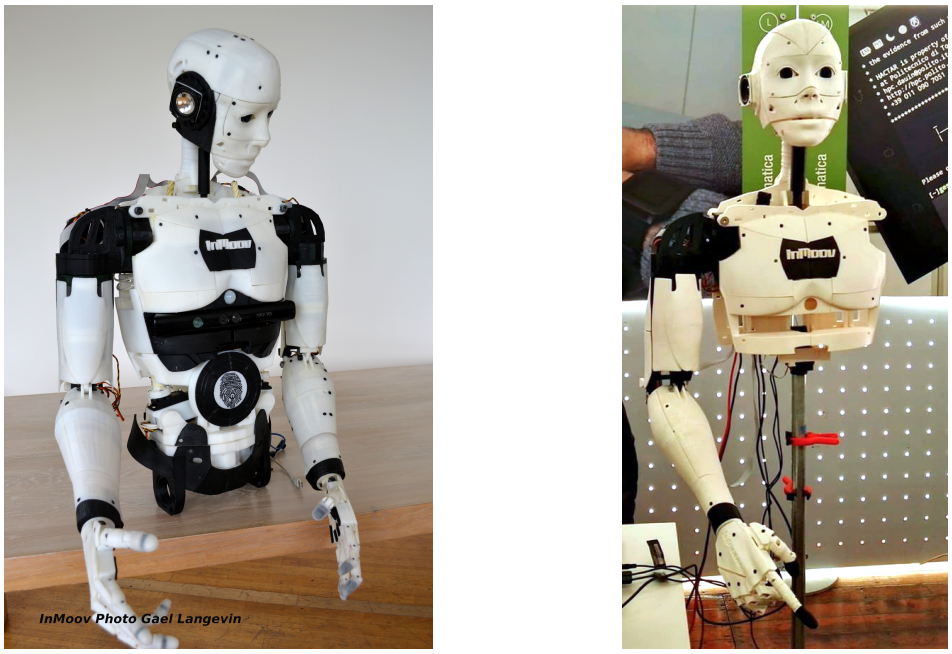
In the following, the embodied (socially interactive) and map-based receptionist systems developed for this study will be introduced, by providing also some implementation details.

Physical Robot This paragraph introduces *InMoov* [221], the robotic platform considered in this application domain to play the role of the physical receptionist, by also providing its hardware and software features.

InMoov is a humanoid robot devised by the French sculptor Gaël Langevin within an open source project initiated in 2012. As illustrated in Figure 3.38a, it is entirely built out of 3D printing ABS (Acrylonitrile Butadiene Styrene) filaments. For the purpose of this study, only the head, the right arm (including the shoulder, biceps, and hand) and the upper torso were printed (Figure 3.38b).

InMoov's assembly consists of 15 servomotors with different torques and speeds distributed on the body with a total of 16 DOFs: 5 DOFs for the arm, 1 DOF for the wrist, 5 DOFs for the fingers and 5 DOFs for the head. It is worth noting that in this study, all the servomotors in the arm were modified in order to let the robot's joints execute unconstrained rotations, which were not permitted in the original design. By focusing on the robot's arm, it consists of a kinematic chain (omoplate, shoulder, bicep, elbow, forearm, and wrist) made up of six revolute joints controlled using inverse kinematics (IK). The central processing unit consists of an Arduino Mega ADK[30] board, which is responsible for collecting data from software modules (discussed later) and sending

[30]https://store.arduino.cc/arduino-mega-adk-rev3



(a) (b)

Figure 3.38: Robotic platform used in this study: (a) original design of the *InMoov* robot, and (b) robotic receptionist implementation.

them to the servomotors. Two cameras were positioned in the robot's eyes for reproducing the vision system, whereas two speakers were positioned in the robot's ears, and attached to an amplifier board, for sound reproduction. An external microphone was used for the auditory system in order to attenuate the impact of noise produced by the servomotors on the perceived audio.

The high-level architecture, which was inherited by the *InMoov* project and revised/adapted to implement the receptionist system considered in this study, is illustrated in Figure 3.39. It is a four-layered framework composed by the following levels: *sensors & actuator*, *control*, *middleware* and *application*. The first layer contains the physical robot's sensors and actuators described in the paragraph above. The *control* layer is made up of eight main modules, which are used to manage the robot's direction-giving functionalities. The *middleware* is the layer responsible for the execution of the aforementioned modules, whereas the *application* layer is devoted to the implementation of the logic for the interaction with human users by making the robot give directions in a natural way. Details of the above layers are provided below.

Concerning the *control* layer, the eight modules are reported in the following.

- *Face tracking*: this module is responsible for elaborating the video stream obtained from the robot's cameras and detecting the presence of human faces and their

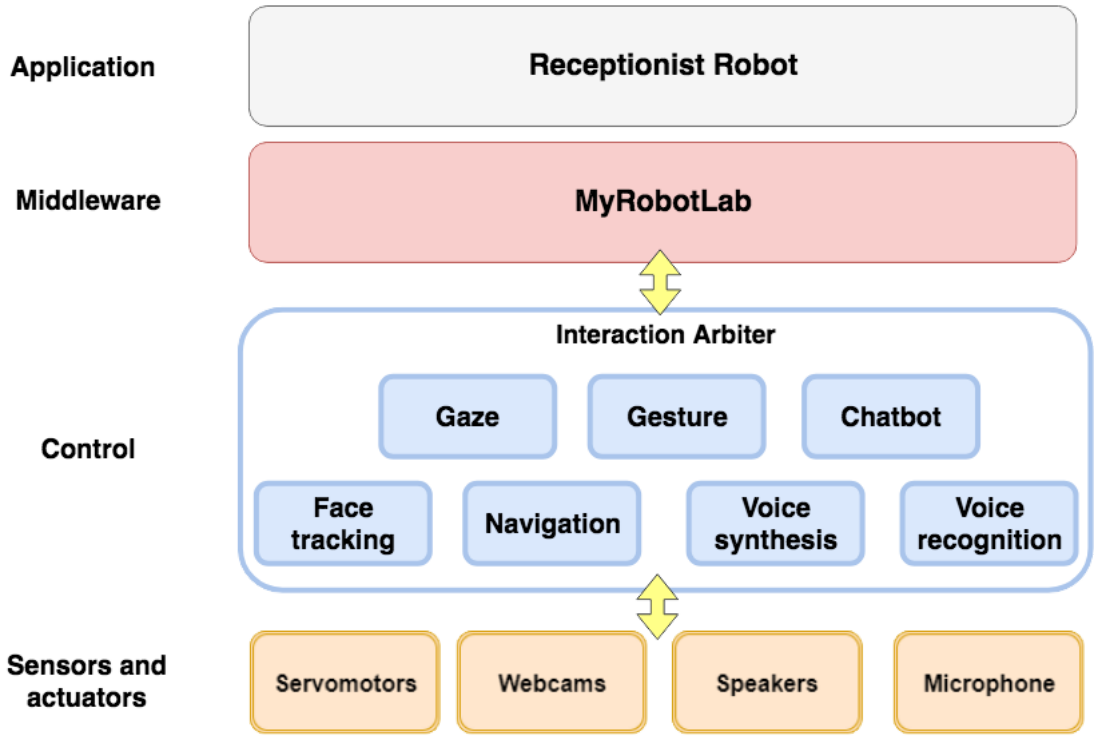


Figure 3.39: High-level architecture of the robotic receptionist system.

position in the robot's FOV. It leverages MyRobotLab[31] Tracking module, which allows to track human faces in real-time via the OpenCV library. In particular, when a face is detected, the robot adjusts the position of its head in order to keep the identified face in the center of its FOV. However, in a reception use case in which the robot is expected to be placed in public and crowded areas, not all the detected people may want to start a conversation. Hence, two events, namely, *foundFace* and *lostFace*, were added to the original tracking module in order to limit the number of unwanted activations. The *foundFace* event is triggered when a human face is detected in a given number of consecutive frames; similarly, the *lostFace* event is fired when no human face is detected in a pre-defined number of frames. Tracking data produced by this module is sent the to *Gaze* module.

- *Gaze*: this module is devoted to manage robot's head and eyes movements during the interaction with a human user. As a matter of example, in the greetings and farewell phases, the robot's head and gaze are directed on the user's face,

[31] http://www.myrobotlab.org

whereas, in one of the considered configurations (i.e., the robot provides directions by pointing destinations on a map) the gaze and the head are oriented towards the map.

- *Chatbot*: this module represents the brain of the system. It is responsible for generating text responses according to received text stimuli. In particular, it relies on the A.L.I.C.E. bot[32], an open source natural language processing chatterbot that allows the customization of a conversation by using a XML schema called AIML (Artificial Intelligence Markup Language). With AIML it is possible to define the phrases/keywords that the robot should capture and understand (associated with greeting/farewell phases, as well as to destinations) and the answers it should provide (greeting/farewell expressions and directions). In addition, when a phrase/keyword related to a destination is spotted, the AIML language can be used to activate actions. In this case, robot's arm gestures for giving directions are triggered, by sending required information to the *Navigation* module. The conversation logic adopted by the robot is illustrated in Figure 3.40: purple clouds represent examples of possible user inputs, whereas grey clouds are examples of possible robot's answers.

- *Voice recognition*: this module gains voice commands from the microphone, converts them to text through the WebKit speech recognition APIs by Google, and sends the result to the *Chatbot* module.

- *Voice synthesis*: this module allows the robot to speak. It receives text messages from the *Chatbot* module, converts them into audio files using the MaryTTS[33] speech synthesis engine, and sends them to the speakers. Moreover, when a message is received, it triggers a *moveMouth* event, which makes the robot's mouth move by synchronizing with words pronounced.

- *Navigation*: this is the main module that was developed in this study and integrated in MyRobotLab for providing users with directions for the requested location. It relies on the IK module available in MyRobotLab, which was adapted for moving the EE of robot's arm and reproduce pointing gestures (in-the-air or on the map). However, the IK module does not guarantee that a runtime-computed solution always exhibits the same sequence of movements for the EE to reach the intended position. Moreover, should the solution fall in a kinematic singularity point, it would cause the robot's arm to lose its ability to move, making it unusable. For these reasons, the module was modified by adding an IK pre-calculation phase, in which the EE is moved to the desired position by means of a controlled

[32] http://www.alicebot.org

[33] http://mary.dfki.de

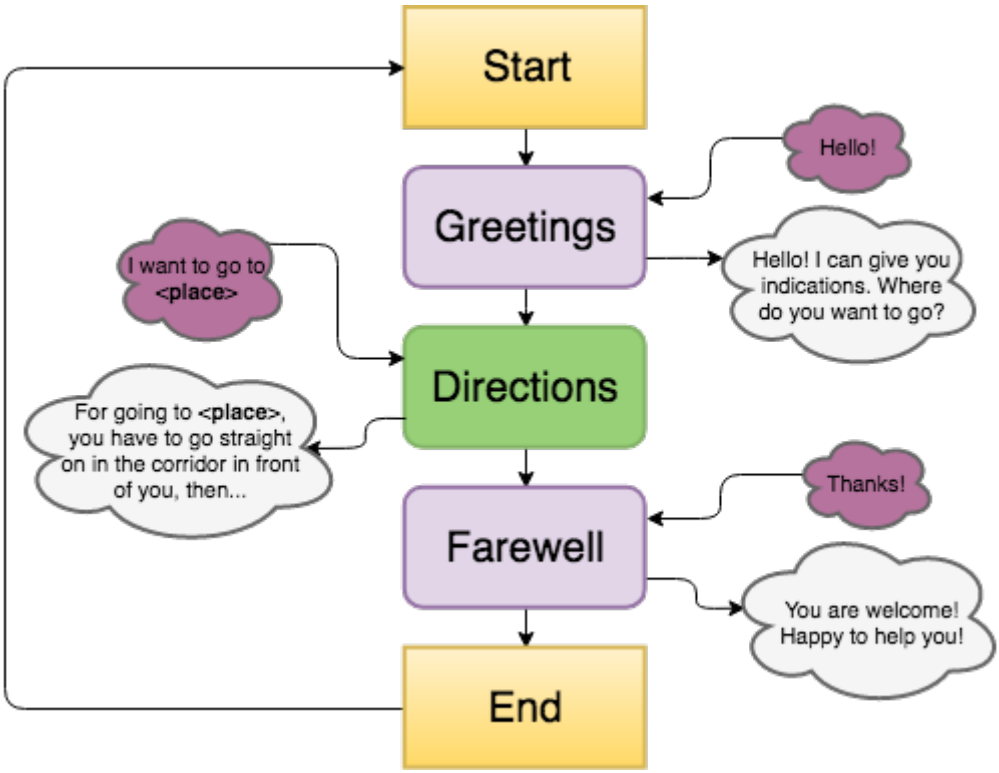


Figure 3.40: Robotic receptionist's conversation logic.

sequence of gestures made up of small displacements along the xyz axes. When the intended position is reached, the gesture sequence is stored in the system and linked to the given destination. In order to make the robot move in a natural way, tracking data of a human arm while executing the required gestures were first recorded via a Microsoft Kinect depth camera and the OpenNI software[34] and then properly adapted to the experimented scenario. When the destination is determined by the *Chatbot* module, the *Navigation* module loads the corresponding sequence and sends gestures to the *Gesture* module, which executes them.

- *Gesture*: this module was developed within the present study. It is devoted to make the robot actuate the gesture sequence suitable for the particular direction-giving modality being considered (in-the-air or map pointing gestures) and the specific destination selected.

- *Interaction Arbiter*: this module "arbitrates" all the previous ones according to the flow of human-robot interaction and the direction-giving modality in use. As a matter of example, while the robot is speaking using the *Voice synthesis* module,

[34] http://openni.ru/

the *Voice recognition* module needs to be stopped to avoid misbehaviors.

The *middleware* layer, which consists of MyRobotLab Java service, orchestrates the execution of the above modules and acts as an intermediary between the robot's functionalities and the *application* layer.

The *application* layer completes the stack by actually implementing the reception logic illustrated in Figure 3.41. In particular, when the system starts, it initializes the MyRobotLab modules and waits for external stimuli for initiating the interaction. Stimuli may be both a detected face or a voice command issued by the human user. In the first case, the receptionist robot starts the interaction with the greetings phrase: "Hello! I can give you indications! Where do you want to go?". Afterwards, the user can continue the interaction as shown in Figure 3.40. If no answer is detected, the robot returns in the waiting phase. In the second case, the user begins the interaction by greetings the robot or asking it about a given place. Interaction continues as illustrated in Figure 3.40.

Virtual Robot The virtual robot used in this study for acting as a virtual receptionist consists of an open source 3D model of the physical *InMoov* robot displayed on a 27-inch monitor with 1920×1080 resolution.

The 3D model was selected to exhibit the same appearance of the physical robot (rather than a virtual human-like character) in order to limit the presence of biases in the users' evaluation due to the receptionist's appearance. Notwithstanding, it is worth saying that the use of a (2D) screen to display the 3D model could introduce differences in the way users perceive the virtual and the physical robots (e.g., in terms of size, tridimensionality, etc.). Hence, future work should be aimed to further dig into the above aspects.

As illustrated in Figure 3.42a, the 3D model, later referred to as *VinMoov*, consists of the following parts: head, upper torso, right arm (omoplate, shoulder, bicep, elbow, forearm, and wrist), and hand. Similarly to the physical robot, where the movements of the various body parts are controlled by servomotors, *VinMoov* is endowed with a virtual skeleton, or "rig", made up of bones and joints which are articulated for making the 3D model assume intended poses (Fig. 3.42b). All the DOFs in the physical robot were maintained.

A webcam was placed on top of the workstation screen hosting *VinMoov*, in order to replicate the vision system of the physical robot. In the same way, an external microphone was used for implementing the auditory system as well as external speakers were placed on the sides of the screen for reproducing sound.

Like for the *InMoov* robot, MyRobotLab provides a so-called *Virtual InMoov* service to manage the operation of the virtual robot in two modalities: user and developer. The former allows users working with the exact features available in *InMoov*, whereas the latter allows them to develop or in particular test new functionalities without plugging the physical robot to the workstation. However, with the physical robot not connected

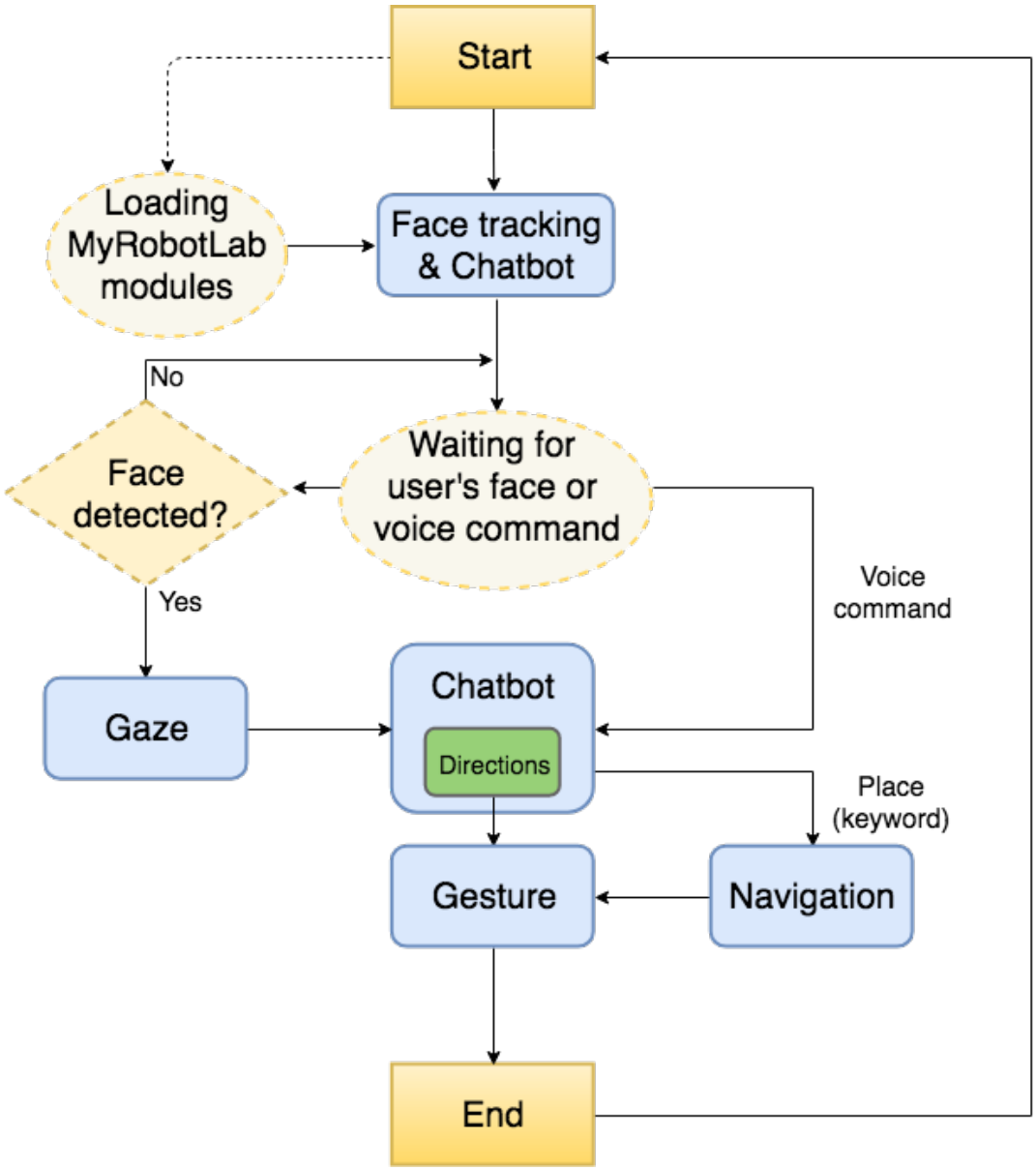
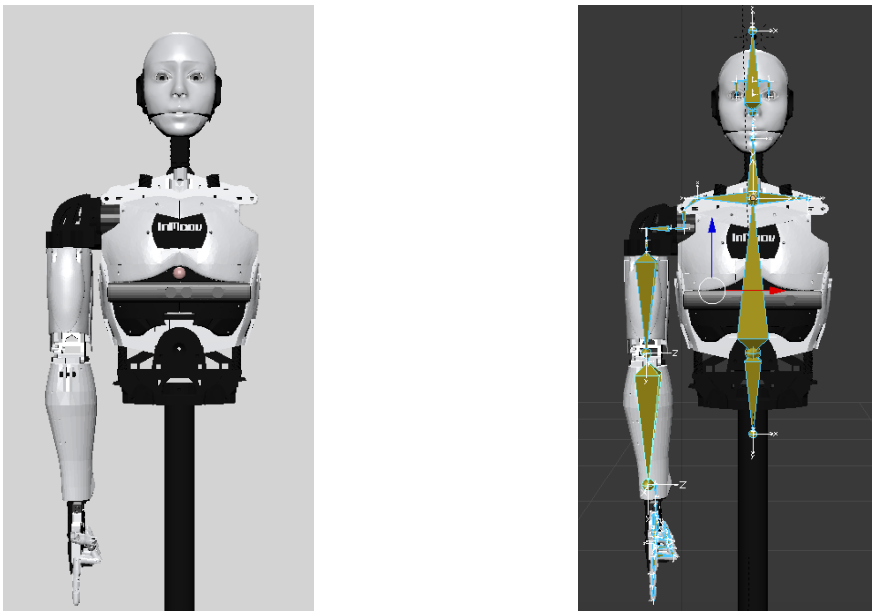


Figure 3.41: Robotic receptionist's application logic.

to the workstation, the timing of the virtual replica's movements are not recreated in a proper way. This is due to the fact that, in the physical robot, the timing of movements is dictated by the real presence, type and physical features of the servomotors, which are only virtually recreated in the above service. Furthermore, although the user mode reached a stable version, the developer mode (the appropriate mode for the integration of the *Navigation* module developed in this study) is currently under development, and interfaces are changing rapidly.

Thus, in this study, it was decided to discard the MyRobotLab for the virtual robot



(a) (b)

Figure 3.42: *VinMoov*: (a) aspect; and (b) rig used for controlling deformations (poses).

and to develop a distinct software for implementing the *VinMoov*-based reception service. Software architecture, developed according to the client-server paradigm, is illustrated in Figure 3.43. The client side hosts a .NET application implementing the modules for managing the interaction: *Face tracking*, *Chatbot*, and *Voice recognition*. The server side consists of the BGE, which is responsible for rendering the 3D model and the execution of the modules providing the interaction feedback: *Gaze*, *Voice synthesis*, *Gesture*, and *Navigation*. The choice to implement the modules on the client or the server side was based on the availability of third-party libraries to ground developments onto.

Like in the physical robot, the *Face tracking* module was implemented through the OpenCV image processing library by applying the same filters and handling the same events. Once the 2D position of the interacting user's face is detected in the robot's FOV, tracking data are sent to the *Gaze* module in order to let the *VinMoov*'s head/eyes follow it.

The *Gaze* module maps these data (two coordinates) on a 3D object (not visible during the interaction) placed in front of the virtual robot (third coordinate is fixed on the 3D world), which mimics the position of the face in the BGE reference system. A *track-to* constraint is applied to the *VinMoov*'s head bone to emulate the face tracking of the physical robot. The difference in the position of the camera(s) between the physical and the virtual robots did not affect the accuracy of the head's movements.

The *Chatbot* module controls and customizes the interaction with the users by defining the inputs to be spotted and sending them to the *Voice synthesis* and *Navigation*

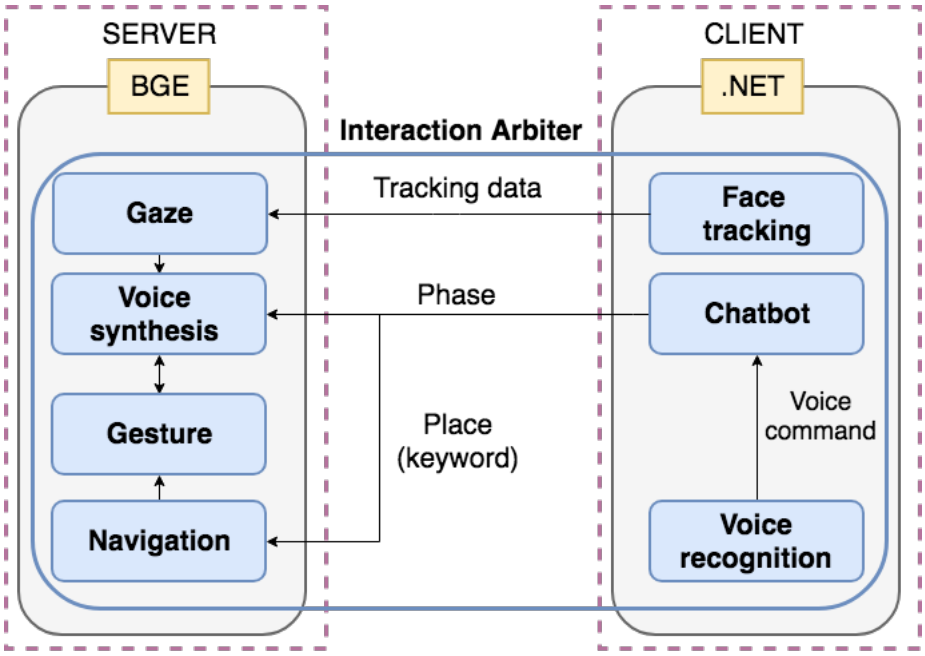


Figure 3.43: High-level architecture of the *VinMoov* receptionist system's control logic.

modules, like in the physical robot. The module was implemented through the Microsoft Grammar Builder service[35], which was used to specify the dictionary and the phrases/keywords to be detected. The *Voice recognition* module was built by leveraging the Microsoft Speech to Text API[36], which converts voice inputs received by the users to text to be sent to the *Chatbot* module. The *Voice synthesis* module receives detected keywords (greetings/farewell expressions and destinations) from the *Chatbot* module and makes *VinMoov* speak by playing audio files resulting from the conversion of text answers (greetings/farewell expressions and directions) into audio answers through the Microsoft Text to Speech API[37]. When a sound is to be reproduced, the robot's face movements synchronization is triggered, like in the physical robot.

The *Navigation* module is responsible for identifying the appropriate gestures from a pre-defined list. Like in the physical robot, gestures were recorded in Blender by leveraging the keyframing technique, in which key poses of the bones and joints in the *VinMoov*'s rig can be defined on a timeline. It is worth observing that, in one of the considered configurations (in which the virtual robot gives directions on a map), the map was slightly tilted towards the user in order to make it visible to him or her (the map positioned on a plane orthogonal to the physical robot could not be seen on a

[35] http://msdn.microsoft.com/en-us/library/microsoft.speech.recognition.grammar.aspx

[36] https://docs.microsoft.com/en-us/azure/cognitive-services/speech/home

[37] https://docs.microsoft.com/en-us/azure/cognitive-services/speech/home

screen). For this reason, the same IK solutions determined in the physical robot could not be directly applied to its virtual replica. Thus, with the aim to guarantee that the *VinMoov*'s gestures were as much similar as possible to the physical robot's sequence both in terms of bones' and joints' position as well as of timing, a functionality of Blender known as *video reference* was used. This technique leverages a set of videos recorded while the physical robot is giving directions by projecting them on a semi-transparent plane (adjustable in size and position) placed on the *VinMoov*'s 3D model background and referencing them for the *VinMoov* animation stage (basically leveraging the same idea of tracing paper to copy a drawing). Once the *Navigation* module receives the keywords spotted by the *Chatbot* module, it loads the appropriate sequence of keyframes (an *action*, in the Blender's terminology, representing a gesture) and the *Gesture* module selects the correct direction-giving modality according to the current configuration.

Lastly, Figure 3.44 depicts the *Interaction Arbiter* module, which was implemented by using the drag & drop-based *Logic Editor* integrated in Blender. Specifically, the yellow blocks on the left side represent the Blender's actions recorded in the *Navigation* module. The bottom part shows how the outputs of modules on the server side are connected to the inputs of modules on the client side through the scripts shown to the right.

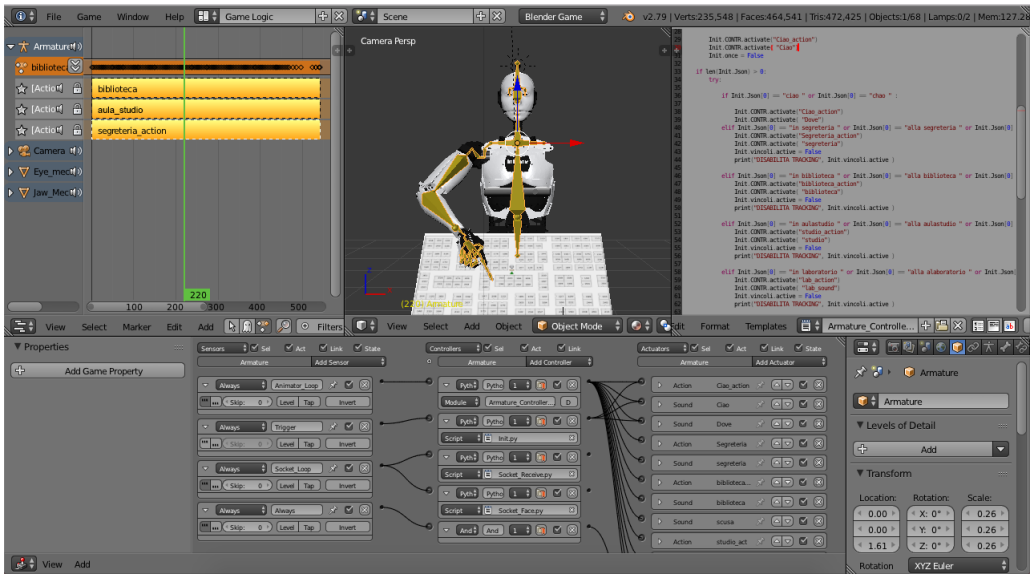


Figure 3.44: Implementation of the *VinMoov*'s interaction arbiter in Blender.

Interactive Map A further receptionist system that was considered in this study is a 2D interactive audio-map, the appearance of which is illustrated in Figure 3.45.

The map was displayed on the same monitor used for the virtual robot (though with a horizontal orientation, in this case). It relies on a minimalistic and generalized visualization based on the architectural blueprint technique discussed in Section 3.2.2.

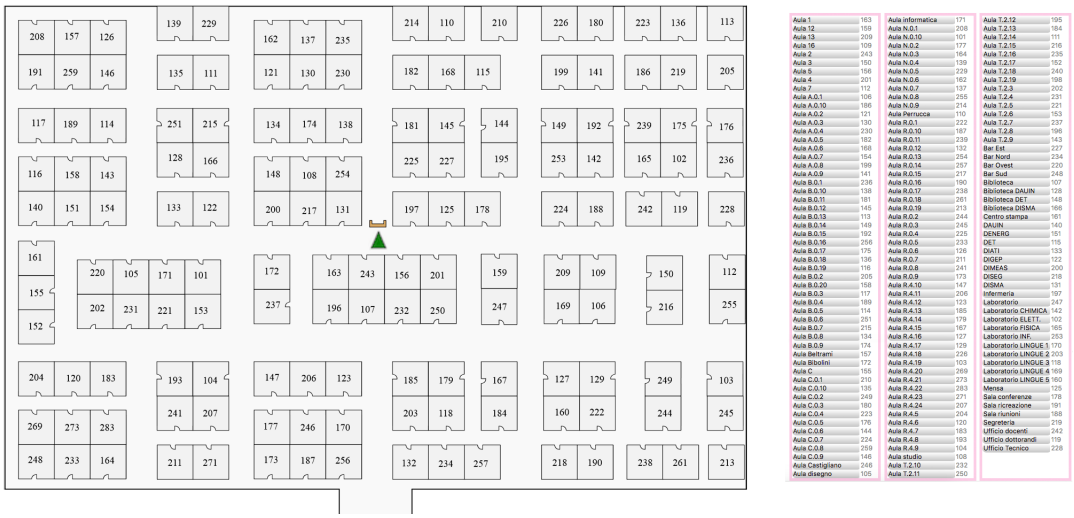


Figure 3.45: Interactive map-based receptionist system.

Concerning user's location and orientation, the map uses a YAH representation and a forward-up orientation. In particular, as illustrated in Figure 3.45, a green triangle is used to indicate both the user's location and orientation with respect to the environment, whereas a light brown rectangle depicts the reception desk where the map system is installed onto. The system is able to refer to a number of destinations (rooms), which are listed in alphabetical order to the right of the map. Should the user select a destination by either clicking the name in the list or the corresponding box on the map, an animated colored path would display on the map from the YAH mark to the requested destination.

By moving from the considerations emerging from the review in Section 3.2.2, it was decided to accompany the visual representation of the path with the same voice directions uttered by the two receptionist robots. This choice was made with the aim to both limit differences among receptionist systems considered in this study to their actual peculiarities as well as to avoid biases as much as possible. The audio files of directions were developed through the Artyom JavaScript-based voice synthesis library[38]. Voice input was not considered.

The interactive map was implemented as a Web application by leveraging the wayfinding jQuery plugin[39], which allows for the creation of interactive SVG (Scalable Vector Graphics) maps. It supports the computation of the shortest path to a target location on the map based on information encoded in the SVG file and its visualization

[38] https://sdkcarlos.github.io/sites/artyom.html

[39] https://github.com/ucdavis/wayfinding

on the map. Information is stored in layers, which refer to rooms (defining clickable areas on the map), paths (line segments defining possible routes), and doors (end-points associating room names to paths' ends).

Hypotheses and Experimental Results

By moving from the works discussed in Section 3.2.2 and in particular, by leveraging findings already obtained from the comparison of different direction-giving systems, three hypotheses were formulated.

The first hypothesis (reported below) originates from the assumption that the introduction of a map on which a physically embodied receptionist can trace the path to follow while uttering directions might improve users' correct understanding of the indications received and their ability to reach the place of interest compared to the best approach proposed in the literature, i.e., physical humanoid robot equipped with (listener's perspective) arm pointing gestures. Figure 3.46 depicts the aforementioned configurations which will be later referred to as *PRM* (Physical Robot with Map) and *PRG* (Physical Robot with Gestures). For the sake of completeness, a virtual embodiment for the latter configuration (later referred to as *VRG*) was also considered in the study.

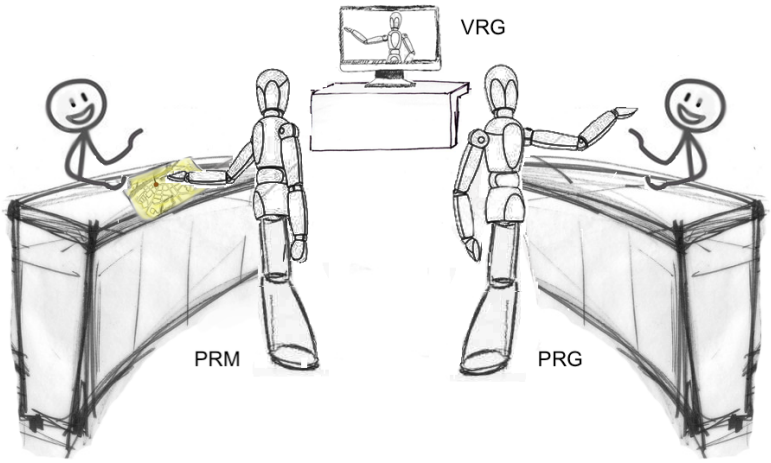


Figure 3.46: Configurations considered in the first user study to evaluate the suitability and effectiveness of direction-giving systems developed so far: *PRM* (Physical Robot with Map), *PRG* (Physical Robot with Gestures), and—for the sake of completeness—*VRG* (Virtual Robot with Gestures).

Hypothesis 1. *The integration of a map in a physically embodied receptionist system and the possibility for the receptionist to trace the route on it would be more effective in direction-giving tasks compared to the use of arm pointing gestures.*

Assuming that the use of a receptionist system showing the path to follow on a map can increase performance and likeability, this study also aims to examine the role played both by different embodiments (virtual, physical, unembodied) and social behaviors (with and without) in the considered domain. To this aim, like in [220], further configurations were considered and compared in this study, namely, the aforementioned PRM system, a Virtual Robot with Map or later abbreviated as *VRM* (i.e., the virtual version of *PRM*) and an Interactive-audio Map (later abbreviated *IM*), as illustrated in Figure 3.47. The second hypothesis is given below.

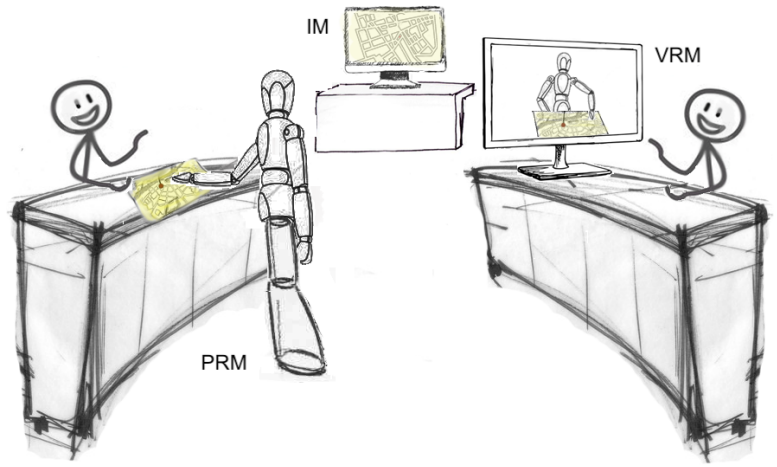


Figure 3.47: Map-based configurations considered in the second user study to assess the role of embodiment and social interaction behaviors in direction-giving systems: *PRM*, *VRM* (Virtual Robot with Map), and *IM* (Interactive-audio Map).

Hypothesis 2. *Embodied receptionist systems (both physical and virtual) endowed with social behaviors (face tracking, gaze) would be evaluated in a better way and would lead to better performance in wayfinding tasks compared to a map-only (unembodied) solution without social behaviors.*

Should the latter hypothesis be verified, one could expect to observe measurable differences in users' judgments between virtual and physical receptionist robots. Thus, the third hypothesis was formulated to dig into these possible differences, by studying the impact on human users of the physical presence (robot co-located with the user) with respect to the virtual presence (robot displayed on a screen). For this reason, the two embodied receptionist systems considered in this domain were designed to share the same (robotic) visual appearance, though in one case the robot was displayed on a screen (Figure 3.47). The third hypothesis reads as reported below.

Hypothesis 3. *The physical presence through a co-located robot would have a higher measurable impact on participants' performance, as well as on their perception of social*

interactions, in wayfinding tasks compared to a virtual robot displayed on a screen [222, 223].

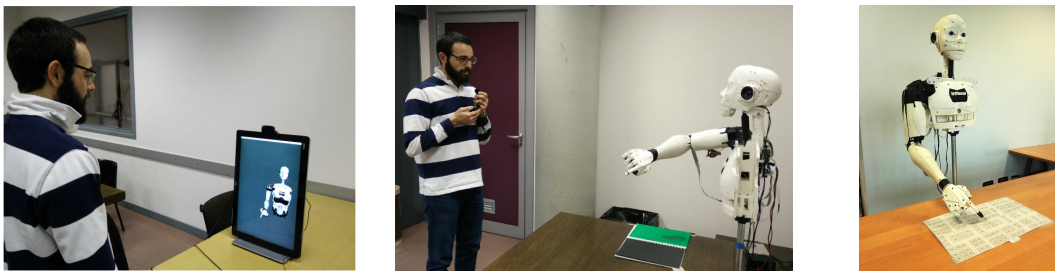
As discussed above, in this study three hypotheses were formulated, which were evaluated by means of two user studies. The first study was aimed at determining whether the introduction of a map in a physically embodied robotic receptionist system could improve users' wayfinding performance and understanding of the received directions compared to an accepted direction-giving system found in the literature leveraging arm pointing gestures (Hypothesis 1). By building on results of the first study, the second study purported, firstly, to assess whether social behaviors and embodiment can improve users' acceptance of a receptionist system and their performance compared to no social behavior/no embodiment (Hypothesis 2) and, secondly, to investigate the added value of physical presence compared to a virtual representation (Hypothesis 3).

First User Study: Arm Pointing Gestures or Route Tracing on a Map?

The user study presented in this section was carried out to test the Hypothesis 1. Before digging into detail, it can be recalled from previous works discussed in Section 3.2.2 that, the listener's perspective is considered the best option for embodied receptionist systems using arm pointing gestures. Moreover, physical robots are (to be) preferred to the virtual ones, when these gestures are employed. Based on the above considerations, a comparative analysis was performed in this study involving a physical robot featuring a paper map on which to trace the route while uttering directions (*PRM*) and a receptionist system featuring the same embodiment but using listener's perspective arm pointing gestures (*PRG*). As mentioned previously, for the sake of completeness, a virtual version for the *PRG* configuration was also considered (*VRG*).

Participants involved in the study (11 males and 4 females) aged between 21 and 29 years (M = 25.57 SD = 2.40), were recruited among Italian-speaking students from Politecnico di Torino. At the beginning of the experiment, participants were informed that during the test, they would have to interact with the three receptionist systems illustrated in Figure 3.48a–c (corresponding to configurations sketched in Figure 3.46) to ask for directions of three different rooms—namely, secretary, library, and study room. As illustrated in Figure 3.49, a map depicting a fictional university environment was designed for soliciting students' wayfinding abilities in an unknown site.

Participants were also told that, at the end of the interaction with each receptionist system, they would have to recall directions received for performing a spatial navigation task. For this purpose, a virtual environment was modeled and imported on a 3D simulator to let participants explore it, thus giving them the impression to navigate a realistic, though simplified, university site. Moreover, in order to create a strong connection between participants and the experience just completed, at the beginning of each simulation, a reception desk depicting the receptionist system just experimented was placed in front of them. The 3D representation of the map used in the simulator is depicted in Figure 3.50.



(a) (b) (c)

Figure 3.48: First user study: (a) *VRG*; (b) *PRG*; and (c) *PRM* systems experimented.

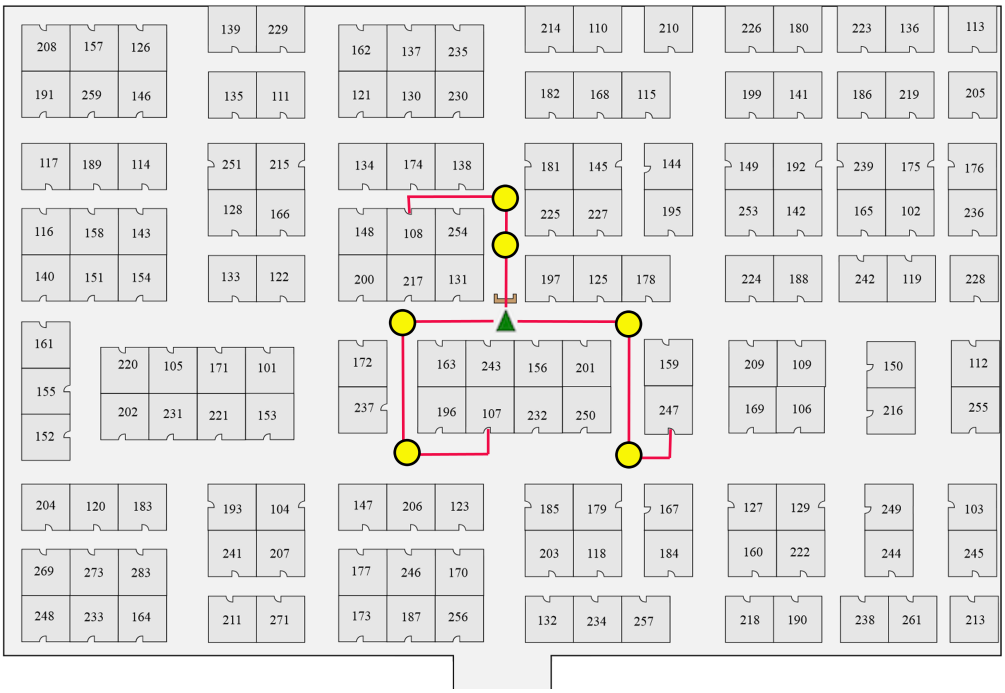


Figure 3.49: First user study: configuration of the fictional environment considered to give directions.

It is worth noting that, in this study, a virtual navigation-based evaluation strategy was selected among others (e.g., compared to strategies requiring the users to draw the route on a map, like in [220]) mainly for two purposes. Firstly, to cope with possible biases that could be introduced by route drawing tasks during the evaluation phase. In fact, differently than users who see the robot showing the path on the map first, users who get gesture-based directions first would not see the map before being required to draw it, with clear consequences on evaluation fairness. Secondly, as discussed in [174],

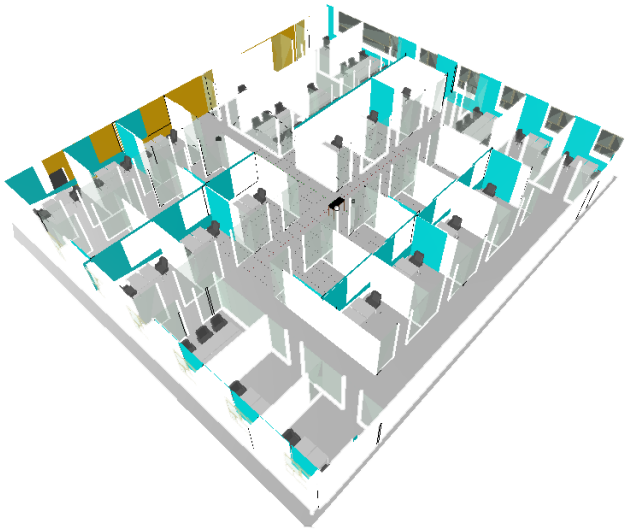


Figure 3.50: First user study: 3D reconstruction of the fictional university used during simulated navigation.

a virtual navigation-based approach represents a more demanding task than path drawing one; hence, it should make differences between the various receptionist systems emerge in a clearer way.

To compensate for possible learning effects, a random order was used to select the sequence of both receptionist systems and rooms (one per receptionist system). The locations of the rooms and the routes to reach them were specifically designed to exhibit the same difficulty and to guarantee that possible differences in participants' performance were only due to the particular receptionist system experimented. As illustrated in Figure 3.49, the route to each room always consisted of two consecutive crossing points (yellow circles in the figure).

During the experiment, time required by participants to perform the room-wayfinding task with each receptionist system, as well as the correctness in reaching the intended room (binary score, i.e., room correctly identified or not) were measured. The temporal organization of a session (all receptionist systems tested) is reported in Figure 3.51.

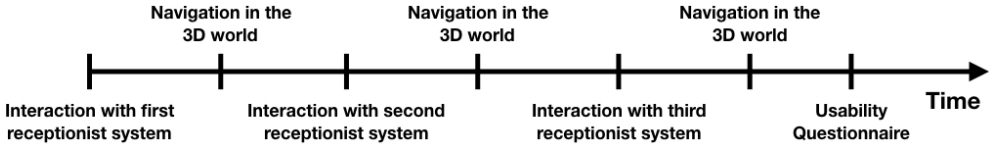


Figure 3.51: First user study: temporal organization of the experiments.

As illustrated in Figure 3.51, at the end of the session, participants were invited to compile a usability questionnaire by expressing their agreement with 21 statements on

a 5-point Likert scale. Sentences used in this study were grouped in four categories, each of which refers to a specific evaluated aspect, as illustrated in Table 3.10.

Table 3.10: First user study: questions/statements in the questionnaire used for the subjective evaluation.

Evaluated Aspect	Question/Statement
Satisfaction	
Q1	The system is pleasant to use
Q2	The system works the way I want it to work
Q3	The system is fun to use
Q4	I am satisfied with the system
User Interaction	
Q5	I was able to understand how to interact with the system
Q6	I think that the way to interact with the system is simple and uncomplicated
Q7	I was impressed with the way I could interact with the system
Q8	I had the right level of control on the system
Q9	I was always aware of what the system was doing
Q10	I felt disorientated while using the system
Q11	System behavior was self-explanatory
Q12	I thought the user interface negatively influenced my performance
Q13	The system response time did negatively affect my performance
Q14	The system appeared to freeze or pause at intervals
Visual Feedback	
Q15	I think the aspect of the receptionist was visually pleasant
Q16	I thought that the aspect of the receptionist negatively influenced my performance
Q17	The aspect of the receptionist reduced my sense of being connected with it
Receptionist Role	
Q18	Good as a receptionist
Q19	Give useful indications
Q20	How much did you enjoy receiving guidance from the system?
Q21	How much would you like to receive guidance from the system in the future?

The first aspect referred to participants' *satisfaction* in using the systems, which was evaluated through the questions Q1–Q4 extracted (and adapted) from the USE questionnaire [82] (only questions on satisfaction were considered). The second aspect was related to *users' interaction* with the receptionist system, which was analyzed by leveraging questions Q5–Q14 derived (and adapted) from the Virtual Reality Usability (VRUSE) questionnaire [224]. The third aspect concerned the *visual feedback* of the receptionist systems. Questions used to this aim (Q15–Q17) were extracted from a modified version of the VRUSE questionnaire [224]. The fourth aspect concerned participants' perception of each configuration considered in the study in the *role* of *receptionist*. Questions Q18–Q21 were used, like in [218, 219, 225]. As in previous studies, scores for questions related to negative aspects, (in this case Q10, Q12, Q13, Q14, Q16, and Q17) were reversed so that higher scores reflect positive opinions.

Collected data were analyzed first through one-way repeated measures ANOVA test (significance level of 0.05) and a two-tailed paired t-test (significance level of 0.05) in post-hoc analysis, in order to check for possible differences in participants' subjective evaluation and objective performance among the three receptionist systems.

Objective evaluations in terms of time required to complete the task as well as the percentage of participants able to successfully identify the intended room are illustrated

in Figure 3.52 and 3.53, respectively.

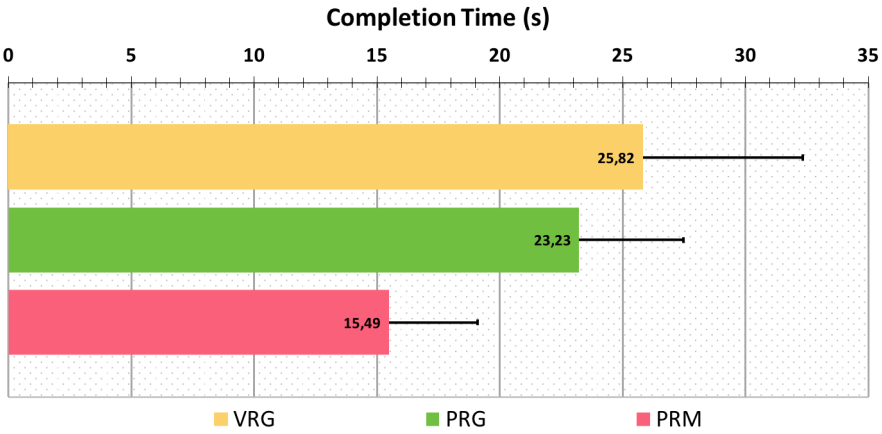


Figure 3.52: First user study: objective results in terms of time required to complete the room-wayfinding tasks. Bar lengths report average values (lower is better), whereas whiskers report standard deviation.

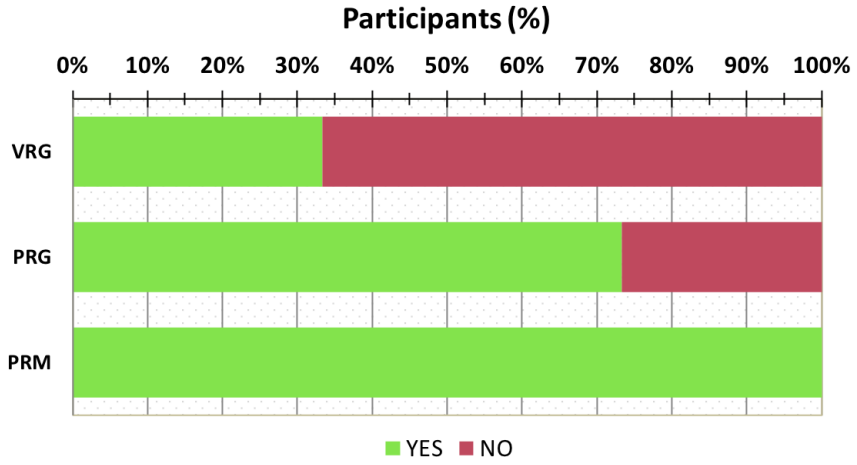


Figure 3.53: First user study: objective results in terms of percentage of participants able to correctly identify the room.

At first sight, it appears from Figure 3.52 that, participants who experimented the configuration featuring the map (*PRM*) took less time to reach the room compared to those who experimented configurations giving directions with arm pointing gestures i.e., *PRG* and *VRG* systems (ANOVA: $p = 1.72 \times 10^{-6}$). Post-hoc analysis revealed no statistically significant differences between *PRG* and *VRG* ($t[14] = 1.89$, $p = 0.0799$), confirming findings reported in [74]. However, differences between *PRM* and both *PRG* ($t[14] = 6.12$, $p = 2.66 \times 10^{-5}$) and *VRG* ($t[14] = 4.96$, $p = 2.11 \times 10^{-4}$), were found to be significant.

Concerning the success rate in finding the correct location of the room, it is clearly evident from Figure 3.53 that, the introduction of a map on which the *PRM* receptionist could trace the route, allowed 100% of the participants to successfully find the intended room. However, only the 73% and 33% of participants were able to find the correct room by getting directions from receptionist systems using arm pointing gestures (*PRG* and *VRG* systems, respectively).

Subjective observations in terms of participants' evaluations of the three receptionist systems gathered through the designed questionnaire are illustrated in Figures 3.54 (+ symbols report ANOVA tests results, i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$) and 3.55.

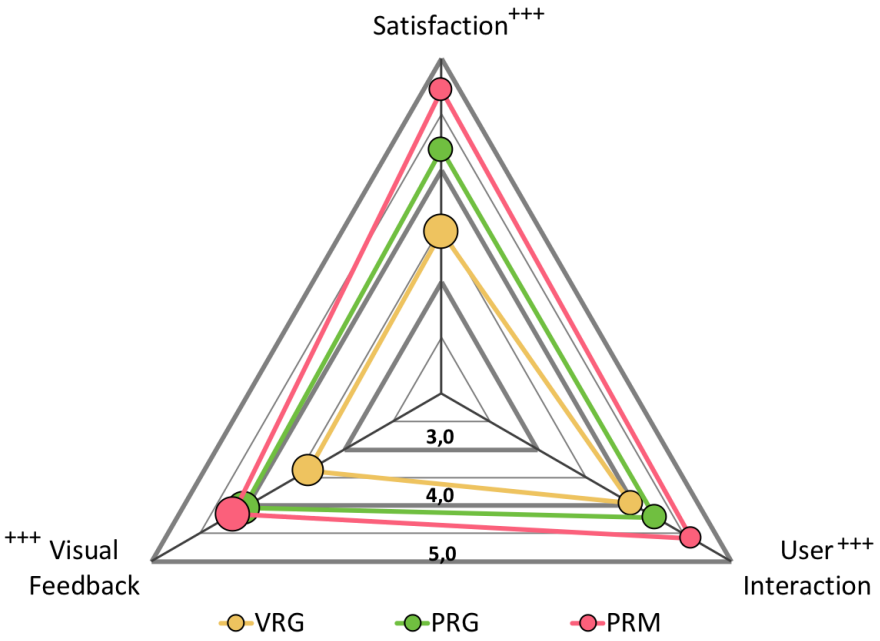


Figure 3.54: First user study: subjective results in terms of usability, circle positions report average values (higher is better), circle dimensions report standard deviation, whereas + symbols report statistical significance determined with ANOVA tests (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

From Figure 3.54, it can be immediately seen that, the *PRM* configuration performed better than the *PRG* and *VRG* ones, for all the usability factors. In particular, concerning the *satisfaction* aspect, the *PRM* system was judged by participants as the most satisfactory receptionist solution in terms of pleasantness (Q1), operation expectations (Q2), fun (Q3), and satisfaction (Q4). By focusing on the comparison between the two configurations leveraging arm pointing gestures, the physical version was judged more positively compared to the virtual one. Post-hoc analysis, reported in Table 3.11, confirmed this evidence by showing that all the questions in this category were found to be statistically significant.

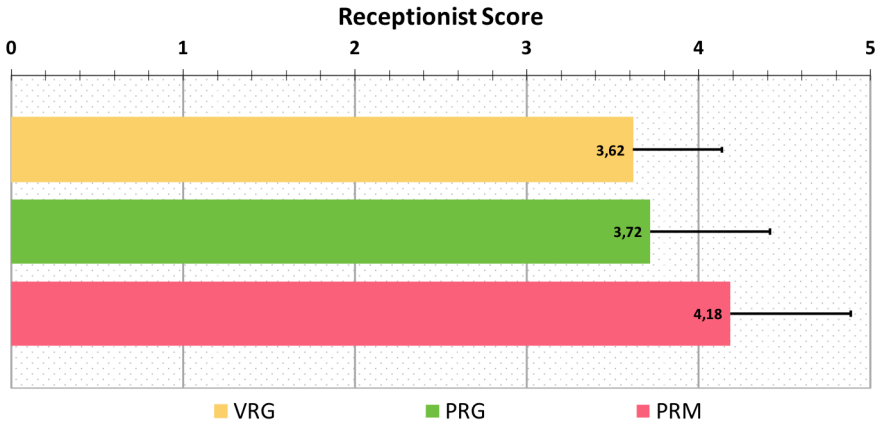


Figure 3.55: First user study: subjective results in terms of suitability of the systems in the receptionist role. Bar lengths report average values (higher is better), whereas whiskers report standard deviation.

Considering the *user interaction* aspect and focusing on statistically significant questions, participants were impressed by how simply they could interact with the *PRM* system compared to the *PRG* and *VRG* ones (Q6 and Q7), by also judging *PRM* system's behavior more self-explanatory (Q11) than that of the other solutions. In fact, they felt disoriented when arm pointing gestures were employed (Q10), with worse performance (Q12). However, according to post-hoc analysis (Table 3.11), the difference between the virtual and physical receptionist systems giving directions with arm pointing gestures (*VRG* vs. *PRG*) was less marked compared to the differences between map-based and gesture-based receptionist systems.

Participants' evaluations in terms of receptionist systems' *visual feedback* showed that configurations exhibiting physical embodiments (*PRG* and *PRM*) were more appreciated compared to the virtual configuration displayed on a screen (*VRG*). In particular, the physical robots were evaluated as more visually pleasing than the virtual one (Q15), which was considered to be responsible for worse performance (Q16) and less engaging (Q17). These observations were also confirmed by the post-hoc analysis reported in Table 3.11, where the differences between the *VRG* and both *PRG* and *PRM* configurations were found to be statistically significant, differently than the difference between *PRM* and *PRG* systems. All the questions belonging to this category proved to be statistically significant.

Concerning the evaluation of the last aspect in the questionnaire, i.e., the suitability of the considered systems in the *role* of the *receptionist*, Figure 3.55 shows that the *PRM* configuration achieved higher scores than the other two systems (ANOVA: p = 8.53 $\times 10^{-5}$). Post-hoc analysis (Table 3.11) revealed that, differences between the *PRM* and both the *PRG* and *VRG* configurations reached statistical significance, whereas no significant difference was found between the *PRG* and *VRG* configurations (all questions

belonging to this category were statistically significant).

Table 3.11: First user study: post-hoc analysis on subjective results and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

Evaluated Aspect	VRG vs. PRG	PRG vs. PRM	VRG vs. PRM
Usability			
Satisfaction	t[14] = -3.33, $p = 4.96 \times 10^{-3}$ (++)	t[14] = -5.49, $p = 7.99 \times 10^{-5}$ (+++)	t[14] = -4.64, $p = 3.86 \times 10^{-4}$ (+++)
User Interaction	t[14] = -2.31, $p = 3.64 \times 10^{-2}$ (+)	t[14] = -5.68, $p = 5.64 \times 10^{-5}$ (+++)	t[14] = -4.87, $p = 2.48 \times 10^{-4}$ (+++)
Visual Feedback	t[14] = -4.12, $p = 1.04 \times 10^{-3}$ (++)	t[14] = -2.09, $p = 0.0552$	t[14] = -4.12, $p = 1.04 \times 10^{-3}$ (++)
Receptionist Role	t[14] = -0.81, $p = 0.4332$	t[14] = -5.33, $p = 1.06 \times 10^{-4}$ (+++)	t[14] = -4.21, $p = 8.77 \times 10^{-4}$ (+++)

By moving from results obtained above, it can be stated that Hypothesis 1 was confirmed and supported both in terms of subjective and objective observations.

Second User Study: The Role of Embodiment and Social Behavior

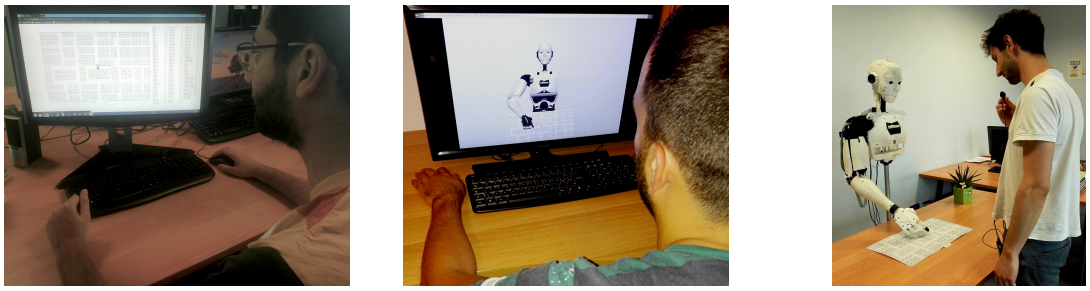
The study presented in this section was conducted to test Hypotheses 2 and 3. To this end, findings obtained from the study performed in [220], and discussed in Section 3.2.2, need to be recalled. In particular, the aforementioned results showed no difference in the effectiveness of direction-giving solutions between voice-enabled map-based systems and virtual/physical robots using arm pointing gestures. By considering only the type of embodiment, authors deduced that this result was probably due to the fact that, in the scenario considered, the embodied systems were employed to provide users with directions towards a destination which was not visible to participants. Thus, differences between virtuality and physicality did not come into play. Rather, considering the results obtained in the study discussed in Section 3.2.2, statistically significant preference was found for a robotic receptionist system showing the route to follow on a map compared to a robot-based direction-giving solution leveraging arm pointing gestures. The application scenario experimented in this case, is comparable to that tackled in [220], as a not visible, even non-existing (fictional), destination was considered. The difference, here, was in the direction-giving system used rather than in the type of embodiment adopted.

Based on these results, the second user study was designed to investigate whether different embodiments (i.e., virtual or physical) of a socially interactive receptionist robot may affect users' performance and perception in wayfinding tasks compared to a map-only (unembodied) solution without social behaviors (Hypothesis 2). Moreover, like in [220], the type of embodiment was also considered in order to investigate possible changes in participants' performance or preference when virtually vs. physically embodied systems are used in wayfinding tasks (Hypothesis 3). Hence, the receptionist systems (all featuring a map) experimented in this study are the *PRM*, the *VRM*, and the *IM* ones.

Participants involved in the study (11 males and 7 females), aged between 21 and 26

years (M = 23.17, SD = 1.42), were selected among university students from Politecnico di Torino by avoiding overlaps with the group of subjects recruited in the first study.

At the beginning of the experiment, participants were provided with the same instructions given in the first study, and were asked to interact with all the three receptionist systems illustrated in Figure 3.56a–c (corresponding to configurations depicted in Figure 3.47). As in the first study, a random order was used to choose the sequence of the receptionist systems and rooms.



(a) (b) (c)

Figure 3.56: Second user study: (a) *IM*; (b) *VRM*; and (c) *PRM* systems experimented.

During the experiment, like in the first study, time needed to complete the room-wayfinding task (for each receptionist system) in the virtual environment as well as the percentage of participants who successfully identify the intended room, were recorded. After having tested all the three receptionist systems, participants were invited to fill in a questionnaire including the same statements used in the first study plus new statements aimed to explore possible differences between the receptions systems' embodiments in greater detail.

The questionnaire was split in four parts. The first part was meant to investigate receptionist systems' usability by means of a number of factors evaluated on a 5-point Likert scale. First, *ease of use*, *ease of learning*, and participants' *satisfaction* in using the receptionist systems were examined. In particular, for the first two factors, questions were extracted (and adapted) from the USE questionnaire [82], as listed in Table 3.12, whereas for users' *satisfaction* factor, questions are the same of those exploited in the first study (Table 3.10). Afterwards, the NAU methodology [79] were considered, by requesting participants to assess the various configurations through four out of five usability factors, namely, *efficiency*, *learnability*, *errors*, and *memorability*. Lastly, a revised version of the VRUSE questionnaire [224] was used to assess participants' experience by means of five factors, i.e., *visual feedback*, *user interaction*, *functionality*, *consistency*, and *engagement*. Specifically, regarding visual feedback and user interaction factors, the same questions already used in the first study were employed in this case (Table 3.10). For the functionality factor (i.e., control that participants had over the systems), questions Q7–Q10 in Table 3.12 were used. Consistency factor was evaluated by means of questions Q11–Q15 in Table 3.12. Lastly, the engagement factor was measured by using

questions Q16-Q23 in Table 3.12.

Table 3.12: Second user study: selection of statements in the questionnaire used for the subjective evaluation (not including those re-used from the first user study or concerning aspects addressed via other methods).

Evaluated Aspect	Question/Statement
Ease of Use	
Q1	The system is easy to use
Q2	The system is simple to use
Q3	The system is user friendly
Ease of Learning	
Q4	I learned how to use the system quickly
Q5	I easily remember how to use the system
Q6	It is easy to learn to use the system
Functionality	
Q7	The level of control provided by the system was appropriate for the task
Q8	The control provided by the system was ambiguous
Q9	I found it easy to access all the functionalities of the system
Q10	It was difficult to remember all the functions available
Consistency	
Q11	The system behaved in a manner that I expected
Q12	It was difficult to understand the operation of the system
Q13	The information presented by the system was consistent
Q14	I was confused by the operation of the system
Q15	The sequence of inputs to perform a specific action matched my understanding of the task
Engagement	
Q16	I felt successful to get involved in
Q17	I felt bored
Q18	I found it impressive
Q19	I forgot everything around me
Q20	I felt frustrated
Q21	I felt completely absorbed
Q22	I felt good
Q23	System's appearance reduced my sense of being involved
Social Attraction	
Q24	I think the robot could be a friend of mine
Q25	I think I could spend a good time with the robot
Q26	I could establish a personal relationship with the robot
Q27	I would like to spend more time with the robot
Social Presence	
Q28	While you were interacting with the robot, how much did you feel as if it was a social being?
Q29	While you were interacting with the robot, how much did you feel as if it was communicating with you?

The second part was designed to assess the *entertainment of interaction* with the considered systems and their suitability in the *role* of the *receptionist* (on a 5-point Likert scale). In particular, the former aspect was assessed (like in [225]) by requesting participants to express their evaluations by means of six attributes, i.e., *enjoyable*, *interesting*, *fun*, *satisfying*, *entertaining*, *boring*, and *exciting*. The latter aspect was analyzed using the same questions of the first study in the same category (Table 3.10).

The third part was designed to perform a direct comparison of the three configurations. In particular, like in [225], participants were requested to rank them with respect to nine assessment categories: *more useful*, *more intelligent*, *enjoy more*, *more interesting*, *more entertaining*, *more boring*, *more frustrating*, *prefer as receptionist*, and *chose from now on*.

The fourth part was meant to study the role of the embodiment and its possible impact on perceived systems' social attractiveness and effectiveness. To this aim, only the physical and virtual embodiments were considered in this part, by requesting participants to assess the *PRM* and *VRM* systems with respect to five dimensions. The first

three dimension, namely,*usefulness*, *companionship* and *intelligence* were to be assessed according to the semantic scale defined in [226, 225]. The fourth dimension, *social attraction*, was evaluated by means of questions Q24–Q27 in Table 3.12. The last dimension, *social presence*, was assessed through questions Q28–Q29 in Table 3.12 and an adapted version of the Interpersonal Attraction Scale [227] (*unsociable/sociable*, *impersonal/personal*, *machine-like/life-like*, *insensitive/sensitive*). A 10-point scale was used for all the dimensions.

Collected data were analyzed first through one-way repeated measures ANOVA test (significance level of 0.05) and a two-tailed paired t-test (significance level of 0.05) in post-hoc analysis, like in the first study.

Objective evaluations in terms of completion time in performing the tasks (for each receptionist system) as well as the percentage of participants who correctly identified the intended room, are illustrated in Figures 3.57 and 3.58, respectively. Concerning time required to complete the room-wayfinding tasks, it can be observed from Figure 3.57 that, participants experimenting the *PRM* system were faster compared to those who experimented the *IM* and *VRM* ones (ANOVA: $p = 0.0286$). This evidence was also confirmed by post-hoc analysis by showing that the differences between the *PRM* and both the *IM* and *VRM* systems were found to be statistically significant ($t[17] = 2.72$, $p = 0.0146$ and $t[17] = 2.13$, $p = 0.0476$, respectively).

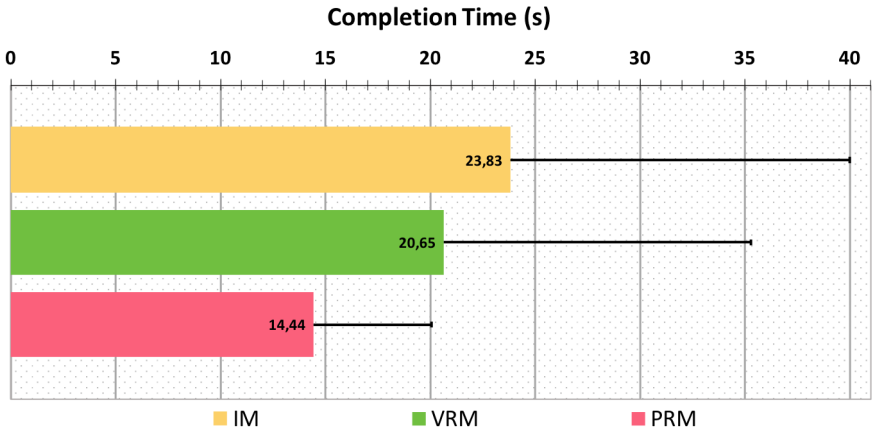


Figure 3.57: Second user study: objective results in terms of time required to complete the wayfinding tasks. Bar lengths report average values (lower is better), whereas whiskers report standard deviation.

By focusing on the success rate in finding the correct room, it is evident from Figure 3.58 that, as in the first study, 100% of the participants were able to reach the intended destination with the *PRM* configuration. However, only 94% and 83% of the participants were able to successfully complete the task with the *VRM* and *IM* configurations.

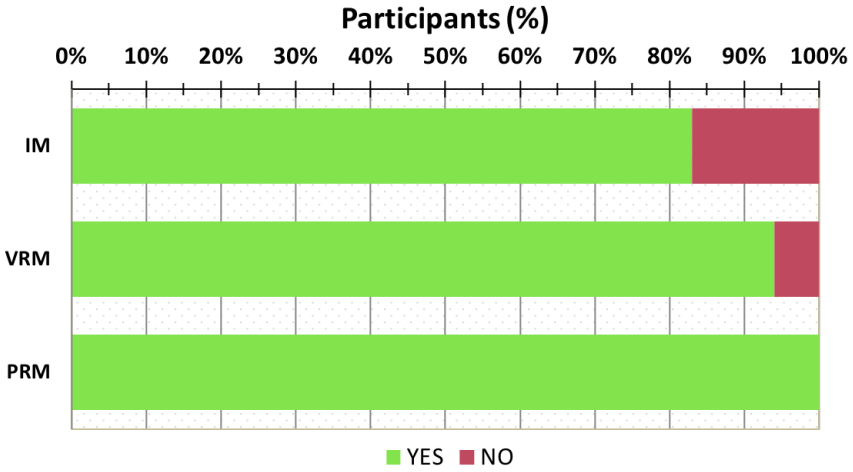


Figure 3.58: Second user study: objective results in terms of success rate in finding the correct room.

Subjective observations in terms of participants' evaluations of the two embodied receptionist systems (with social skills) and the map-based system (without social skills) gathered by the first part of the questionnaire are illustrated in Figures 3.59, 3.60 and 3.61 (+ symbols report ANOVA tests results, i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

By observing the aforementioned figures on the whole, it can be noticed that the *PRM* configuration performed better compared to the *VRM* and *IM* ones for all the usability factors considered, except the *ease of learning* (Fig. 3.59) and *errors* (Fig. 3.60) ones. In particular, all the receptionist systems were judged by participants as easy and quick to learn, and allowing them to easily recover from errors.

By digging more in detail and considering the results obtained with the USE questionnaire (Fig. 3.59), it can be observed that participants judged the *PRM* system as more user-friendly and easier to use than the *IM* and *VRM* ones. However, post-hoc analysis reported in Table 3.13 revealed that, although the difference between the *PRM* and *VRM* configurations was pronounced, it did not reaching statistical significance. A similar consideration holds also for the *satisfaction* factor, where the *PRM* system was judged more pleasant, satisfactory and fun to use than the *IM* and *VRM* ones; however, also in this case no significant difference was found between the two embodied robots (Table 3.13).

Considering subjective results obtained with the NAU methodology and illustrated in Figure 3.60, it can be observed that, overall, the *PRM* system achieved higher scores compared to the other two solutions for all the usability factors. The *VRM* system was judged more positively than the *IM* one in terms of *efficiency*, whereas the *VRM* and *IM* systems were evaluated similarly in terms of *learnability*. This evidence was also confirmed by post-hoc analysis reported in Table 3.14, where no statistically significant differences were found between the *IM* and *VRM* systems. More in detail, the *PRM* system was perceived by participants to be capable to let them complete the task more

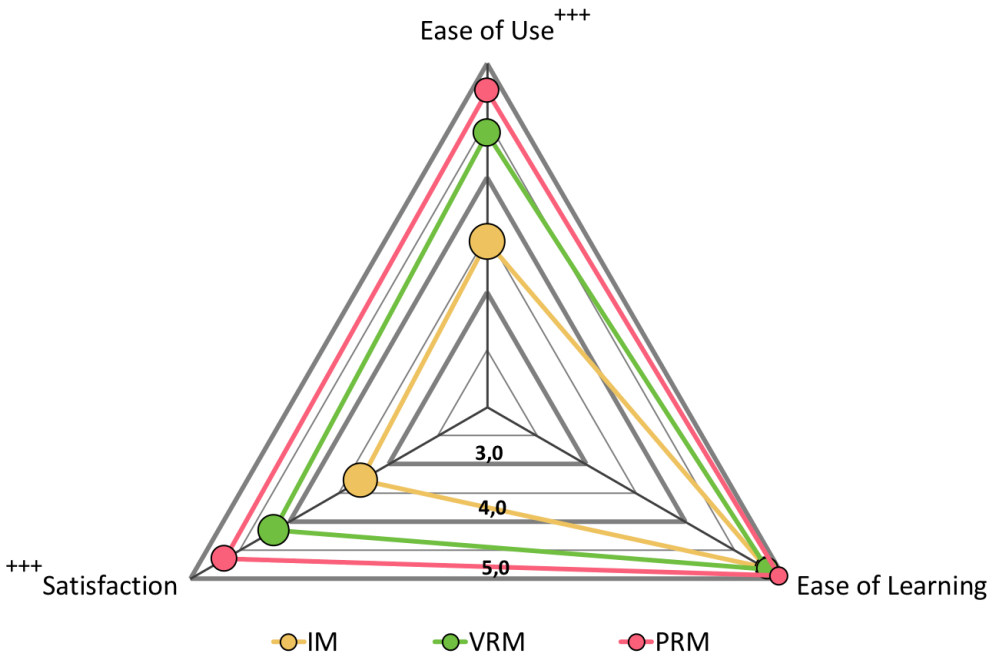


Figure 3.59: Second user study: results concerning the usability of the three receptionist systems based on USE questionnaire. Circle position reports average values (higher is better), circle dimension reports standard deviation, whereas + symbols report statistical significance determined with the ANOVA tests (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

Table 3.13: Second user study: post-hoc analysis on results obtained with the USE questionnaire and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

	IM vs. VRM	**VRM vs. PRM**	**IM vs. PRM**
Ease of Use	t[17] = -3.04, $p = 7.46 \times 10^{-3}$ (++)	t[17] = -1.80, $p = 0.0911$	t[17] = -4.48, $p = 3.28 \times 10^{-4}$ (+++)
Ease of Learning	t[17] = -1.51, $p = 0.1492$	t[17] = -1.68, $p = 0.1106$	t[17] = -0.25, $p = 0.8045$
Satisfaction	t[17] = -3.92, $p = 1.11 \times 10^{-3}$ (++)	t[17] = -1.87, $p = 0.0788$	t[17] = -4.75, $p = 1.86 \times 10^{-4}$ (+++)

easily the first time they used it and to be more effective once they learned to use it compared to with the *IM* and *VRM* ones.

Concerning the *memorability* dimension, participants evaluated the physical robot as the receptionist solution allowing them to re-establishing proficiency in the use of the system with a greater ease compared to the *VRM* and *IM* ones. However, post-hoc analysis revealed no statistically significant difference between the *PRM* and *VRM* systems (Table 3.14).

Data about participants' evaluations collected through the VRUSE methodology are

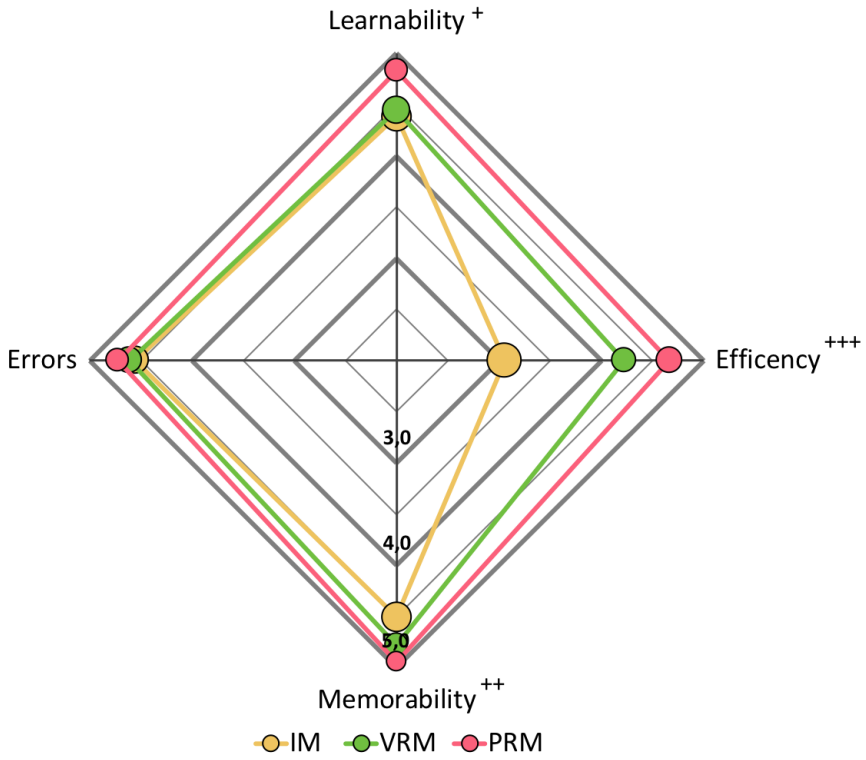


Figure 3.60: Second user study: results concerning the usability of the three receptionist systems based on NAU questionnaire. Circle position reports average values (higher is better), circle dimension reports standard deviation, whereas + symbols report statistical significance determined with the ANOVA tests (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

Table 3.14: Second user study: post-hoc analysis on results obtained with the NAU questionnaire and statistical significance determined with t-tests (+ $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

	IM vs. VRM	**VRM vs. PRM**	**IM vs. PRM**
Learnability	t[17] = -0.43, p = 0.6676	t[17] = -2.36, $p = 3.02 \times 10^{-2}$ (+)	t[17] = -2.41, $p = 2.78 \times 10^{-2}$ (+)
Efficiency	t[17] = -3.70, $p = 1.80 \times 10^{-3}$ (++)	t[17] = -2.68, $p = 1.60 \times 10^{-2}$ (+)	t[17] = -4.68, $p = 2.16 \times 10^{-4}$ (+++)
Memorability	t[17] = -2.56, $p = 2.04 \times 10^{-2}$ (+)	t[17] = -1.84, p = 0.0827	t[17] = -2.68, $p = 1.60 \times 10^{-2}$ (+)
Errors	t[17] = -0.57, p = 0.5786	t[17] = -1.46, p = 0.1631	t[17] = -1.37, p = 0.1871

illustrated in Figure 3.61.

At first sight, it is immediately clear that, also in this case, the *PRM* system was judged more positively by participants compared to the other two solutions. In particular, participants expressed positive evaluations for the *PRM* system both in terms of ease of access to its functionalities and the provided level of control (*functionality*

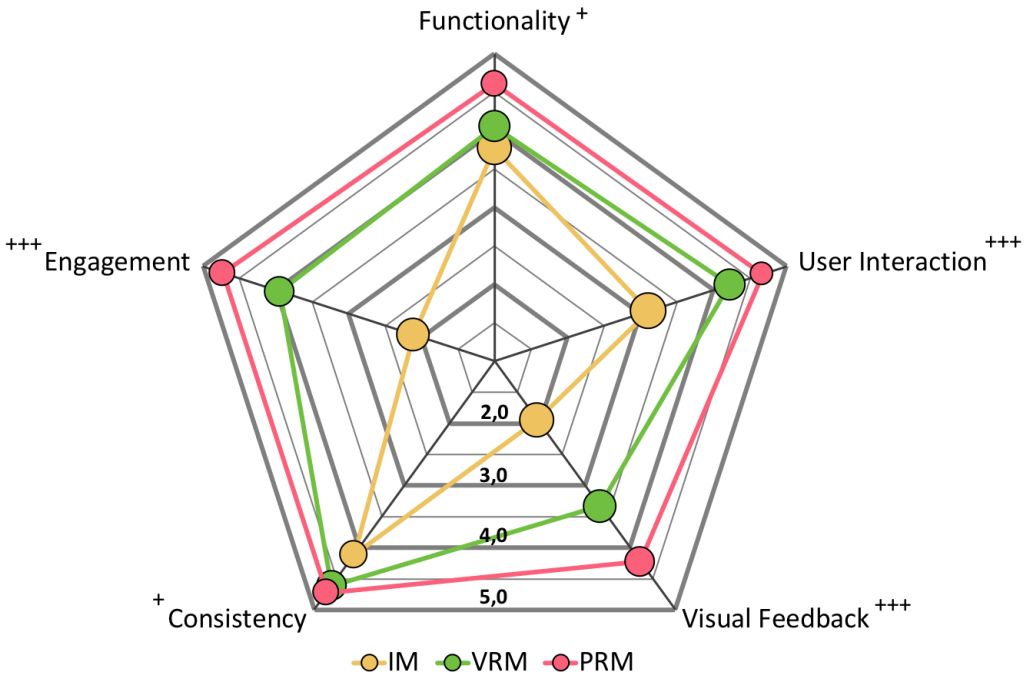


Figure 3.61: Second user study: results concerning the usability of the three receptionist systems based on VRUSE questionnaire. Circle position reports average values (higher is better), circle dimension reports standard deviation, whereas + symbols report statistical significance determined with the ANOVA tests (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

dimension) as well as of the consistency of its operations (*consistency* dimension). No significant difference emerged between the *VRM* and *IM* systems for these two dimensions (Table 3.15). By taking into account the *visual feedback* of the three receptionist systems, a strong preference was obtained for the aspect of the *PRM* than of the *IM* and *VRM* systems. In fact, participants judged it as more visually pleasant and not negatively influencing their performance, whereas the *IM* system was judged to have a negative influence on participants' performance and to be responsible for the reduction of the feeling with the system. Concerning *user interaction* with the systems, participants found it impressive, simple, and uncomplicated the way they could interact with the physical robot; on average, during the interaction, they did not feel disoriented and were always aware of what the system was doing. Rather, the *IM* system was perceived as disorienting, and the way to interact with it was judged by participants as negatively influencing their performance. Finally, the *PRM* system was largely rated as the most impressive, engaging, and completely immersive receptionist solution, whereas the *IM* system was judged as the most frustrating and boring one (*engagement* dimension).

Table 3.15: Second user study: post-hoc analysis on results obtained with the VRUSE questionnaire and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	IM vs. VRM	VRM vs. PRM	IM vs. PRM
Functionality	t[17] = -0.89, p = 0.3843	t[17] = -2.56, $p = 2.04 \times 10^{-2}$ (+)	t[17] = -3.22, $p = 5.03 \times 10^{-3}$ (++)
User Interaction	t[17] = -2.87, $p = 1.06 \times 10^{-2}$ (+)	t[17] = -2.20, $p = 4.16 \times 10^{-2}$ (+)	t[17] = -3.83, $p = 1.34 \times 10^{-3}$ (++)
Visual Feedback	t[17] = -4.15, $p = 6.71 \times 10^{-4}$ (+++)	t[17] = -3.19, $p = 5.38 \times 10^{-3}$ (++)	t[17] = -6.33, $p = 7.53 \times 10^{-6}$ (+++)
Consistency	t[17] = -2.03, p = 0.0579	t[17] = -0.52, p = 0.6073	t[17] = -3.33, $p = 3.93 \times 10^{-3}$ (++)
Engagement	t[17] = -7.46, $p = 9.39 \times 10^{-7}$ (+++)	t[17] = -4.08, $p = 7.77 \times 10^{-4}$ (+++)	t[17] = -2.18, $p = 4.65 \times 10^{-8}$ (+++)

Data gathered in the second part of the questionnaire and related to the *entertainment of interaction* and the suitability of the considered solutions in the *role* of the *receptionist* are illustrated in Figure 3.62.

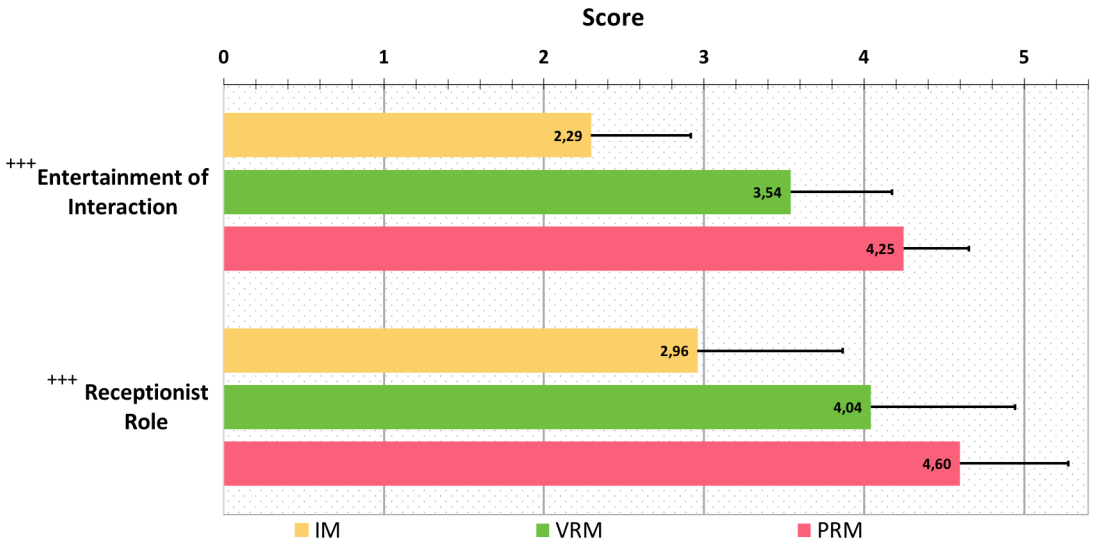


Figure 3.62: Second user study: subjective results obtained from the second part of the questionnaire. Bar lengths report average values (higher is better), whiskers report standard deviation, whereas + symbols report statistical significance determined with the ANOVA tests (i.e., + $p < 0.05$, ++ $p < 0.01$, +++ $p < 0.001$).

Concerning the *entertainment of interaction*, it can be observed that, participants' agreed in promoting the *PRM* solution as the most entertaining, stimulating, enjoyable, interesting, satisfying and exciting receptionist system. Considering the *receptionist role* evaluation, the participants judged the *PRM* system as the best receptionist, reporting that they had fun in receiving directions from it and they would like to use it again in the future. Notwithstanding, the information provided by the three different receptionist systems was considered as equally useful by the participants. Post-hoc analysis reported

in Table 3.16, confirmed the above observations: all the differences between the three receptionist systems were found to be statistically significant.

Table 3.16: Second user study: post-hoc analysis on subjective results obtained from the second part of the questionnaire and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	IM vs. VRM	**VRM vs. PRM**	**IM vs. PRM**
Entertainment of Interaction	t[17] = -6.65, $p = 4.06 \times 10^{-6}$ (+++)	t[17] = -4.29, $p = 4.98 \times 10^{-4}$ (+++)	t[17] = -11.95, $p = 1.07 \times 10^{-9}$ (+++)
Receptionist Role	t[17] = -4.41, $p = 3.86 \times 10^{-4}$ (+++)	t[17] = -2.63, $p = 1.77 \times 10^{-2}$ (+)	t[17] = -5.99, $p = 1.46 \times 10^{-5}$ (+++)

Data about participants' preferences gathered by means of the third part of the questionnaire are reported in Table 3.17. Results show a clear preference for the *PRM* solution over the other two solutions. In particular, the high percentage of participants rating the *PRM* system as more interesting (94%), more entertaining (83%), more intelligent (72%), more enjoyable (89%), preferred as receptionist (78%), and chose from now on (56%), confirming findings obtained with the previous part of the questionnaire.

Table 3.17: Second user study: subjective results (third part of the questionnaire concerning participants' preferences).

Evaluated Aspect	**IM**	**VRM**	**PRM**	**IM = VRM**	**VRM = PRM**	**IM = PRM**
Enjoy more	0%	11%	**89%**	0%	0%	0%
More intelligent	0%	11%	**72%**	0%	17%	0%
More useful	28%	22%	39%	0%	6%	6%
Prefer as receptionist	6%	11%	**78%**	0%	6%	0%
More frustrating	**78%**	17%	6%	0%	0%	0%
More boring	**94%**	6%	0%	0%	0%	0%
More interesting	0%	6%	**94%**	0%	0%	0%
More entertaining	0%	11%	**83%**	0%	6%	0%
Chose from now on	17%	22%	**56%**	0%	6%	0%

In the fourth part of the questionnaire, the two embodied systems were compared. As illustrated in Figure 3.63, it appears that the physical robot was perceived as more friendly and closer to the human users as well as a better companion than the virtual one. The physical robot was also judged by the participants as smarter and more intelligent than the virtual one. Participants felt more socially attracted by the physical robot than by the virtual robot, and expressed rather concordant opinions on the will and pleasure of spending more time with it. Lastly, the physical robot was perceived as a social being possessing a greater social, personal, and realistic presence than its virtual counterpart. As illustrated in Table 3.18, significant differences emerged in all the categories except in the *usefulness* dimension (though differences were pronounced).

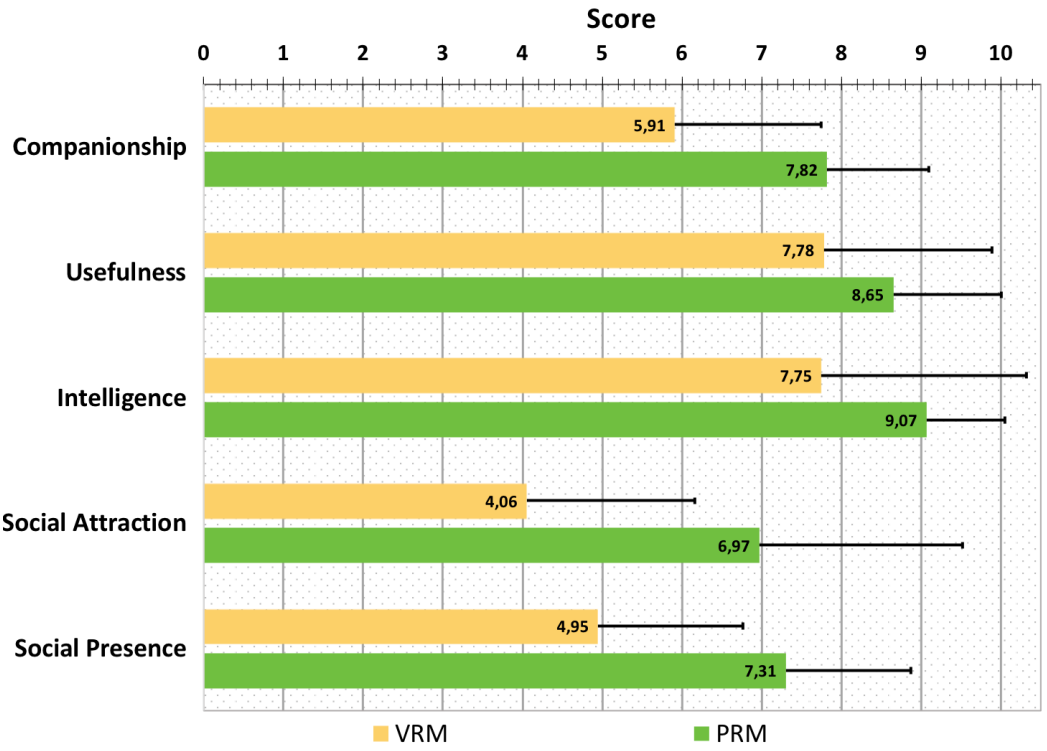


Figure 3.63: Second user study: subjective results obtained from the comparison of the two embodied (i.e., virtual and physical) systems. Bar lengths report average values (higher is better), whereas whiskers report standard deviation.

Table 3.18: Second user study: post-hoc analysis on subjective results obtained from the fourth part of the questionnaire and statistical significance determined with t-tests ($+\ p < 0.05$, $++\ p < 0.01$, $+++\ p < 0.001$).

	VRM vs. PRM
Companion	$t[17] = -3.86,\ p = 1.26 \times 10^{-3}$ (++)
Usefulness	$t[17] = -1.68,\ p = 0.1110$
Intelligence	$t[17] = -2.16,\ p = 4.53 \times 10^{-2}$ (+)
Social Attraction	$t[17] = -5.25,\ p = 6.49 \times 10^{-5}$ (+++)
Social Presence	$t[17] = -4.99,\ p = 1.12 \times 10^{-4}$ (+++)

By summarizing results obtained above, it can be stated that Hypotheses 2 and 3 were largely confirmed. In particular, the beneficial impact of embodied receptionist systems with social behaviors compared to interactive audio-maps in wayfinding applications supported the Hypothesis 2, both in terms of subjective and objective observations. From results obtained in the comparative analysis between the two robotic

receptionists, it can be observed that, physical presence had a measurable impact on humans' performance and perception when compared to a virtual agent displayed on a screen, thus supporting Hypothesis 3.

Chapter 4

Discussion

In this thesis, two HRI spatial proximity patterns, i.e., remote and colocated, were explored within the service robotics domain with the aim of addressing the problems identified in Section 1.3 by developing and implementing some HRI techniques/interfaces that could be appropriate, effective and suitable for ordinary people. This chapter purports to summarize the solutions proposed in the previous sections and the related findings, in order to illustrate their significance and comprehend the overall contributions in the two considered proximity patterns. The next chapter concludes this discussion by articulating the lessons learned from the performed activities and the scope of obtained contributions and will finish the chapter with an outline of where this dissertation points in terms of future work for HRI in service robotics applications.

4.1 Consideration of Remote Spatial Proximity Pattern

In the remote spatial proximity pattern, two application domains (i.e., *robotic telepresence* and *robotic aerial traffic monitoring*) were explored for investigating the two different interfaces that are prevalent in this model of interaction, namely, *teleoperation* and *supervisory control*.

In the *robotic telepresence* domain, two user studies were conducted. The former study consisted of a comparative analysis of user interfaces for robotic telepresence. The evaluation focused on two main control modalities used today in most experimental and commercial solutions, i.e., keyboard and point-and-click video navigation. The combination of the two methods was also considered. The second study was conducted to assess the impact on users' SA and performance of different camera configurations (in terms of FOV size and pan-tilt availability) in remote robots navigation tasks, both considering objective and subjective factors. Experimental results obtained through the first user study provided precious indications about user experience with the three interfaces, both in objective and subjective terms. In particular, objective observations

indicated that, the time required to complete a complex (navigation) task as well as the number of needed interactions can be significantly reduced when using the combined interface (i.e., keyboard plus point-and-click video navigation) rather than one single navigation interfaces at a time. Similarly, subjective observations showed that the favorite interface was the combined one. However, based on the feedback gathered during the tests, it could be observed that users' preference for the combined interface was due to the fact that they were allowed to switch between the two interfaces when needed, thus benefiting from the advantages of both of them. Specifically, considering the performances obtained and preferences expressed in the execution of specific navigation operations, it was found that the keyboard-based interface actually provided significant advantages when accurate control was needed, whereas point-and-click video navigation was more effective when robot's autonomous navigation capabilities could be exploited. In the second analysis, experimental results in terms of objective observations showed that camera configurations characterized by wider FOVs are more effective, as they allow users to carry out tasks in less time and with fewer navigation commands. Similarly, results obtained through subjective observations suggested that the configuration exhibiting a fisheye FOV was the less cognitive demanding configuration compared to those featuring a narrow FOV or a wide FOV with a pan-tilt camera. The configuration endowed with a wide FOV and a pan-tilt camera was judged by the users as the configuration providing the highest SA. Based on the preferences expressed after the experiments, this evidence was mainly motivated by the fact the fisheye-FOV configuration allowed users to spot a larger portion of the robot's surroundings by making it easier for them to exploit the semi-autonomous point-and-click video navigation modality. The configuration exhibiting a wide FOV and a pan-tilt camera was preferred because of the pan-tilt capability, which allowed users to better explore the scene without necessarily having to move the robot.

In the *robotic aerial traffic monitoring* domain, an adjustable autonomy system exploiting decision-making capabilities was developed to assist UAV controllers by predicting the appropriate LOA relying on operators' MW measurements in drone monitoring scenarios. Three different LOAs, each of which with its corresponding interface, were considered in this domain, namely, *warning*, *suggestion* and *autonomous*. In the warning LOA, the system warns the UAVs controller if critical situations occur; the suggestion LOA suggests feasible actions to him or her; the autonomous LOA monitors and performs actions autonomously without any human intervention. A Bayesian Network classifier was first exploited as learning probabilistic model and the NASA-TLX questionnaire as subjective workload assessment technique. Obtained results showed that the proposed model was able to predict the appropriate LOA with an accuracy of 83.44%. However, based on feedbacks collected from users during the tests, a limitation of the proposed approach was represented by the subjective technique used to gain the training data. In fact, users had to answer a large number of questions in order to evaluate their cognitive load in all different test conditions, thus requiring a considerable amount of time. For this reason, a second study exploiting EEG signals as physiological

measures and a SVM with two different kernels (linear and RBF) as learning model was conducted to build a MW prediction model based on UAV operators' cognitive demand by evaluating operators' performance in accomplishing the assigned tasks. A classification and validation procedure was then performed to both categorize the cognitive workload measured by EEG signals and evaluate the obtained patterns from the point of view of accuracy. Obtained results showed that the SVM with the linear kernel is able to predict the operator's MW with an average accuracy equal to 95.8%, whereas, an accuracy equal to 94.1% is reached with the SVM - RBF kernel. In summary, it can be observed that the physiological measures combined with the SVM resulted to perform better compared to the subjective measures and the BN classifier. This is reasonably due to the fact that physiological measurements capture cognitive information in real-time and continually thus with higher reliability in the measurements. The outcome of this study also suggests that small devices with wireless acquisition systems are promising BCI technologies for monitoring and assessing MW used (in this case) to assist UAV controllers.

4.2 Consideration of Colocated Spatial Proximity Pattern

In the colocated spatial proximity pattern, two application domains, namely, *robotic gaming* and *assistive robotics* were explored to investigate the different interaction paradigms with proximate robots as well as users' perception about colocated robotic systems.

In the *robotic gaming* domain, two robotic games were developed by sticking to state-of-the-art game design guidelines. In the first game, called *Protoman Revenge*, a COTS drone was used as a robotic player. The game was conceived by taking into account the role of autonomous robot's behaviors and emotional features. The impact of these elements on the user experience was evaluated by means of a user study, in which three different versions of the game (i.e., a full version including both autonomous robot's behaviors and emotional features, a version without autonomy, a version without emotional features) were played to identify the contribution of distinct game elements. Despite the limitation due to the small sample size of volunteers, the obtained results showed that drone autonomous behaviors do not introduce significant improvements in the overall experience, but corroborate the key role played by other elements such as believable natural movements and illusion of life. Emotional features noticeably increased players' engagement and enjoyment of the game while, at the same time, affecting positively HRI under the aspects of emotional distinction, personality and interaction with the drone. In the second game, called *RoboQuest*, it was possible to demonstrate that a MR Phygital Play platform can be exploited to set up cloud robotics-based gaming scenarios capable to reach most of the objectives identified in this domain. In particular, in *RobotQuest* a seamless interaction between virtual

and real game elements is achieved. Digital content projected on the floor are used to display the story, but also to drive the robot (via line following, making it act as an autonomous, intelligent entity) and to augment its physical behaviors (through virtual shoots and explosions that accompany its movements during the battle). The use of role play game logic combined with proximity interaction is regarded as a means to limit players' sedentary and solitary behaviors. In fact, during the game players are requested to move in the play area to position some TUIs, but also to collaborate with other players in order to build up the required support tools (through combinations of TUIs). The devised solution also complies with the PIRG game design principles, particularly regarding the possibility to use existing hardware and to keep costs down. It is worth observing that many other game designs could be easily developed to tackle the issues of interest by reusing some or all of the features exploited in *RobotQuest*, like the presence of intersections (choices to make), enemies (obstacles to pass), support tools (players' inputs), etc. The focus could be on pure entertainment, like in this case, but also, e.g., on education (by simply changing the story and game elements). In particular, the devised solution could be ideally deployed at home or at school, where a projector may be already available.

In the *assistive robotics* domain, two use cases were considered involving two different robots that vary in the interaction paradigms used to assist human users. The first case involved a mobile robotic assistant featured by a pyramidal shape able to guide people in an unknown environment by accompanying them along the path while providing AR hints. The second use case involved a socially interactive robotic assistant exhibiting a humanoid shape and placed in (fixed to) a reception desk able to greet people, provide them with vocal-, arm pointing gestures-based directions and with human-like social behaviors (e.g., gaze, face tracking) as well as to say goodbye.

In the first use case, a simulation framework was created with the aim of supporting the study of interaction paradigms for robotic applications. The proposed framework was exploited to help the development of NUIs to be used in a service robotic scenario. To this aim, two interfaces were designed: one based on head/gaze motion tracking and AR hints, the other based on body motion tracking and non AR hints, displaying them on a tablet mounted on the top of the robotic assistant. Both interfaces provided voice commands. In this study, the above interfaces were tailored to the control of a set of selected office-oriented robotic tasks. Experimental results obtained through a user study provided precious indications about user experience with the two interfaces, both in objective and subjective terms. Results obtained with the objective evaluation showed that users were faster in performing tasks with the interface exhibiting AR hints with respect to that displaying them on the tablet. Results gathered from the subjective evaluation showed a strong users' preference for the AR interface in terms of ease of use, time and support information provided in completing the task assigned. This preference arises from the fact that with the AR interface human users could interact with the robot just by using their gaze and/or their voice. Moreover, they could always receive feedback from the robot or useful information for the execution of the task by looking

at the robot even at distance.

In the second use case, the aim was to investigate how embodied and socially interactive robotic assistant used as direction-giving systems can be perceived by human users and can impact on their wayfinding performance. To this aim, two user studies were conducted.

In the first study, the impact of using a map integrated with a physical robot compared both to a physical robot giving directions with arm pointing gestures and (for the sake of completeness) the virtual version of this last configuration, was addressed. Findings showed that participants found it easier and faster to locate their destination when the robot showed them directions on a map rather than when it used arm pointing gestures. Based also on the comments provided after the experiments, preference appears to be mainly motivated by the fact that, with the introduction of the map, all the participants were able to better understand and remember the route described by the robot. The difference among the three configurations emerges in a rather clear way also considering the usability analysis. In fact, the robot with the map was the preferred configuration for what it concerns both satisfaction and user interaction, followed by the physical robot using arm pointing gestures. No significant difference was found in terms of visual feedback between map and arm pointing gestures for the physically embodied systems, whereas the virtual embodied robot was judged as responsible for worse performance. Concerning the robots' suitability as receptionists, the map-based configuration was evaluated as the best one, whereas no difference was found between the two systems using arm pointing gestures.

In the second study, the role of different embodiments (namely virtual and physical) and social behaviors for a receptionist system were investigated from different perspectives and compared to unembodied (not socially interactive) wayfinding means represented by an interactive audio-map. Objective observations showed that participants were faster in finding the intended destination and less prone to get lost when using the embodied systems than the unembodied one. Concerning subjective evaluation, participants judged the virtual and physical robots as easier to use, more satisfying, and more efficient than the interactive audio-map; the two embodied systems were also considered as capable to provide a better visual feedback, as well as to be more fun and entertaining. The interactive audio-map was perceived as disorienting, frustrating, reducing the engagement, and worsening performance; these findings were reasonably due to the fact that, with this system, participants had to visually spot and select the intended destination by either clicking it on the map or in the list. These evidences are confirmed also by participants' judgments concerning systems' suitability as receptionist solutions. In fact, a clear preference for the physical robot was recorded, followed by the virtual one and, lastly, by the map. Although usefulness of provided information was considered as comparable for the three systems, results above confirmed that embodied receptionist systems with social behaviors lead to better performance in wayfinding tasks compared to interactive audio-maps. When it comes to compare the two robotic receptionists, it can be observed that, when interacting with the physical receptionist

system, participants were always able to correctly remember and identify the requested destination, whereas the use of the virtual one negatively influenced their performance. Feedback collected through the questionnaire indicated that it was easier for them to learn how to interact with the physical robot than with the virtual one the first time they used them; moreover, they judged the physical receptionist system as more efficient and offering a simpler access to its functionalities, and were more engaged and impressed by the way they could interact with it than with the virtual receptionist system. Focusing on social perception of the robots, the physical robot was found to exhibit a higher companionship than the virtual one, and to be closer to participants; it was also perceived as more intelligent, and to have a more realistic aspect, social attraction, and presence. These findings confirm that physical presence has a measurable impact on humans' performance and perception when compared to a virtual agent displayed on a screen, as suggested by [222, 223]. This evidence is also supported by comments the participants provided at the end of the experiment: in fact, many of them stated that interacting with the physical receptionist system increased their level of attention and made them feel more focused, allowing them to better understand and remember the received directions.

Chapter 5

Conclusions and Future Work

In this thesis, different aspects of HRI field in service robotics applications were investigated. The starting point was to survey the state-of-the-art in the field by spanning the space between semi-autonomous and fully tele-controlled solutions. Despite the broadness of the explored field, through these activities it was possible to realize that HRIs could be classified according to two important correlated dimensions: the spatial proximity between humans and robots (remote or physically colocated) and the type of interaction (i.e., indirect or direct), thus defining the remote and colocated spatial proximity patterns. Afterwards, research activities were focused on exploring these patterns to detect the open problems as well as the arising challenges, in order to identify those interfaces that could be regarded as more appropriate and effective. As a result, on the one hand, it was possible to observe that research issues arising from remote spatial proximity pattern are related to poor human users' SA of the remote environment and their high MW arising both from the execution of tasks at distance or from the supervision of a large number of robots. On the other hand, it was found that common research issues in colocated spatial proximity pattern are related to the usability of the paradigms used by human users to interact with robots and to users' acceptability of these robotic systems populating the same environment. With the aim of addressing the aforementioned problems and proposing new approaches to advance the state-of-the-art in the field, some application domains were explored and studied as representative examples of the above proximity categories, and different interfaces were designed and developed.

Specifically, the robotic telepresence and the robotic aerial traffic monitoring domains were selected as representative of remote HRI scenarios exploiting teleoperation and supervisory control interfaces, respectively. By digging more in details, the former domain was explored to study other possible teleoperation interfaces for robotic applications than those investigated in previous works. The latter domain was investigated to study possible supervisory control interfaces leveraging adaptive automation systems. Furthermore, the robotic gaming and assistive robotics domains were chosen for representing colocated HRI scenarios dealing with real use cases human users are or will be

expected to be involved into. More in detail, the first domain was chosen to study user acceptability of toy service robots endowed or not with autonomous behaviors and exhibiting or not emotional features when engaged in recreational activities. The second domain was explored, on the one hand, to investigate the usability of different natural interaction paradigms (NUIs) with a local semi-autonomous robot by using virtual and AR techniques, hand and body gestures as well as touch, speech and eye gaze-based commands. On the other hand, it was explored to study users' acceptability when involved in robotic receptionist scenarios with robots exhibiting or not human-like appearance, endowed or not with social interaction skills and being physically present with human users or virtually displayed on a screen.

By leveraging the above application domains and in particular the results obtained from the various experiments and tests performed to validate each implemented solution, it was possible to identify and outline a set of UI design requirements to improve the HRI and make it more effective in both the remote and colocated spatial proximity patterns.

Concerning remote spatial proximity patterns, what emerged from the performed activities, and is interesting to highlight, is the fact that providing the robots (or the robot supervision systems) with different *LOAs* and the capability to flexibly slide between this *LOAs* range, leads to better users' performance and lowers their MW. Specifically, by recalling the results obtained from the telepresence application domain, it is possible to observe how human users were able to accomplish the navigation tasks in a more effective way and with lower MW when they were allowed to switch between an assisted teleoperation interface (using a keyboard) and a semi-autonomous teleoperation interface (point-and-click on the screen) and preferred this combination also in terms of usability. Regarding the robotic aerial traffic monitoring domain, obtained results showed that systems featured by supervisory control interfaces equipped with decision-making capabilities and able to flexibly adjust their *LOAs* are capable to assist human users in the role of supervisors by lowering their MW and improving their performance in accomplishing the given task. By focusing on the human users in the role of operators, it is possible to highlight that robots equipped with *wide FOV* cameras and *pan-tilt* capabilities lead to higher users' SA while keeping their MW relative low. From the above analysis, two design requirements have emerged clearly and are listed below.

- *Adjustable Autonomy*: importance of providing robots with different *LOAs* and the ability to flexibly balance the amount of control that users have over them;
- *Flexible Wide Vision*: need to equip robots with *wide FOV* cameras combined with *pan-tilt* capabilities in order to provide users with better sensory and/or contextual feedback while keeping the interaction with the robot as simple as possible.

Concerning colocated spatial proximity patterns, two interesting aspects have emerged and need to be highlighted. The former is that robots exhibiting *human-like*

behaviors and/or *human-like appearance* are more acceptable, engaging and lead to better performance. The latter is that those robots not equipped with the above human-like features but with *NUIs/TUIs* and exploiting *AR* and/or *MR* techniques were found to be effective and easy to use. Specifically, by recalling the studies involving robotic systems equipped with *human-like behaviors* and/or *appearance* (i.e., the robotic game leveraging a drone and the socially interactive robotic assistant) users exhibited a strong preference for these systems as well as better performance rather than for the versions without these human-like features. In particular, for the drone-based robotic game, users reported a better user experience and more appreciation for the version of the game showing emotional features and with the robot behaving as a rational agent. Considering the socially interactive robotic assistant use case, results showed that people performed better when they received directions on a map by humanoid receptionist systems. Moreover, they suggested that socially interactive embodied systems are more appreciated and perceived as social being by human users and have a positive impact on users' preference as well as on their performance in the execution of wayfinding tasks compared to an unembodied interactive map. Lastly, they indicated that the physicality of the experience (in addition to the *human-like appearance* and *behaviors*) with a physically present humanoid robot can help people to achieve better performance with respect to the same robot but displayed on a screen. By focusing on the studies exploiting *AR* or *MR* techniques combined with *NUIs/TUIs*, it can be observed that the users experimenting with these interfaces judged them to be more usable in carrying out given tasks with better performance. Moreover, it was demonstrated that the combination of real and digital elements can be used to foster users' engagement and collaboration in robotic gaming usage contexts. From the above analysis, further design requirements have emerged clearly and are listed below.

- *Human-like Robot Interaction*: need to design and build robots exhibiting *human-like appearance* and the ability to interact with users through human-like paradigms in order to allow people to treat and think to robots as social entities that are easy to understand and to work with, thus improving the quality of the interaction;

- *Augmented N(T)UIs*: importance of enhancing the *NUIs/TUIs* through the use of *AR* and/or *MR* techniques for creating robots capable to both get people involved and focused on the tasks to be performed by achieving the objective of the intended activities (e.g., carrying out tasks or have fun).

In conclusion, the aforementioned findings provide evidence that to achieve a more effective and satisfactory HRI both in remote and colocated spatial proximity patterns, different requirements need to be considered. This dissertation aims at providing researchers with these design recommendations in order to be used as simple tools allowing them to apply the overall lessons learned without having to investigate the domain in depth again.

5.0.1 Future Work

Although the studies conducted in this thesis allowed to delineate a set of guidelines/requirements to be taken into account for providing effective HRIs in service robotics applications, work is still needed to investigate issues that have been identified in each explored pattern.

Concerning remote proximity patterns, and in particular the robotic telepresence domain, future work could be aimed, for instance, to explore the effect of dynamically combining the considered wide-angle FOV configurations, by letting the users switch between them depending on the situation, thus benefiting from the advantages offered by both of them. In the robotic aerial traffic monitoring domain, future work may focus on exploiting alternative data analysis and classification procedures in order to achieve a real-time (rather than offline) evaluation of the collected data.

In collocate spatial proximity patterns, future work in the robotic gaming domain could be aimed to extend player-robot interaction, e.g., by using voice and gesture commands and improving the synesthetic experience by adding haptic feedback to a player's game tools. Further activities could be performed on the *Phygital Play* platform, with the goal to improve its capabilities and make it possible to devise other game concepts coping with new challenges. From a technical point of view, alternative methods to localize and track robots may be explored. Lastly, different ways to combine physical and digital content could be tested, by using, for instance, holographic projections based on head-mounted displays. In the mobile robotic assistant use case, alternative simulation environments considering, for instance, robot's dynamics and the physics laws of the real world could be experimented to gather further indications about feasibility and suitability of the proposed NUIs before actually passing to the production and/or renovation phases of the real robots. In the socially interactive robotic assistant use case, albeit comparing virtual agents and robots with diverse sizes is a common practice in works found in the literature, future work could be aimed to investigate the influence of different visual appearances on comparative experiments. The same consideration holds also for the effect of social behaviors and embodiment; since these two factors were jointly considered in this domain, follow-up studies may be designed in order to gain a more in-depth knowledge on how they may have separately influenced the experiments reported in this thesis. Moreover, further configurations not considered in the performed studies (e.g., encompassing a virtual map with a physical robot or vice versa, a different way to use audio feedback, etc.) and other application scenarios with diverse types of users could be explored, in order to determine whether there are solutions that could be better suited to carry out the tasks of interest depending on the specific scenario considered.

Bibliography

[1] IFR - International Federation of Robotics. *World Robotics 2017 Service Robots*. Available online: https://ifr.org/downloads/press/Executive_Summary_WR_Service_Robots_2017_1.pdf (accessed on 10 April 2018).

[2] Kerstin Dautenhahn. "Socially intelligent robots: dimensions of human–robot interaction". In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362.1480 (2007), pp. 679–704.

[3] Jennifer Casper and Robin R. Murphy. "Human-robot interactions during the robot-assisted urban search and rescue response at the world trade center". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 33.3 (2003), pp. 367–385.

[4] Brad Cain. *A review of the mental workload literature*. Tech. rep. Defence Research and Development Toronto, (2007), p. 35.

[5] Anca D. Dragan and Siddhartha S. Srinivasa. "A policy-blending formalism for shared control". In: *The International Journal of Robotics Research* 32.7 (2013), pp. 790–805.

[6] Shervin Javdani, Siddhartha S. Srinivasa, and J. Andrew Bagnell. "Shared autonomy via hindsight optimization". In: *Robotics: Science and Systems XI*. (2015).

[7] Leila Takayama, Eitan Marder-Eppstein, Helen Harris, and Jenay M. Beer. "Assisted driving of a mobile remote presence system: system design and controlled user evaluation". In: *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. (2011), pp. 1883–1889.

[8] Jijun Wang. "Human control of cooperating robots". PhD thesis. University of Pittsburgh, (2008).

[9] Lauren Milliken and Geoffrey A. Hollinger. "Modeling user expertise for choosing levels of shared autonomy". In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. (2017), pp. 2285–2291.

[10] Julie A Adams. "Critical considerations for human-robot interface development". In: *Proceedings of 2002 AAAI Fall Symposium*. (2002), pp. 1–8.

[11] International Federation of Robotics. *World Robotics 2016. Definition of service robots*. Available online: http://www.ifr.org/service-robots/ (accessed on 10 April 2018).

[12] Holly A. Yanco and Jill Drury. "Classifying human-robot interaction: an updated taxonomy". In: *2004 IEEE International Conference on Systems, Man and Cybernetics*. Vol. 3. IEEE. (2004), pp. 2841–2846.

[13] Jenay M. Beer, Arthur D. Fisk, and Wendy A Rogers. "Toward a framework for levels of robot autonomy in human-robot interaction". In: *Journal of Human-Robot Interaction* 3.2 (2014), pp. 74–99.

[14] International Federation of Robotics. *World Robotics 2016. Definition of industrial robots*. Available online: https://ifr.org/industrial-robots (accessed on 10 April 2018).

[15] Hit Robot Group. *Service Robots Business Unit*. Available online: http://www.hitrobotgroup.com/en/public/image/fuwu_bg.jpg (accessed on 26 November 2018).

[16] Karol Niechwiadowicz and Zahoor Khan. "Robot based logistics system for hospitals-survey". In: *IDT Workshop on Interesting Results in Computer Science and Engineering*. Citeseer. (2008).

[17] Eitan Marder-Eppstein, Eric Berger, Tully Foote, Brian Gerkey, and Kurt Konolige. "The office marathon: robust navigation in an indoor office environment". In: *2010 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. (2010), pp. 300–307.

[18] Horst-Michael Gross and Hans-Joachim Boehme. "Perses - a vision-based interactive mobile shopping assistant". In: *2000 IEEE International Conference on Systems, Man, and Cybernetics*. Vol. 1. IEEE. (2000), pp. 80–85.

[19] Illah R. Nourbakhsh, Clayton Kunz, and Thomas Willeke. "The mobot museum robot installations: A five year experiment". In: *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003. (IROS 2003)*. Vol. 4. IEEE. (2003), pp. 3636–3641.

[20] Lauren Goode. *More Roombas are getting Wi-Fi, so you can control your robot vacuum remotely*. Available online: https://www.theverge.com/2017/5/2/15512844/irobot-roomba-890-690-wifi-alexa-robot-vacuum (accessed on 03 January 2019).

[21] Kirk Miller. *This Robo-sentry is basically a Roomba for home security*. Available online: https://www.insidehook.com/nation/like-a-roomba-for-home-security (accessed on 03 January 2019).

[22] Sony Corporation. *Entertainment Robot "aibo"*. Available online: https://www.sony.net/SonyInfo/News/Press/201711/17-105E/index.html (accessed on 03 January 2019).

[23] The Guardian. *Robear: the bear-shaped nursing robot who'll look after you when you get old*. Available online: https://www.theguardian.com/technology/2015/feb/27/robear-bear-shaped-nursing-care-robot (accessed on 03 January 2019).

[24] Simon Cocking. *Robots podcast ladybird, with james underwood from australian centre for field robotics*. Available online: https://irishtechnews.ie/robots-podcast-ladybird-with-james-underwood-from-australi-an-centre-for-field-robotics (accessed on 03 January 2019).

[25] Avidbots. *This is Neo*. Available online: https://www.avidbots.com (accessed on 03 January 2019).

[26] All on robots.com. *Search and rescue robots*. Available online: http://www.allonrobots.com/rescue-robots.html (accessed on 03 January 2019).

[27] Marine Knowledge.com. *ROV: An underwater robot*. Available online: http://www.marine-knowledge.com/types_of_ships/an-underwater-robot (accessed on 03 January 2019).

[28] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. "Common metrics for human-robot interaction". In: *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot interaction*. ACM. (2006), pp. 33–40.

[29] Jennifer L. Burke, Robin R. Murphy, Erika Rogers, Vladimir J. Lumelsky, and Jean Scholtz. "Final report for the DARPA/NSF interdisciplinary study on human-robot interaction". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34.2 (2004), pp. 103–112.

[30] Holly A. Yanco and Jill L. Drury. "A taxonomy for human-robot interaction". In: *Proceedings of the AAAI Fall Symposium on Human-Robot Interaction*. (2002), pp. 111–119.

[31] Justin Storms, Kevin Chen, and Dawn Tilbury. "Modeling teleoperated robot driving performance as a function of environment difficulty". In: *International Federation of Automatic Control (IFAC)-PapersOnLine* 49.32 (2016), pp. 216–221.

[32] V. Molino, Rajmohan Madhavan, E. Messina, T. Downs, A. Jacoff, and S Balakirsky. "Traversability metrics for urban search and rescue robots on rough terrain". In: *Proceedings of the Performance Metrics for Intelligent Systems* (2006), pp. 77–84.

[33] Jean Scholtz. "Theory and evaluation of human robot interactions". In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*. IEEE. (2003), 10–pp.

[34] Clarence A. Ellis, Simon J. Gibbs, and Gail Rein. "Groupware: some issues and experiences". In: *Communications of the ACM* 34.1 (1991), pp. 39–58.

[35] Sebastian Thrun. "Toward a framework for human-robot interaction". In: *Human-Computer Interaction* 19.1 (2004), pp. 9–24.

[36] Michael A. Goodrich and Alan C. Schultz. "Human–robot interaction: a survey". In: *Foundations and Trends® in Human–Computer Interaction* 1.3 (2007), pp. 203–275.

[37] Kazuhiko Kawamura, Phongchai Nilas, Kazuhiko Muguruma, Julie A. Adams, and Chen Zhou. "An agent-based architecture for an adaptive human-robot interface". In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*. IEEE. (2003), 8–pp.

[38] Brenden Keyes, Mark Micire, Jill L. Drury, and Holly A Yanco. "Improving human-robot interaction through interface evolution". In: *Human-Robot Interaction*. InTech. (2010).

[39] Daniel Wigdor and Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch And Gesture*. 1st. Morgan Kaufmann Publishers Inc., (2011).

[40] Hans-Joachim Böhme, Torsten Wilhelm, Jürgen Key, Carsten Schauer, Christof Schröter, Horst-Michael Groß, and Torsten Hempel. "An approach to multimodal human–machine interaction for intelligent service robots". In: *Robotics and Autonomous Systems* 44.1 (2003), pp. 83–96.

[41] Mauro Dragone, Thomas Holz, and Gregory MP O'Hare. "Mixing robotic realities". In: *Proceedings of the 11th International Conference on Intelligent User Interfaces*. ACM. (2006), pp. 261–263.

[42] Ian Yen-Hung Chen, Bruce MacDonald, and Burkhard Wunsche. "Mixed reality simulation for mobile robots". In: *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. (2009), pp. 232–237.

[43] Cheng Guo and Ehud Sharlin. "Exploring the use of tangible user interfaces for human-robot interaction: a comparative study". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. (2008), pp. 121–130.

[44] Jill L. Drury, Jean Scholtz, and Holly A. Yanco. "Awareness in human-robot interactions." In: *Proceedings of the 2003 IEEE International Conference on Systems Man and Cybernetics*. Vol. 1. (2003), pp. 912–918.

[45] Mary Beth Rosson and John M Carroll. *Usability Engineering: Scenario-Based Development of Human-Computer Interaction*. Morgan Kaufmann, (2002).

[46] Candace L. Sidner, Christopher Lee, Cory D. Kidd, Neal Lesh, and Charles Rich. "Explorations in engagement for humans and robots". In: *Artificial Intelligence* 166.1-2 (2005), pp. 140–164.

[47] Andrew Dillon. "User acceptance of information technology". In: *Encyclopedia of Human Factors and Ergonomics*. London: Taylor and Francis, (2001).

[48] Federica Bazzano, Fabrizio Lamberti, Andrea Sanna, Gianluca Paravati, and Marco Gaspardone. "Comparing usability of user interfaces for robotic telepresence." In: *12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*. (2017), pp. 46–54.

[49] Federica Bazzano, Angelo Grimaldi, Fabrizio Lamberti, Gianluca Paravati, and Marco Gaspardone. "Adjustable autonomy for UAV supervision applications through mental workload assessment techniques". In: *International Conference on Intelligent Human Computer Interaction*. Springer. (2017), pp. 32–44.

[50] Federica Bazzano, Paolo Montuschi, Fabrizio Lamberti, Gianluca Paravati, Silvia Casola, Gabriel Ceròn, Jaime Londoño, and Flavio Tanese. "Mental workload assessment for UAV traffic control using consumer-grade BCI equipment". In: *International Conference on Intelligent Human Computer Interaction*. Springer. (2017), pp. 60–72.

[51] Federica Bazzano, Fabrizio Lamberti, Andrea Sanna, and Marco Gaspardone. "The Impact of Field of View on Robotic Telepresence Navigation Tasks". In: *Computer Vision, Imaging and Computer Graphics – Theory and Applications*. Springer International Publishing, (2019), pp. 66–81.

[52] Marvin Minsky. "Telepresence". In: *Omni* 2.9 (1980), pp. 44–52.

[53] Thomas B Sheridan. "Teleoperation, telerobotics and telepresence: a progress report". In: *Control Engineering Practice* 3.2 (1995), pp. 205–214.

[54] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. "A review of mobile robotic telepresence". In: *Advances in Human-Computer Interaction* 2013 (2013), p. 3.

[55] Irene Rae, Gina Venolia, John C. Tang, and David Molnar. "A framework for understanding and designing telepresence". In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM. (2015), pp. 1552–1566.

[56] Jessie YC Chen, Ellen C. Haas, and Michael J Barnes. "Human performance issues and user interface design for teleoperated robots". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37.6 (2007), pp. 1231–1245.

[57] James S. Tittle, Axel Roesler, and David D Woods. "The remote perception problem". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 46. 3. SAGE Publications Sage CA: Los Angeles, CA. (2002), pp. 260–264.

[58] Paulo G. De Barros and Robert W Linderman. *A survey of user interfaces for robot teleoperation*. Tech. rep. Worcester Polytechnic Institute, (2009).

[59] Curtis W Nielsen and Michael A Goodrich. "Comparing the usefulness of video and map information in navigation tasks". In: *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*. ACM. (2006), pp. 95–101.

[60] David D. Woods, James Tittle, Magnus Feil, and Axel Roesler. "Envisioning human-robot coordination in future operations". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34.2 (2004), pp. 210–218.

[61] Brenden Keyes. "Evolution of a telepresence robot interface". PhD thesis. University of Massachusetts., (2007).

[62] Terrence Fong and Charles Thorpe. "Vehicle teleoperation interfaces". In: *Autonomous Robots* 11.1 (2001), pp. 9–18.

[63] Rainer Stiefelhagen, Hazim Kemal Ekenel, Christian Fugen, Petra Gieselmann, Hartwig Holzapfel, Florian Kraft, Kai Nickel, Michael Voit, and Alex Waibel. "Enabling multimodal human–robot interaction for the karlsruhe humanoid robot". In: *IEEE Transactions on Robotics* 23.5 (2007), pp. 840–851.

[64] Tristan Lewis, Jill Drury, and Brandon Beltz. "Evaluating mobile remote presence (MRP) robots". In: *Proceedings of the 18th International Conference on Supporting Group Work*. ACM. (2014), pp. 302–305.

[65] Jill L. Drury, Brenden Keyes, and Holly A Yanco. "LASSOing HRI: analyzing situation awareness in map-centric and video-centric interfaces". In: *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM. (2007), pp. 279–286.

[66] Curtis W Nielsen, Bob Ricks, Michael A. Goodrich, David Bruemmer, Doug Few, and M Few. "Snapshots for semantic maps". In: *IEEE International Conference on Systems, Man and Cybernetics*. Vol. 3. IEEE. (2004), pp. 2853–2858.

[67] Ludek Zalud. "ARGOS-system for heterogeneous mobile robot teleoperation". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. (2006), pp. 211–216.

[68] Bruce A. Maxwell, Nicolas Ward, and Frederick Heckel. "A human-robot interface for urban search and rescue." In: *AAAI Mobile Robot Competition* 3.01 (2003).

[69] Hajirne Nagahara, Yasushi Yagi, H. Kitamura, and M Yachida. "Super wide view tele-operation system". In: *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE. (2003), pp. 149–154.

[70] Naoji Shiroma, Noritaka Sato, Yu-huan Chiu, and Fumitoshi Matsuno. "Study on effective camera images for mobile robot teleoperation". In: *13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*. IEEE. (2004), pp. 107–112.

[71] Jim Vaughan, Sven Kratz, and Don Kimber. "Look where you're going: visual interfaces for robot teleoperation". In: *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. (2016), pp. 273–280.

[72] Hyewon Lee, Jung J. Choi, and Sonya S. Kwak. "A social agent, or a medium?: the impact of anthropomorphism of telepresence robot's sound interface on perceived copresence, telepresence and social presence". In: *Proceedings of the 7th International Workshops on the Convergent Research Society among Humanities, Sociology, Science, and Technology*. (2015), pp. 19–22.

[73] Katherine M. Tsui, Munjal Desai, Holly A. Yanco, and Chris Uhlik. "Exploring use cases for telepresence robots". In: *Proceedings of the 6th International Conference on Human-Robot interaction*. ACM. (2011), pp. 11–18.

[74] Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. "To Beam or not to Beam: A study of remote telepresence attendance at an academic conference". In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM. (2016), pp. 418–431.

[75] Nuzoo Robotics. *Ra.Ro Robot*. Available online: http://www.nuzoo.it/ (accessed on 28 November 2018).

[76] L. Giuliano, M. E. Kaouk Ng, M. L. Lupetti, and C. Germak. "Virgil, robot for museum experience: Study on the opportunity given by robot capability to integrate the actual museum visit". In: *2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*. (2015), pp. 222–223.

[77] Russell Toris, Julius Kammerl, David V. Lu, Jihoon Lee, Odest Chadwicke Jenkins, Sarah Osentoski, Mitchell Wills, and Sonia Chernova. "Robot web tools: efficient messaging for cloud robotics." In: *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. (2015), pp. 4530–4537.

[78] Sandra G. Hart and Lowell E Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research". In: *Advances in Psychology*. Vol. 52. Elsevier, (1988), pp. 139–183.

[79] Jakob Nielsen. *Usability Engineering*. Elsevier, (1994).

[80] Kate S. Hone and Robert Graham. "Towards a tool for the subjective assessment of speech system interfaces (SASSI)". In: *Natural Language Engineering* 6.3-4 (2000), pp. 287–303.

[81] Mica R Endsley. "Measurement of situation awareness in dynamic systems". In: *Human Factors* 37.1 (1995), pp. 65–84.

[82] Arnold M Lund. "Measuring usability with the use questionnaire12". In: *Usability interface* 8.2 (2001), pp. 3–6.

[83] Zak Sarris. "Survey of UAV applications in civil markets". In: *Proceedings of the 9th IEEE Mediterranean Conference on Control and Automation*. IEEE. Jan. (2001), p. 11.

[84] Hai Chen, Xin-min Wang, and Yan Li. "A survey of autonomous control for UAV". In: *International Conference on Artificial Intelligence and Computational Intelligence (AICI)*. Vol. 2. IEEE. (2009), pp. 267–271.

[85] Jessie Y. Chen, Michael J. Barnes, and Michelle Harper-Sciarini. *Supervisory control of unmanned vehicles*. Tech. rep. Army Research Lab Aberdeen Proving Ground Md Human Research and Engineering Directorate, (2010).

[86] Andrew Kopeikin, Andrew Clare, Olivier Toupet, Jonathan How, and Mary Cummings. "Flight testing a heterogeneous multi-UAV system with human supervision". In: *AIAA Guidance, Navigation, and Control Conference*. (2012), p. 4825.

[87] Bob Jacobs, Ewart De Visser, Amos Freedy, and Paul Scerri. "Application of intelligent aiding to enable single operator multiple UAV supervisory control". In: *Association for the Advancement of Artificial Intelligence, Palo Alto* (2010).

[88] PN Squire and R Parasuraman. "Effects of automation and task load on task switching during human supervision of multiple semi-autonomous robots in a dynamic environment". In: *Ergonomics* 53.8 (2010), pp. 951–961.

[89] Raymond Holsapple, John Baker, Phillip Chandler, Anouck Girard, and Meir Pachter. "Autonomous decision making with uncertainty for an urban intelligence, surveillance and reconnaissance (ISR) scenario". In: *AIAA Guidance, Navigation and Control Conference and Exhibit*. (2008), p. 6310.

[90] Kevin B. Bennett, Jeffrey D. Cress, Lawrence J. Hettinger, Dean Stautberg, and Michael W Haas. "A theoretical analysis and preliminary investigation of dynamically adaptive interfaces". In: *The International Journal of Aviation Psychology* 11.2 (2001), pp. 169–195.

[91] David B. Kaber and Jennifer M Riley. "Adaptive automation of a dynamic control task based on secondary task workload measurement". In: *International Journal of Cognitive Ergonomics* 3.3 (1999), pp. 169–187.

[92] Sarah Miller. "Workload measures". In: *National Advanced Driving Simulator. Iowa City, United States* (2001).

[93] Takahiro Hayashi and Reo Kishi. "Utilization of NASA-TLX for workload evaluation of gaze-writing systems". In: *2014 IEEE International Symposium on Multimedia (ISM)*. IEEE. (2014), pp. 271–272.

[94] Daniel Cannon and Mel Siegel. "Perceived mental workload and operator performance of dexterous manipulators under time delay with master-slave interfaces". In: *2015 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. IEEE. (2015), pp. 1–6.

[95] Susana Rubio, Eva Díaz, Jesús Martín, and José M Puente. "Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods". In: *Applied Psychology* 53.1 (2004), pp. 61–86.

[96] Scerbo Mark W, Freeman Frederick G, Mikulka Peter J, Parasuraman Raja, Nocero Francesco Di, and III Lawrence J. Prinzel. *The efficacy of psychophysiological measures for implementing adaptive technology*. Tech. rep. 2001.

[97] Glenn F. Wilson, Corrina T. Monett, and Chris A Russell. "Operator functional state classification during a simulated ATC task using EEG". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 41. 2. SAGE Publications Sage CA: Los Angeles, CA. (1997), pp. 1382–1382.

[98] Patricia Besson, Erick Dousset, Christophe Bourdin, Lionel Bringoux, Tanguy Marqueste, DR Mestre, and Jean-Louis Vercher. "Bayesian network classifiers inferring workload from physiological features: Compared performance". In: *2012 IEEE Intelligent Vehicles Symposium*. IEEE. (2012), pp. 282–287.

[99] Staffan Magnusson. "Similarities and differences in psychophysiological reactions between simulated and real air-to-ground missions". In: *The International Journal of Aviation Psychology* 12.1 (2002), pp. 49–61.

[100] Michel Besserve, Matthieu Philippe, Geneviève Florence, François Laurent, Line Garnero, and Jacques Martinerie. "Prediction of performance level during a cognitive task from ongoing EEG oscillatory activities". In: *Clinical Neurophysiology* 119.4 (2008), pp. 897–908.

[101] Glenn F. Wilson. "Real-time adaptive aiding using psychophysiological operator state assessment". In: *Engineering Psychology and Cognitive Ergonomics: Industrial Economics, HCI, and Applied Cognitive Psychology* 6 (2001), pp. 175–182.

[102] Chin-Teng Lin, Li-Wei Ko, Meng-Hsiu Chang, Jeng-Ren Duann, Jing-Ying Chen, Tung-Ping Su, and Tzyy-Ping Jung. "Review of wireless and wearable electroencephalogram systems and brain-computer interfaces–a mini-review". In: *Gerontology* 56.1 (2010), pp. 112–119.

[103] David B. Kaber, Carlene M. Perry, Noa Segall, and Mohamed A Sheik-Nainar. "Workload state classification with automation during simulated air traffic control". In: *The International Journal of Aviation Psychology* 17.4 (2007), pp. 371–390.

[104] Shouyi Wang, Jacek Gwizdka, and W Art Chaovalitwongse. "Using wireless EEG signals to assess memory workload in the *n*-back task". In: *IEEE Transactions on Human-Machine Systems* 46.3 (2016), pp. 424–435.

[105] David Sharek. *Online NASA-TLX*. Available online: http://www.nasatlx.com (accessed on 28 November 2018).

[106] V Mihajlovic and Milan Petkovic. "Dynamic bayesian networks: A state of the art". In: *University of Twente Document Repository* (2001).

[107] Carina Walter, Stephanie Schmidt, Wolfgang Rosenstiel, Peter Gerjets, and Martin Bogdan. "Using cross-task classification for classifying workload levels in complex learning tasks". In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE. (2013), pp. 876–881.

[108] Wei Lun Lim, Olga Sourina, Yisi Liu, and Lipo Wang. "EEG-based mental workload recognition related to multitasking". In: *2015 10th International Conference on Information, Communications and Signal Processing (ICICS)*. IEEE. (2015), pp. 1–4.

[109] Thomas C. Bulea, Atilla Kilicarslan, Recep Ozdemir, William H. Paloski, and Jose L. Contreras-Vidal. "Simultaneous scalp electroencephalography (EEG), electromyography (EMG), and whole-body segmental inertial recording for multimodal neural decoding". In: *Journal of Visualized Experiments: JoVE* 77 (2013).

[110] Christian Kothe. *The artifact subspace reconstruction method*. Available online: http://sccn.ucsd.edu/eeglab/plugins/asr.pdf (accessed on September 2017).

[111] Kornraphop Kawintiranon, Yanika Buatong, and Peerapon Vateekul. "Online music emotion prediction on multiple sessions of EEG data using SVM". In: *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE. (2016), pp. 1–6.

[112] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. Vol. 112. Springer, (2013).

[113] Fabrizio Lamberti, Filippo G. Pratticò, Davide Calandra, Giovanni Piumatti, Federica Bazzano, and Thiago RK Villani. "Robotic gaming and user interaction: impact of autonomous behaviors and emotional features". In: *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE. (2018), pp. 1–9.

[114] Fabrizio Lamberti, Davide Calandra, Federica Bazzano, Filippo G. Prattico, and Davide M Destefanis. "RobotQuest: A robotic game based on projected mixed reality and proximity interaction". In: *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE. (2018), pp. 1–9.

[115] Federica Bazzano, Federico Gentilini, Fabrizio Lamberti, Andrea Sanna, Gianluca Paravati, Valentina Gatteschi, and Marco Gaspardone. "Immersive virtual reality-based simulation to support the design of natural human-robot interfaces for service robotic applications". In: *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer. (2016), pp. 33–51.

[116] Federica Bazzano and Fabrizio Lamberti. "Human-robot interfaces for interactive receptionist systems and wayfinding applications". In: *Robotics* 7.3 (2018), p. 56.

[117] Entertainment Software Association. "Essential facts about the computer and video game industry". In: (2013).

[118] Nicolas Esposito. "A short and simple definition of what a videogame is". In: *Proceedings of the 2005 DiGRA International Conference: Changing Views: Worlds in Play*. (2005).

[119] Mark Stephen Tremblay, Rachel Christine Colley, Travis John Saunders, Genevieve Nissa Healy, and Neville Owen. "Physiological and health implications of a sedentary lifestyle". In: *Applied Physiology, Nutrition, and Metabolism* 35 (2010), pp. 725–740.

[120] *The International Classification of Diseases (ICD-11): Gaming Disorder*. World - Health Organization, (2018).

[121] L. Groves Christopher and Craig A. Anderson. "Negative effects of video game play". In: *Handbook of Digital Games and Entertainment Technologies* 49 (2017).

[122] Maria Luce Lupetti, Giovanni Piumatti, and Federica Rossetto. "Phygital play HRI in a new gaming scenario". In: *2015 7th International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*. IEEE. (2015), pp. 17–21.

[123] Diego Martinoia, Daniele Calandriello, and Andrea Bonarini. "Physically interactive robogames: definition and design guidelines". In: *Robotics and Autonomous Systems* 61.8 (2013), pp. 739–748.

[124] Gordon Beavers and Henry Hexmoor. "Types and limits of agent autonomy". In: *International Workshop on Computational Autonomy*. Springer. (2003), pp. 95–102.

[125] Rodney Brooks. "A robust layered control system for a mobile robot". In: *IEEE Journal on Robotics and Automation* 2.1 (1986), pp. 14–23.

[126] Thomas B. Sheridan and William L. Verplank. *Human and computer control of undersea teleoperators*. Tech. rep. Massachusetts Institute of Technology Cambridge Man-Machine Systems Lab, (1978).

[127] Ana Paiva, Iolanda Leite, and Tiago Ribeiro. "Emotion modeling for social robots". In: *Handbook of Affective Computing* (2014).

[128] Cynthia Breazeal and Juan Velásquez. "Toward teaching a robot "infant" using emotive communication acts". In: *Proceedings of the 1998 Simulated Adaptive Behavior Workshop on Socially Situated Intelligence*. Citeseer. (1998), pp. 25–40.

[129] Kimberly J. Montgomery and James V Haxby. "Mirror neuron system differentially activated by facial expressions and social hand gestures: a functional magnetic resonance imaging study". In: *Journal of Cognitive Neuroscience* 20.10 (2008), pp. 1866–1877.

[130] Joseph Bates. "The role of emotion in believable agents". In: *Communications of the ACM* 37.7 (1994), pp. 122–125.

[131] Robin Hunicke, Marc LeBlanc, and Robert Zubek. "MDA: A formal approach to game design and game research". In: *Proceedings of the AAAI Workshop on Challenges in Game AI*. Vol. 4. 1. (2004), p. 1722.

[132] Giovanni Piumatti, Filippo G. Prattico, Gianluca Paravati, and Fabrizio Lamberti. "Enabling autonomous navigation in a commercial off-the-shelf toy robot for robotic gaming". In: *2018 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE. (2018), pp. 1–6.

[133] Giovanni Piumatti, Andrea Sanna, Marco Gaspardone, and Fabrizio Lamberti. "Spatial augmented reality meets robots: human-machine interaction in cloud-based projected gaming environments". In: *2017 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE. (2017), pp. 176–179.

[134] Giovanni Piumatti, Fabrizio Lamberti, Andrea Sanna, and Paolo A Montuschi. "Robust robot tracking for next-generation collaborative robotics-based gaming environments". In: *IEEE Transactions on Emerging Topics in Computing* (2017).

[135] UAV Commercial. "Commercial drone market analysis by product (fixed wing, rotary blade, nano, hybrid), by application (agriculture, energy, government, media & entertainment) and segment forecasts to 2022". In: *Grand View Research: San Francisco, CA, USA* (2016).

[136] David Robert, Ryan Wistorrt, Jesse Gray, and Cynthia Breazeal. "Exploring mixed reality robot gaming". In: *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM. (2011), pp. 125–128.

[137] Minoru Kojima, Maki Sugimoto, Akihiro Nakamura, Masahiro Tomita, Hideaki Nii, and Masahiko Inami. "Augmented coliseum: An augmented game environment with small vehicles". In: *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*. IEEE. (2006), 6–pp.

[138] Daniel Calife, João Luiz Bernardes Jr, and Romero Tori. "Robot Arena: An augmented reality platform for game development". In: *Computers in Entertainment (CIE)* 7.1 (2009), p. 11.

[139] Byron Lahey, Winslow Burleson, Camilla Nørgaard Jensen, Natalie Freed, and Patrick Lu. "Integrating video games and robotic play in physical environments". In: *Proceedings of the 2008 ACM SIGGRAPH Symposium on Video Games*. ACM. (2008), pp. 107–114.

[140] Masanori Sugimoto. "A mobile mixed-reality environment for children's storytelling using a handheld projector and a robot". In: *IEEE Transactions on Learning Technologies* 4.3 (2011), pp. 249–260.

[141] Drew Fudenberg and Jean Tirole. *Game Theory*. Vol. 393. 12. Massachusetts Institute of Technology (MIT) Press, (1991), p. 80.

[142] Rodney Allen Brooks. *Cambrian Intelligence: The Early History of the New AI*. Vol. 97. MIT press Cambridge, MA, (1999).

[143] Jessica R. Cauchard, Jane L. E, Kevin Y. Zhai, and James A. Landay. "Drone & me: an exploration into natural human-drone interaction". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. (2015), pp. 361–365.

[144] Felipe Andres Cid, Luis J. Manso, Luis V. Calderita, Agustín Sánchez, and Pedro Nuñez. "Engaging human-to-robot attention using conversational gestures and lip-synchronization". In: *Journal of Physical Agents* 6.1 (2012), pp. 3–10.

[145] Brandon Heenan, Saul Greenberg, Setareh Aghel-Manesh, and Ehud Sharlin. "Designing social greetings in human robot interaction". In: *Proceedings of the 2014 Conference on Designing Interactive Systems*. ACM. (2014), pp. 855–864.

[146] Daniel Szafir, Bilge Mutlu, and Terry Fong. "Communicating directionality in flying robots". In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM. (2015), pp. 19–26.

[147] Wijnand A. IJsselsteijn, Yvonne A.W. de Kort, and Karolien Poels. *The Game Experience Questionnaire*. Technische Universiteit Eindhoven, (2013).

[148] John Brooke. "SUS-A quick and dirty usability scale". In: *Usability Evaluation in Industry* 189.194 (1995), pp. 4–7.

[149] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". In: *International Journal of Social Robotics* 1.1 (2009), pp. 71–81.

[150] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. "Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction". In: *International Journal of Human-Computer Studies* 64.10 (2006), pp. 962–973.

[151] Pratibha Adkar. "Unimodal and multimodal human computer interaction: a modern overview". In: *International Journal of Computer Science Engineering and Technology (IJCSET)* 2.3 (2013), pp. 1–8.

[152] Sharon Oviatt. "Multimodal interfaces". In: *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* 14 (2003), pp. 286–304.

[153] Thomas Kollar, Anu Vedantham, Corey Sobel, Cory Chang, Vittorio Perera, and Manuela Veloso. "A multi-modal approach for natural human-robot interaction". In: *International Conference on Social Robotics*. Springer. (2012), pp. 458–467.

[154] Marjorie Skubic, Dennis Perzanowski, Samuel Blisard, Alan Schultz, William Adams, Magda Bugajska, and Derek Brock. "Spatial language for human-robot dialogs". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34.2 (2004), pp. 154–167.

[155] Rainer Stiefelhagen, C. Fugen, R. Gieselmann, Hartwig Holzapfel, Kai Nickel, and Alex Waibel. "Natural human-robot interaction using speech, head pose and gestures". In: *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Vol. 3. IEEE. (2004), pp. 2422–2427.

[156] Rainer Bischoff and Volker Graefe. "Dependable multimodal communication and interaction with robotic assistants". In: *Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication*. IEEE. (2002), pp. 300–305.

[157] Björn Giesler, Tobias Salb, Peter Steinhaus, and Rüdiger Dillmann. "Using augmented reality to interact with an autonomous mobile platform". In: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. Vol. 1. IEEE. (2004), pp. 1009–1014.

[158] Tathagata Chakraborti, Sarath Sreedharan, Anagha Kulkarni, and Subbarao Kambhampati. "Alternative modes of interaction in proximal human-in-the-loop operation of robots". In: *Computing Research Repository (CoRR)* (2017).

[159] Jacob W. Crandall, Michael A. Goodrich, Curtis W. Nielsen, and Dan R. Olsen Jr. "Validating human-robot interaction schemes in multitasking environments". In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 35.4 (2005), pp. 438–449.

[160] Thomas B. Sheridan. *Humans and Automation: System Design and Research Issues*. John Wiley & Sons, Inc., (2002).

[161] Mica Endsley, Betty Bolte, and Debra Jones. *Designing for Situation Awareness: An Approach to User-Centered Design*. CRC press, Jan. (2003).

[162] Joan H. Johnston, Stephen M. Fiore, Carol Paris, and C. A. P. Smith. "Application of Cognitive Load Theory to Develop a Measure of Team Cognitive Efficiency". In: *Military Psychology* 25.3 (2013), pp. 252–265.

[163] Gary Klein, Paul J. Feltovich, Jeffrey Bradshaw, and David Woods. *Common Ground and Coordination in Joint Activity*. June 2005.

[164] Brian P. Gerkey, Richard Vaughan, and Andrew Howard. "The Player/Stage project: tools for multi-robot and distributed sensor systems". In: *Proceedings of the International Conference on Advanced Robotics* (Aug. 2003).

[165] Kerstin Eklundh, Anders Green, and Helge Hüttenrauch. "Social and collaborative aspects of interaction with a service robot". In: *Robotics and Autonomous Systems* 42 (Mar. 2003), pp. 223–234.

[166] Jesus Savage-Carmona, Mark Billinghurst, and Alistair Holden. "The VirBot: A virtual reality robot driven with multimodal commands". In: *Expert Systems with Applications* 15 (Oct. 1998), pp. 413–419.

[167] Gilberto Echeverria, Séverin Lemaignan, Arnaud Degroote, Simon Lacroix, Michael Karg, Pierrick Koch, Charles Lesire, and Serge Stinckwich. "Simulating Complex Robotic Scenarios with MORSE". In: *Simulation, Modeling, and Programming for Autonomous Robots*. Vol. 7628. Springer Berlin Heidelberg, (2012), pp. 197–208.

[168] James R Lewis. "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use". In: *International Journal of Human-Computer Interaction* 7.1 (1995), pp. 57–78.

[169] Takashi Minato, Michihiro Shimada, Hiroshi Ishiguro, and Shoji Itakura. "Development of an android robot for studying human-robot interaction". In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer. (2004), pp. 424–434.

[170] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. "A survey of socially interactive robots". In: *Robotics and Autonomous Systems* 42.3-4 (2003), pp. 143–166.

[171] David Feil-Seifer and Maja J Mataric. "Defining socially assistive robotics". In: *9th International Conference on Rehabilitation Robotics (ICORR)*. IEEE. (2005), pp. 465–468.

[172] Rudolph P. Darken and Barry Peterson. "Spatial orientation, wayfinding, and representation". In: *In K. M. Stanney (Ed.), Handbook of Virtual Environments: Design, Implementation, and Applications*. Erlbaum, (2001), pp. 493–518.

[173] Marvin Levine. "You-are-here maps: Psychological considerations". In: *Environment and Behavior* 14.2 (1982), pp. 221–237.

[174] Daniel R. Montello and Corina Sas. "Human factors of wayfinding in navigation". In: *International Encyclopedia of Ergonomics and Human Factors* (2006).

[175] Toru Ishikawa, Hiromichi Fujiwara, Osamu Imai, and Atsuyuki Okabe. "Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience". In: *Journal of Environmental Psychology* 28.1 (2008), pp. 74–82.

[176] Lynn A. Streeter, Diane Vitello, and Susan A Wonsiewicz. "How to tell people where to go: Comparing navigational aids". In: *International Journal of Man-Machine Studies* 22.5 (1985), pp. 549–562.

[177] Volker Coors, Christian Elting, Christian Kray, and Katri Laakso. "Presenting route instructions on mobile devices: From textual directions to 3D visualization". In: *Exploring Geovisualization*. Elsevier, (2005), pp. 529–550.

[178] Alexandra Lorenz, Cornelia Thierbach, Nina Baur, and Thomas H Kolbe. "App-free zone: Paper maps as alternative to electronic indoor navigation aids and their empirical evaluation with large user bases". In: *Progress in Location-Based Services*. Springer, (2013), pp. 319–338.

[179] Julie Dillemuth. "Map design evaluation for mobile display". In: *Cartography and Geographic Information Science* 32.4 (2005), pp. 285–301.

[180] Rodney Fewings. "Wayfinding and airport terminal design". In: *The Journal of Navigation* 54.2 (2001), pp. 177–184.

[181] Roope Raisamo. "A multimodal user interface for public information kiosks". In: *Proceedings of the Workshop on Perceptual User Interfaces*. (1998), pp. 7–12.

[182] Simon Bergweiler, Matthieu Deru, and Daniel Porta. "Integrating a multitouch kiosk system with mobile devices and multimodal interaction". In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ACM. (2010), pp. 245–246.

[183] Andreea I. Niculescu, Kheng Hui Yeo, and Rafael Enrique Banchs. "Designing MUSE: A multimodal user experience for a shopping mall kiosk". In: *Proceedings of the International Conference on Human Agent Interaction*. ACM. (2016), pp. 273–275.

[184] Hakan Tüzün, Esra Telli, and Arman Alır. "Usability testing of a 3D touch screen kiosk system for way-finding". In: *Computers in Human Behavior* 61 (2016), pp. 73–79.

[185] Patricia Wright, Anthony Soroka, Steve Belt, Duc T. Pham, Stefan Dimov, David De Roure, and Helen Petrie. "Using audio to support animated route information in a hospital touch-screen kiosk". In: *Computers in Human Behavior* 26.4 (2010), pp. 753–759.

[186] Vasileios Toutziaris. "Usability of an adjusted IndoorTubes map design for indoor wayfinding on mobile devices". MA thesis.

[187] *HERE WeGo*. Available online: http://wego.here.com (accessed on 10 April 2018).

[188] *Cartogram*. Available online: http://www.cartogram.com (accessed on 10 April 2018).

[189] *Google Maps*. Available online: http://www.google.com/maps/ (accessed on 10 April 2018).

[190] *Mapwize*. Available online: http://www.mapwize.io/en/ (accessed on 10 April 2018).

[191] *Wayfinding kiosk with 3D wayfinder*. Available online: http://3dwayfinder.com (accessed on 10 April 2018).

[192] *Kiosk wayfinder*. Available online: http://www.kioskwebsite.com (accessed on 10 April 2018).

[193] Tomoko Koda and Pattie Maes. "Agents with faces: The effect of personification". In: *Proceedings of the IEEE International Workshop on Robot and Human Communication*. IEEE. (1996), pp. 189–194.

[194] Justine Cassell, Tim Bickmore, Lee Campbell, Hannes Vilhjalmsson, and Hao Yan. "More than just a pretty face: conversational protocols and the affordances of embodiment". In: *Knowledge-based Systems* 14.1-2 (2001), pp. 55–64.

[195] Justine Cassell. *Embodied Conversational Agents*. MIT press, (2000).

[196] Mariët Theune, Dennis Hofs, and Marco van Kessel. "The Virtual Guide: A direction giving embodied conversational agent". In: *Proceedings of the Annual Conference of the International Speech Communication Association*. (2007).

[197] Justine Cassell, Tom Stocky, Tim Bickmore, Yang Gao, Yukiko Nakano, Kimiko Ryokai, Dona Tversky, Catherine Vaucelle, and Hannes Vilhjálmsson. "Mack: Media lab autonomous conversational kiosk". In: *Proceedings of the Imagina: Intelligent Autonomous Agents*. Vol. 2. (2002), pp. 12–15.

[198] Thomas August Stocky. "Conveying routes: Multimodal generation and spatial intelligence in embodied conversational agents". PhD thesis. Massachusetts Institute of Technology, (2002).

[199] Yukiko I. Nakano, Gabe Reinstein, Tom Stocky, and Justine Cassell. "Towards a model of face-to-face grounding". In: *Proceedings of the Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics. (2003), pp. 553–561.

[200] Stefan Kopp, Paul A. Tepper, Kimberley Ferriman, Kristina Striegnitz, and Justine Cassell. "Trading spaces: How humans and humanoids use speech and gesture to give directions". In: *Conversational Informatics: An Engineering Approach* (2007), pp. 133–160.

[201] Sabarish Babu, Stephen Schmugge, Tiffany Barnes, and Larry F Hodges. ""What would you like to talk about?" An evaluation of social conversations with a virtual receptionist". In: *International Workshop on Intelligent Virtual Agents*. Springer. (2006), pp. 169–180.

[202] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, and Jue Wang. "Designing robots for long-term social interaction". In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. (2005), pp. 1338–1343.

[203] Marek P. Michalowski, Selma Sabanovic, and Reid Simmons. "A spatial model of engagement for a social robot". In: *Proceedings of the IEEE International Workshop on Advanced Motion Control*. IEEE. (2006), pp. 762–767.

[204] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. "Receptionist or information kiosk: how do people talk with a robot?" In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. ACM. (2010), pp. 31–40.

[205] Lauren Cairco, Dale-Marie Wilson, Vicky Fowler, and Morris LeBlanc. "AVARI: Animated virtual agent retrieving information". In: *Proceedings of the Annual Southeast Regional Conference*. ACM. (2009), p. 16.

[206] Andrea L. Thomaz and Crystal Chao. "Turn-taking based on information flow for fluent human-robot interaction". In: *AI Magazine* 32.4 (2011), pp. 53–63.

[207] Maha Salem, Micheline Ziadee, and Majd Sakr. "Effects of politeness and interaction context on perception and experience of HRI". In: *Proceedings of the International Conference on Social Robotics*. Springer. (2013), pp. 531–541.

[208] Talha Rehmani, Sabur Butt, Inam-ur-Rehman Baig, Mohammad Zubair Malik, and Mohsen Ali. "Designing robot receptionist for overcoming poor infrastructure, low literacy and low rate of female interaction". In: *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM. (2018), pp. 211–212.

[209] Ryuichi Nisimura, Takashi Uchida, Akinobu Lee, Hiroshi Saruwatari, Kiyohiro Shikano, and Yoshio Matsumoto. "ASKA: Receptionist robot with speech dialogue system". In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vol. 2. IEEE. (2002), pp. 1314–1319.

[210] Hartwig Holzapfel and Alex Waibel. "Behavior models for learning and receptionist dialogs". In: *Proceedings of the Annual Conference of the International Speech Communication Association*. (2007).

[211] Akansel Cosgun, Alexander JB Trevor, and Henrik I Christensen. "Did you mean this object?: detecting ambiguity in pointing gesture targets". In: *10th ACM/IEEE International Conference on Human-Robot Interaction (HRI) Workshop on Towards a Framework for Joint Action*. (2015).

[212] Yasuhiko Hato, Satoru Satake, Takayuki Kanda, Michita Imai, and Norihiro Hagita. "Pointing to space: modeling of deictic interaction referring to regions". In: *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*. IEEE Press. (2010), pp. 301–308.

[213] Catherina Burghart, Hartwig Holzapfel, Roger Haeussling, and Stephan Breuer. "Coding interaction patterns between human and receptionist robot". In: *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*. IEEE. (2007), pp. 454–460.

[214] Takuya Hashimoto, Sachio Hiramatsu, Toshiaki Tsuji, and Hiroshi Kobayashi. "Realization and evaluation of realistic nod with receptionist robot SAYA". In: *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*. IEEE. (2007), pp. 326–331.

[215] Yusuke Okuno, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. "Providing route directions: Design of robot's utterance, gesture, and timing". In: *ACM/IEEE International Conference on Human-Robot Interaction*. IEEE. (2009), pp. 53–60.

[216] Dan Bohus, Chit W Saw, and Eric Horvitz. "Directions robot: In-the-wild experiences and lessons learned". In: *Proceedings of the International Conference on Autonomous Agents and Multi-agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems. (2014), pp. 637–644.

[217] Patrick Holthaus and Sven Wachsmuth. "The receptionist robot". In: *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM. (2014), pp. 329–329.

[218] Gabriele Trovato, Josué G. Ramos, Helio Azevedo, Artemis Moroni, Silvia Magossi, Hiroyuki Ishii, Reid Simmons, and Atsuo Takanishi. ""Olá, my name is Ana": A study on brazilians interacting with a receptionist robot". In: *2015 International Conference on Advanced Robotics (ICAR)*. IEEE. (2015), pp. 66–71.

[219] Gabriele Trovato, JG Ramos, Helio Azevedo, Artemis Moroni, Silvia Magossi, Hiroyuki Ishii, R. Simmons, and Atsuo Takanishi. "Designing a receptionist robot: Effect of voice and appearance on anthropomorphism". In: *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. (2015), pp. 235–240.

[220] Dai Hasegawa, Justine Cassell, and Kenji Araki. "The role of embodiment and perspective in direction-giving systems." In: *AAAI Fall Symposium: Dialog with Robots*. (2010).

[221] G. Langevin. *InMoov - Open source 3D printed life-size robot.* Available online: http://www.inmoov.fr (accessed on 10 April 2018).

[222] Wilma A. Bainbridge, Justin W Hart, Elizabeth S. Kim, and Brian Scassellati. "The benefits of interactions with physically present robots over video-displayed agents". In: *International Journal of Social Robotics* 3.1 (2011), pp. 41–52.

[223] Jamy Li. "The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents". In: *International Journal of Human-Computer Studies* 77 (2015), pp. 23–37.

[224] Roy S Kalawsky. "VRUSE—a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems". In: *Applied Ergonomics* 30.1 (1999), pp. 11–25.

[225] Juan Fasola and M Mataric. "Comparing physical and virtual embodiment in a socially assistive robot exercise coach for the elderly". In: *Center for Robotics and Embedded Systems, Los Angeles, CA* (2011).

[226] Robert H. Poresky, Charles Hendrix, Jacob E. Mosier, and Marvin L. Samuelson. "The companion animal bonding scale: Internal reliability and construct validity". In: *Psychological Reports* 60.3 (1987), pp. 743–746.

[227] James C. McCroskey and Thomas A. McCain. "The measurement of interpersonal attraction". In: *Speech Monographs* 41 (1974), pp. 261–266.

[228] Arnaud Delorme and Scott Makeig. "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis". In: *Journal of Neuroscience Methods* 134.1 (2004), pp. 9–21.