UNIVERSITY OF GENOVA

PHD PROGRAM IN BIOENGINEERING AND ROBOTICS

Mixed Reality: an Efficient Communication Channel for Human-Robot Collaboration

by

Simone Macciò

Thesis submitted for the degree of *Doctor of Philosophy* (36° cycle)

April 2024

Dr. Alessandro Carfì	Supervisor
Prof. Fulvio Mastrogiovanni	Supervisor
Prof. Paolo Massobrio	Head of the PhD program

Thesis Jury:

Prof. Antonio Chella, *Università degli Studi di Palermo*	External examiner
Prof. Dimitri Ognibene, *Università degli Studi Milano-Bicocca*	External examiner
Prof. Antonio Sgorbissa, *Università degli Studi di Genova*	Internal examiner

Dibris

Department of Informatics, Bioengineering, Robotics and Systems Engineering

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Simone Macciò
April 2024

Abstract

Collaborative robots represent a technological leap forward, and their adoption could benefit many small and medium-sized enterprises (SMEs). Such robots are cost-effective and allow humans for safe, close-proximity, and highly flexible interactions with the machine. Nonetheless, industrial collaborative robots nowadays lack a key requirement for efficient collaboration, namely the possibility to effectively communicate with human teammates. To tackle this open and challenging aspect in collaborative robotics, the present Ph.D. work has drawn inspiration from social studies on human-human collaboration, where other researchers have demonstrated how efficient interaction is achieved through implicit communication, made up of a series of cues (e.g., gaze, gestures, etc.), which lead individuals to convey their own intentions and infer their teammate's ones dynamically.

Building on this principle, this Ph.D. project's objective has been attempting to bridge such a communication gap by developing novel interfaces to enable a more intuitive, seamless interaction between humans and robots and to endow the latter with the ability to project their intentions, defined as upcoming planned actions, in a straightforward way. To achieve such a result, various communication alternatives have been evaluated and eventually Mixed Reality has been chosen and thoroughly explored as a suitable channel for building an efficient and intuitive human-robot communication layer. To this extent, a novel robot system architecture has been developed and refined throughout the three years, integrating Mixed Reality with modern and powerful Head-Mounted Display devices. Such architecture brings forth a comprehensive bi-directional, holographic communication interface which can be employed in various collaborative scenarios.

On the one hand, robot-to-human communication enables projecting robot's intentions as holographic, visual cues in a direct way to the human teammate. Specifically, a virtual counterpart of the robot can be superimposed to the real one in the Mixed Reality layer and used to anticipate upcoming robot's actions via dynamic, holographic animations, potentially offering useful insights and improving human teammate's awareness throughout the collaborative process. The proposed interface has been tested in multiple user studies under different collaborative contexts, including assembly tasks with fixed robot manipulators and scenarios of mobile collaboration. The results have highlighted that such form of holographic communication ensures a smoother collaboration process, where human and robot are less likely to obstruct and hinder each other, due to the improved awareness of the human, while at the same time increasing the rate of success of joint actions (e.g., handovers).

On the other hand, human-to-robot communication can be used to ensure a more direct interaction and to easily control and teach tasks to the robotic teammate. In particular, by interacting with the holographic robot in the Mixed Reality layer through a combination of voice and gestures, the human teammate can intuitively achieve Kinesthetic Teaching and teach both simple motions and complex pick-and-place or handover tasks to the robot.

Overall, the result of this Ph.D. work is an open-source, modular architecture which can be employed by other researchers and companies to take advantage of the proposed holographic communication scheme, both in industrial collaborative contexts and in more social scenarios of human-robot interaction.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation

With the coming of the Fourth Industrial Revolution, a new human-centered paradigm started emerging in the manufacturing industry, in which the role and well-being of operators and technicians are prioritized [75]. Such a paradigmatic shift, in combination with technological advancements in the fields of multi-modal perception, control and actuation [139], fostered the development of a new generation of robots, designed not to replace human operators but rather to coexist and collaborate with them (hence the name *cobots*) in shared, possibly unconstrained, work-spaces [2, 5]. The adoption of cobots in industrial environments has shown positive effects in terms of productivity [56, 45], with hybrid human-robot teams combining the benefits of automation, i.e., repeatability, speed, and precision, with human cognitive skills [163, 107]. Furthermore, such collaborative platforms are particularly appreciated and increasingly adopted by Small and Medium Enterprises (SMEs) [118], given their limited costs, high versatility and the possibility to be easily repurposed for different tasks. As a result, Human-Robot Collaboration (HRC) has quickly and recently grown in importance as a research field, both from scientific and industrial standpoints.

Nevertheless, the new HRC paradigm brings about numerous technical challenges, which range from ensuring human operator's safety throughout the collaboration process, to developing efficient communication interfaces for hybrid human-robot teams. On the one hand, multiple works have explored the aspect of human safety in HRC, proposing either solutions that attempt to reduce collision risk, by predicting human space occupancy [130, 176] or approaches that instead focus on detecting and mitigating collisions between agents [134, 138, 103]. Conversely, the communication aspect in HRC still represents an open research problem. An adequate level of collaboration can be achieved only through an effective communication between interacting parties. To this extent, it is fundamental for the human operator to be aware of, and possibly even anticipate, the actions of their robotic teammate. This is of the utmost importance to minimize the cognitive load entailed by the collaboration process and to make sure the human teammate is not caught off-guard by sudden robot movements. Poor communication could in fact lead to undesired circumstances hindering the quality of the collaboration, including accidental collisions between agents, or failures in completing joint actions, e.g., handovers. Such unfavorable outcomes could, in turn, undermine the human trust in the robot [175], possibly increasing their stress if the robot is perceived as useless rather than supportive [91]. Despite the importance of such a research aspect, the literature in HRC still lacks structured frameworks enabling intuitive, efficient and straightforward communication between humans and robots.

Traditionally, communication channels adopted in HRC involve either speech [162, 92], arm and hand gestures [28, 116, 16] or screen-mediated solutions [148, 76]. However, explicit communication is not fully representative of how humans exchange information and coordinate when involved in collaborative activities. As a matter of fact, several social studies conducted in human-human collaboration scenarios [80, 115, 23, 72, 142] have highlighted how individuals tend to rely also on a set of non-verbal cues, resulting from implicit communication, involving gaze and body posture. An example of such behavior is the natural tendency to direct our gaze at an object before grasping it. These implicit yet involuntary signals lead individuals to intuitively understand and predict each other's intentions, favoring synchronization and coordination between agents and enabling them to efficiently complete the collaborative task.

Following such principle, a more natural and seamless interaction between humans and robots could be achieved by means of effective communication interfaces that allow agents to convey and exchange their intentions in a straightforward way. My Ph.D. thesis falls exactly within this context. Drawing inspiration from human behavior, the novelty of this work consists in the development of a structured framework for intuitive and efficient human-robot communication, with a focus on industrial HRC contexts. The fulfillment of such an objective encompassed several challenges, which can be summed up in the following key points: i) identifying suitable channels for effective communication in HRC; ii) modeling and representing agents' intentions through analytical formalization; iii) developing a generalized software architecture able to translate such formalization into empirical cues to be conveyed via the selected channels during HRC processes.

Throughout the chapters of this thesis, I will detail how the aforementioned challenges have been addressed and present the various contributions brought to the state-of-the art, discussing potential applications in real-world industrial scenarios, as well as the limitations and possible future directions for this research.

1.2 Research Overview

Considering the increasing interest that HRC has gained in recent years as research field, numerous works started exploring how communication impacts interaction between humans and robots, and what channels can be leveraged to allow agents to efficiently exchange information and coordinate while working together on a shared task. To this regard, an extensive review of previous, related studies is provided in Chapter 2.

This thesis attempts to provide additional contributions to this field, addressing challenging aspects such as enabling agents to exchange and convey intention cues through intuitive communication during a collaborative process, in order to achieve a more coordinated, fluent interaction. This paragraph provides an overview of my research path.

The first step of my work consisted in exploring communication strategies adopted in literature in HRC contexts, evaluating their pros and cons in light of how they impact collaboration. Generally speaking, collaborative robot platforms lack humanoid characteristics and it is therefore extremely difficult to replicate the distinctive features (e.g., gaze) which make up implicit communication in humans. At the same time, designing algorithms to capture such implicit signals in human behavior is a hard task, given the imperceptible nature of these cues. As a result, the majority of communicative interfaces proposed over the years in HRC takes advantage of explicit communication only, allowing agents to exchange information either through vocal [86, 14] or gestural interaction [93, 49], or even via a combination of the two [71, 39]. However, most of these strategies require agents to pause their tasks to explicitly perform a communication act to their teammate, reducing the overall efficiency of the team. Furthermore, as discussed in more detail in Chapter 2, voice and gesture-mediated communications may suffer several environmental limitations in industrial settings or require precisely calibrated tools and hardware, aspects which could negatively affect their practical implementation in real-world manufacturing scenarios.

To address the aforementioned limitations, research trends recently started veering towards solutions based on Extended Reality (XR) [26], mainly through Augmented Reality (AR) and Mixed Reality (MR). These technologies are particularly captivating as they grant the possibility to create engaging experiences, where digital content is overlaid to the real world, offering a whole new layer for interaction between humans and robots without requiring complex external infrastructure and calibration. On the one hand, AR has been extensively employed to build communication interfaces in HRC, in combination with hand-held devices like smartphones or tablets [54, 110, 122, 32, 38]. These approaches take advantage of the virtual layer to convey information about robot's planned motions and actions to the human operator, projecting digital content directly on top of the device's camera feed. The possibility to visualize in advance, throughout the collaboration, robot's intention cues as digital overlays represented a first step towards exploiting the potential of implicit communication in HRC. Nevertheless, such form of intuitive communication comes at the expense of an efficient interaction between agents, as the human operator is always required to carry the device around, as well as needs to continuously pause their task to inspect the robot's actions on the screen, effectively negating the benefits introduced by implicit communication in the first place.

On the other hand, MR can be thought of as the natural evolution of AR, brought forth by the commercialization of Head-Mounted Display (HMD) devices, such as the *Microsoft HoloLens* family. Unlike AR, which is dependent on hand-held devices, MR relies on these compact, wearable and affordable visors, which offer unmatched possibilities, enabling users to experience 3D digital content directly in their first-person view, while at the same time letting them keep their hands free to interact with the robot teammate. In this context, several studies began exploiting such MR layer to build communicative frameworks for HRC, where robots can, during collaborative activities, convey their motions in advance via holographic representations, providing useful information to the human teammate without pausing their task to explicitly perform a communication act [137, 117]. Although promising in terms of the communicative capabilities they offer, these approaches convey scattered and incomplete cues about the robot's intentions, as the information being communicated lacks a structured formalization. This results in communication acts which are only partially expressive of the robot's internal state and may not appear informative enough to the human operator. Furthermore, they only deal with one-directional communication of intent, namely from the robot to the human teammate. As such, they only exploit partially MR as layer for efficient communication in HRC.

Taking these limitations into account, during this Ph.D. work I chose to concentrate my efforts on expanding such state-of-the-art. I attempted to build a thorough, *bi-directional* communication framework based primarily on MR, but also leveraging other emerging technologies, such as Digital Twins (DTs) [60], in which the virtual layer ensures a more intuitive and simplified interaction between human and robot, as well as where agents' intentions can be conveyed through expressive cues during collaborative tasks, with the final aim of improving team efficiency and teamwork.

The first contribution brought to the field has therefore been the development of an analytical framework that models communication acts in generalized HRC contexts. A similar formalism did not previously exist in literature, although it is fundamental to effectively describe expressive, meaningful communication, aimed at exchanging pieces of information in a clear and unambiguous way. The proposed framework, which I called *Communication Space* (*C-Space* for brevity), provides a methodology to do exactly so. It accounts for the various communicative channels (e.g., gaze, voice, gestures, AR/MR) which exist in a particular instance of HRC, and enables modeling how such channels contribute to the overall communication, modulating and conveying pieces of information. Given its comprehensive and generalized nature, *C-Space* can be leveraged in various domains of HRC to explicitly formalize communicative acts, offering a valuable reference for when communication has to be translated from mere theoretical representation to practical, effective implementation at software-level.

With the introduction of *C-Space*, I was able to tackle the challenge of representing and conveying robots' intentions to human teammates using holographic cues. As such, during the first part of my Ph.D. I focused on the Robot-To-Human (RTH) communication aspect, expanding the aforementioned state-of-the-art. Since my work was focused on building a MR-based communication layer, first I introduced *MR-Space*, a practical instance of *C-Space* tailored specifically for holographic communication. Then, using the formalism of *MR-Space*, I proceeded to provide a wide-ranging, analytical representation of robots' intentions, which is deeply detailed in Chapter 3. Once the modeling phase was completed, the next step in the work consisted in translating such theoretical representation into an empirical software architecture, enabling robots to project their upcoming intentions as holographic cues during collaborative tasks. Specifically, the architecture made possible to spawn a digital counterpart of the robot in the MR layer, which was used to preview upcoming robot's intentions as holographic animations, before the corresponding actions were effectively executed by the real robot agent. A first version of the envisioned architecture, named *MR-HRC-V1*, has been implemented using the popular Unity game engine and interfaced with the Robot Operating System (ROS) [128] middleware. The resulting application has been deployed on a Microsoft HoloLens 2 HMD device and the holographic communication scheme evaluated through two user studies, a preliminary one first followed by an extensive one, carried out in a scenario of collaborative assembly between a human operator and a fixed manipulator robot, namely Baxter [53] from Rethink Robotics. The results, evaluated in terms of team efficiency and User Experience (UX) of the proposed MR interface, highlighted the beneficial effect introduced by the holographic communication of robot's intentions during the collaboration.

In particular, I found overall evidence of improved team coordination during the interaction, with less involuntary collisions occurring between agents, and an increased rate of joint actions (e.g., handovers) completed successfully.

In light of these results, which suggested a positive impact of MR-based communication on HRC in fixed work spaces, the next step in the research involved attempting to generalize the framework to scenarios where humans and robots are not fixed to their workstations. This is particularly relevant in settings like logistics, where agents move around the environment and are not only required to interact but also to simultaneously carry out independent, concurrent activities. Therefore, a subsequent study has been undertaken, starting with the design of an experimental, logistics-like collaborative environment, followed by an updated version of the software architecture, to account for the holographic representation of navigation intentions in mobile robots, according to the formalization of *MR-Space*. Furthermore, this work also acted as first attempt at building a comprehensive, bi-directional communication layer, where not only the RTH aspect was addressed trough holographic cues, but also Human-To-Robot (HTR) communication was accounted for. In particular, the HTR aspect was tackled through the introduction of a DT in the loop, complementing the MR-based architecture. The DT acted as virtual, simulated replica of the collaborative environment, maintaining an updated, online representation of the agents' state during the interaction. Such digital model was then employed to continuously monitor the human throughout the collaboration process, extrapolating potential intention cues from a combination of their implicit signals (i.e., gaze and posture), and consequently triggering certain robot's logic if specific implicit communication patterns were identified in the human's behavior. In other words, the introduction of the DT contributed to the envisioned bi-directional communication layer, providing a tool to analyze humans' implicit cues on the fly, transmitting potential relevant information to the robotic teammate. The overall architecture resulting from this work was renamed *MR-HRC-V2*, in accordance with the previous naming. Coherently with the preceding research, the new architecture and the overall communication framework were tested through user study in the experimental, logistics-like scenario, where a human operator was required to complete a collaborative, warehouse-like task with the robot Tiago++ [121] from Pal Robotics, a well-known mobile manipulator platform. The findings observed in the study were consistent with previous ones, suggesting how MR-based communication improved team efficiency and synchronization between agents. Additionally, the DT proved effective in conveying human's implicit cues to the robot, positively contributing to a more natural, seamless interaction between agents.

Once proving that the MR-based communication framework was an effective mean towards more efficient collaboration, the second part of my Ph.D. journey began with the generalization of the architecture developed so far, in order to build an open-source, modular implementation that could be easily adopted, utilized and possibly extended by other researchers and companies working in the field of HRC. This implied a complete overhaul of the software architecture, to make it independent of the particular HRC context, also extending the implementation to support collaborative scenarios where multiple users and robots are involved at the same time. To this extent, the software architecture has been rebuilt from the ground up, using Unreal Engine 4 (UE4) for the development of the revamped holographic interface. In particular, the adoption of UE4 made possible to exploit its state-of-the-art physics engine, thus enabling the creation of more immersive MR experiences, where holographic entities interact with the real world in a realistic way. This feature, in turn, greatly expanded the communicative capabilities offered by the MR channel, providing more realistic representations of the robot's intentions to the user throughout HRC processes. Alongside UE4, the architecture has been re-designed to accommodate Apache Kafka as main data exchange infrastructure. Given its widespread adoption in industry and the existence of many frameworks that interface with it, Kafka represented an added value to the new version of the architecture, opening up the possibility to integrate the MR-based communication layer in real-world settings. Furthermore, Kafka provided a straightforward interface to support integration with external DTs that could be employed for a generalized communication framework, while also offering scalability, reliability and fault-tolerance features, given its cloud-based nature. This, new, revisited architecture has been named Robot Intent Communication through Mixed Reality (*RICO-MR*) and has been made open-source for all researchers and practitioners interested in adopting the proposed holographic communication scheme.

With a robust architecture at hand, and having focused, for the most part, on intuitive, holographic RTH communication, the last step of my Ph.D. work involved investigating more deeply MR-based HTR communication. Unlike the previous study, where the HTR aspect was tackled and solved using a DT, the present work aimed at exclusively exploiting the *MR-Space* formalism to develop a holographic-based communication scheme, where the virtual layer was used by the human to convey information to the robot teammate. In particular, I concentrated the efforts on designing a novel approach at Kinesthetic Teaching (KT), a popular technique used in HRC settings that allows human operators to teach motions and trajectories to robots via physical interaction. To this extent, KT has been framed into the communicative formalism of the *MR-Space* and an extended version of the *RICO-MR* architecture has been developed, enabling KT in the MR layer, with the human operator teaching motions and actions to a holographic counterpart of the robot through an intuitive combination of vocal and gestural interaction. The main potential of this novel approach resided in the possibility of executing KT on any robotic platform, even on those robots that natively do not possess the hardware/software components necessary to support such teaching feature. The holographic-based KT has been tested in a user study, involving multiple robot models for more generalized results, and compared with traditional, hand-guided KT. The study's findings have demonstrated that the novel, holographic KT approach behaved comparably, in terms of teaching effectiveness, to the traditional one, as well as offered similar UX results.

To conclude, in this thesis I show that a more seamless, natural and efficient collaboration in industrial settings is possible through an intuitive, meaningful communication layer. MR has proven a promising channel to design engaging and informative interfaces, opening up the possibility to superimpose expressive, holographic cues to the real scene, while at the same time enabling a more straightforward interaction between agents through digital overlays. Although the HMD devices needed to experience such extended realities still suffer from several technical limitations (e.g. limited field of view of holographic projections), this technology has increasingly grown in popularity and interest in recent years, given the little hardware required, which makes it most suitable for unstructured collaborative environments.

1.3 Structure

The thesis is composed of 8 chapters and is structured as follows:

This was **Chapter 1**, where I introduced the rationale and the overview of my Ph.D. work.

Chapter 2 provides a general background of the research, starting with notable approaches that attempted to address the communicative gap in human-robot teams, both in terms of RTH and HTR communication (see Section 2.1). Then, a thorough overview of XR-based strategies is presented in Section 2.2, illustrating how AR and MR have been leveraged in literature to build effective communication layers for HRC, taking into account the corresponding limitations which are addressed and overcome with this Ph.D. work. Subsequently, Section 2.3 reports a brief overview of DTs in HRC, whereas Section 2.4 discusses methods to quantitatively evaluate human-robot teams.

Chapter 3 presents the theoretical formalization of the *C-Space* and the corresponding instance of *MR-Space*, tailored specifically for holographic-based communication. Such formalization is then leveraged to describe and analytically represent robots' intentions, and later this modeling is translated into a practical MR-based software architecture. The value of holographic RTH communication is finally put to test in a preliminary user study.

Chapter 4 details an extensive user study carried out in a collaborative assembly scenario to corroborate and generalize the preliminary findings discussed in **Chapter 3**. The proposed RTH holographic communication strategy is tested against similar, related approaches from the literature, appraising the meaningfulness and expressiveness of each MR-based interface through a combination of subjective evaluations and objective metrics.

Chapter 5 attempts to further generalize the results of **Chapter 4** in scenarios of mobile HRC. The software architecture is expanded to account for representing and communicating intentions of mobile robots as holographic cues, while at the same time a DT is introduced in the system's loop to deal with non-verbal HTR communication. A user study is conducted in a simulated logistics-like environment and the bi-directional communication interface evaluated in terms of how it impacts team efficiency.

Chapter 6 details how, in light of previous experimental findings, the software architecture has been re-written from scratch to achieve a fully open-source, modular and general implementation, independent of the particular HRC context and easily re-usable by other researchers and practitioners to take advantage of the developed holographic communication scheme.

Chapter 7 leverages the new implementation to address holographic HTR communication, with particular focus on MR-based KT. Such teaching paradigm is first framed into the communication space developed in **Chapter 3**, then a novel approach at holographic KT is introduced, leveraging the capabilities of the software architecture discussed in **Chapter 6**. The new holographic KT is then put to test in a user study against traditional, physical KT to appraise its communicative effectiveness and perceived participants' UX.

In conclusion, **Chapter 8** summarizes the findings and the limitations of the thesis and outlines potential future research areas based on the progress made in this work.

Chapter 2

Background

Given the prominence of HRC as research field, and the ever increasing adoption of collaborative platforms in manufacturing environments, many studies started dealing with the topic of effective communication, aimed at improving efficiency of the human-robot team. As such, the background starts with an overview of solutions proposed over the years to address the communication gap, leveraging both explicit and implicit cues, and at the same time discusses the advantages and limitations of said approaches in real-world scenarios. Then, given this thesis's interest in HRC and XR, a thorough review of AR and MR-based approaches employed in literature is provided, setting up the context for the various contributions brought forth by the present work, which are detailed in the next chapters. A brief overview of DTs in HRC scenarios follows after that, to account for the precious contribution of such digital models to the development of a comprehensive communication framework. Finally, popular strategies used to assess the interaction and the efficiency of the human-robot team are presented.

2.1 Communication Interfaces in HRC

The introduction of cobots in industrial settings paved the way for a new era of manufacturing and human-centered industry. While the HRC paradigm may virtually offer numerous advantages in terms of productivity, human operators' well-being and reduced stress [43], it also brings forth a series of challenging aspects, which need to be effectively tackled to make sure human-robot teams can be implemented in a safe and acceptable [109] way. One of these barriers is represented by the difficulties in establishing a clear, straightforward communication with artificial agents, especially in industrial environments where the robotic platforms generally lack humanoid features and may appear, therefore, scarcely familiar.

Such lack of familiarity could then reflect in a limited social acceptance of collaborative robots [123], resulting in potential loss of productivity and degraded team efficiency. To cope with these aspects, researchers started proposing solutions to expand the communicative capabilities of collaborative robots, attempting to bridge the lack of familiarity through the integration of straightforward, explicit communication channels, like the verbal one, which aid and simplify the interaction between human and machine. Alongside the vocal layer, a plethora of other channels, both explicit and implicit, has been explored to ensure intuitive, efficient communication, either through RTH interfaces or HTR ones. The following paragraphs provide an overview of successful and recent approaches, distinguishing between interfaces aimed at conveying information from robot to human and vice versa.

2.1.1 RTH Communication

Successful HRC and a more broad social acceptance of collaborative platforms can be achieved only if the human operator is able to grasp the robotic teammate's internal state during the interaction. This implies enabling robots to convey their intentions in a straightforward way to the human teammate, using appropriate communication channels.

To this regard, the most naive, yet direct channel consists in the vocal one, which can be exploited by collaborative robots to convey their intentions and their decision-making processes by means of natural language [119, 59, 161]. Likewise, related approaches have robots share their internal state with the human teammate, asking for help [44] or suggesting actions that the operator should take during the collaboration [20]. Nevertheless, while the vocal interface offers an intuitive channel for interaction, excelling at conveying high-level concepts and actions, it systematically fails at lower-level communication, such as when the robot has to describe how it will move throughout the workspace during the collaboration, or how it will fetch and deliver objects to the human teammate. Furthermore, the desirable degree of familiarity and intuitiveness brought by vocal interfaces comes at the cost of performances of the human-robot team, since such interfaces are generally computationally expensive and may slow down the pace of the collaboration significantly [141]. Finally, when contemplating standard industrial manufacturing environments, like a workshop or a logistics hub, vocal interfaces might prove less than optimal due to elevated levels of environmental noise [35].

An alternative to verbal communication is represented by light-based interfaces, through which robots can convey directionality and navigation intentions using blinking or flashing cues, similarly to what happens with the use of turn signals in vehicles [155, 148, 30, 66].

These approaches are particularly preferred in those scenarios where mobile robots need to signal their movements to human teammates or bystanders in a straightforward way. Nonetheless, analogous techniques recently started being adopted even with fixed manipulator robots, with light cues being used to signal certain internal states of the robot or its inclination to interact with the human colleague [156, 8, 24]. The main drawback of these approaches, however, is that, while seemingly intuitive and direct, communication interfaces relying on blinking and colored light signals might be perceived as insufficiently expressive of robots' internal state. Moreover, the human operators involved in the interaction may need to commit the associations between visual cues and robot actions to memory, leading to a higher cognitive burden, as noted in [89].

Another example of RTH interface involving visual cues is display-mediated communication. Screens and Graphical User Interfaces (GUIs) can be used as straightforward channel to convey information about robot's decision-making process and upcoming actions [148, 177, 3, 16] or to augment the machine's implicit communicative ranges, for example by endowing non-humanoid collaborative platforms with a display-mediated gaze [149, 84, 157]. While the former approach may prove rather distracting and confusing for the human teammate, who is continuously required to divert their attention from the collaborative workspace to the external screen, the latter solution has demonstrated how implicit cues, derived from a simulated robot gaze, can be successfully employed in interactive tasks to signal the machine's intention to pick or hand over a particular item [51].

Speaking of implicit cues, another compelling approach at intuitive RTH communication draws inspiration from human arm movements, which are by nature intent-carrying and not purely functional motions [21]. Specifically, through simple observation of several kinematic properties of arm movements, such as velocity and finger positioning, individuals can easily tell whether their teammate intends to reach for and grasp some object, or rather desires to initialize an handover [152]. In a similar fashion, several studies proposed the usage of motion planners capable of generating legible robot trajectories [47, 46, 147, 87], that is movements aimed at conveying certain intent, in order to provide useful insights to the human operator who is interacting with the machine. As an example, works from Dragan *et al.* [46] and Lastrico *et al.* [87] have demonstrated how predictable and legible robot motions improve human teammate's awareness, and contribute to the individuals behaving more proactively towards completion of the collaborative task. The only downside of these strategies lies in the additional constraints introduced to achieve a legible trajectory of the robot's arm, constraints which could significantly increase the time required by the planner to find a feasible solution, especially when the task takes place in particularly cluttered environments [78], resulting in potential downtime throughout the interaction.

Moving back to more explicit forms of RTH communication, several approaches proposed the usage of projection systems, usually mounted above the collaborative workspace, capable of visualizing relevant information as 2D visual cues, directly in the shared environment. Specifically, works from Andersen *et al.* [4], Sonawani *et al.* [151] and Bolano *et al.* [15] introduce communicative interfaces able to inform human operators about impending robot manipulation actions, by projecting 2D visual warnings directly onto the relevant objects. Similarly, the solution proposed by Ganesan *et al.* [55] involves a projection system that cobots can use to indicate actions that the human should perform during the collaboration, by highlighting relevant portions of workspace or items through visual cues. Finally, several related strategies have been evaluated, where a projection-based interface is employed to visualize the workspace occupancy of upcoming robot's movements, thus contributing to the safety of the human operator during interaction [165, 67, 68, 164]. Nevertheless, such solutions based on projection technologies share a common downside: they depend on a calibrated, structured environment to guarantee accurate projections, rendering them ineffective in dynamic scenarios where robots are frequently repurposed for different tasks. Furthermore, the 2D nature of the projected information may also result in less expressive communication of robots' intentions, reducing the effectiveness of these interfaces. Nonetheless, a successful context in which projection-based interfaces have been adopted is mobile Human-Robot Interaction (HRI) [41, 33, 170, 34], where these approaches can be used to signal robots' navigation intention and direction cues through 2D projections on the floor.

To overcome the inherent limitations of 2D projected information, XR-based approaches, which are analyzed in depth in upcoming Section 2.2, started emerging as natural evolution to these latter solutions, opening up the possibility to exploit 3D, more meaningful digital overlays for RTH communication.

2.1.2 HTR Communication

While the RTH aspect is certainly crucial to make sure collaboration between humans and robots unfolds successfully and efficiently, HTR communication is nevertheless important, to provide operators with straightforward and intuitive mechanisms to interact with the machine.

To this regard, vocal interfaces still represent the most direct solutions. Several approaches have been proposed over the years, to either have humans instruct robots to perform certain actions during the collaborative process [159, 7] or the let operators explicitly ask for robot's aid in completing joint tasks [10]. However, while suffering from environmental limitations, as pointed out in the previous paragraph, verbal interfaces also require a layer of linguistic representation, which is necessary for robots to properly grasp the key concepts and instructions expressed by the human teammate. This particular aspect, combined with the computational complexity of processing human verbal inputs on the fly, may limit the effectiveness of vocal interfaces in industrial collaboration, as recently stated in [88].

An alternative to verbal communication is represented by gesture-based interfaces, which leverage humans' upper limbs movements, that by nature are generally expressive and intent-carrying [27]. In this context, a popular technique employed to acquire and process raw data of human motions consists in using Inertial Measurement Units (IMUs), which are usually attached to the human operator's upper limbs or hands in the form of wearable sensing devices. Multiple studies leveraging this strategy have been proposed in HRC scenarios [42, 116, 18, 9, 131], where the stream of inertial data is used to recognize human activity, therefore enabling robots to react to specific human actions or adapt to the operator's working pace. Nonetheless, IMUs suffer from inherent drift when exposed to prolonged, continuous usage, which degrades the reliability of the acquired data over time. As a result, approaches relying solely on inertial sensors have been gradually set aside over time, in favor of multi-modal approaches, which combine IMUs with other communicative modalities, such as techniques that integrate speech and gestural communication [48]. Continuing on the topic, an alternative to IMUs which recently gained much interest involves vision-based gesture communication, that is an approach at extrapolating gestural cues from video streams, exploiting state-of-the-art Machine Learning (ML) techniques. As an example, Karami *et al.* [77] proposes an interface which makes use of an RGB-D camera to track the human skeleton, exploiting such information to recognize operator's gestures aimed at instructing a robot manipulator during a collaborative inspection task. Similarly, Ferrari *et al.* [52] recently introduced a multi-modal interface, integrating speech and gestural channels, where the human operator can specify items that the robot should pick and hand over using a combination of vocal commands and pointing gestures, acquired and processed through an RGB-D camera. The drawback of these approaches lies in the need for the operator to pause their task to explicitly carry out the gestural communication act, which could possibly represent a bottleneck to team efficiency where continuous exchange of information between human and robot is needed.

Finally, techniques exploiting more implicit cues for HTR communication have been proposed as well. To this regard, gaze represents an intuitive and straightforward channel, which can be used by operators to signal important information to the robot teammate through simple observation of certain objects or places. As an example, a recent work from Zou *et al.* [178] presented a novel interface where speech is combined with gaze to convey human's intentions, enabling the operator to intuitively signal which items the robot has to pick up from the shared workspace to complete a stacking task. The disadvantage of gaze-based interfaces, however, is how expensive eye-tracking devices generally are.

Following on the topic of implicit cues, the effect of body posture has been investigated as well, suggesting how even these implicit cues play an important role in conveying the human operator's willingness at interacting with the robotic teammate [112, 114].

2.2 Extended Reality in HRC

In contrast with the various interfaces discussed so far, in light of their advantages and limitations, XR-based communication approaches have recently gained much interest as they do not need complex projection infrastructure or other expensive devices and can be adopted with minimal calibration procedures. Furthermore, the new layer offered by 3D digital augmentation promises to deliver engaging, immersive experiences, where digital content superimposed to the real world can be used to create expressive, intuitive, novel communication approaches. To this regard, this section provides an overview of XR-based interfaces which have been proposed in literature to address the communication aspect in HRC. However, before proceeding in the discussion, and to avoid confusion in the readers, it is worth mentioning here what is considered to be, for the context of this thesis, the discriminant between AR and MR, given that the two terms tend to overlap in the literature, with similar interfaces being considered AR by some authors and MR by others. In particular, what I recognize for the rest of this dissertation, as separating element between AR and MR, is the hardware employed to experience a given interface. On the one hand, AR involves the pure overlay of digital content onto a scene, with virtual entities that are spatially projected with respect to the surrounding environment, but unable to interact with it. This type of interface is generally realized through hand-held devices, including smartphones or tablets. As a result, from now on the term AR will exclusively refer to approaches that employ everyday hardware for building communication interfaces. On the other hand, MR, whose terminology was originally introduced in [111], involves the creation of hybrid experiences, where digital entities projected in the environment are able to interact with it in a realistic way. Building such additional layer of interaction between real and virtual objects requires constructing a digital representation of the world around the user, which is generally achieved through HMD devices, that possess the necessary sensing capabilities. As a result, I will hereafter refer to MR-based approaches to indicate those interfaces which are experienced through HMD devices, such as the Microsoft HoloLens or the Magic Leap visors.

Overall, the following paragraphs discuss in detail both AR and MR approaches, distinguishing between interfaces aimed at implementing RTH communication or vice versa, consistently with the previous section.

2.2.1 RTH Interfaces

The development of technologies like AR, and its integration in everyday devices such as smartphones and tablets, paved the way for the design of new strategies, able to overcome the structural and communicative limitations of projection-based interfaces. To this regard, one of the pioneering approaches in HRC consists in work from Michalos *et al.* [110], where an AR interface, powered by a hand-held device, is used to enhance the human operator's awareness during a collaborative assembly process, superimposing 3D information about robot's imminent movements directly onto the device's camera feed. In a similar fashion, Palmarini *et al.* [122] proposed a tablet-based AR interface, where the upcoming movements of a tabletop manipulator robot can be previewed in the form of 3D animated, digital overlays. A related, yet relevant and more recent strategy has been proposed by Mourtzis *et al.* [113], with an AR-based interface, powered by a smartphone application, that can improve operators' awareness by previewing the imminent motions of a robot manipulator, along with the corresponding safety zones for the human teammate. In the context of collaboration with mobile robots, on the other hand, Chandan *et al.* [38] recently proposed a similar tablet-based interface, through which several robots can convey their current state and imminent, planned navigation trajectories using intuitive and expressive holographic overlays. A common, compelling point about the aforementioned works is that the user studies conducted by these authors to assess their solutions suggested a positive impact of AR-based RTH communication, particularly on human's trust in the robot teammate, given the possibility to preview imminent robots' movement through intuitive visual cues. Nevertheless, the downside of interfaces based on hand-held devices is represented by the impossibility, for the human operator, to collaborate hands-free with the robot. As such, the efficiency of these communicative approaches is limited to scenarios where concurrent activity between agents is not required.

The turning point is this context has been the introduction of commercial HMD devices, which enabled researchers to overcome the inherent limitation of hand-held displays. These wearable, compact visors provide several advantages with respect to display-mediated interfaces, such as the possibility for operators to experience holographic content directly in their own first-person perspective [120], while at the same time maintaining hands free to cooperate with the robot teammate. Additionally, the aforementioned ability of most commercial HMD devices, capable of mapping the surrounding environment by means of embedded RGB-D cameras, opened up the possibility to create more immersive MR experiences, where holograms and real world merge. To this regard, the usage of this new MR layer in HRC as channel to convey robot's internal state and intentionality has been first addressed by Ruffaldi *et al.* [139], where a holographic arrow is superimposed on a robot manipulator's wrist and used to signal the robot's imminent movements. A similar approach has been proposed in [158], where a static trail of holographic spheres is used to convey the planned trajectory of a manipulator's end-effector, providing useful information to the user about an imminent robot's movement. Finally, works from Rosen *et al.* [136] and Williams *et al.* [174] presented MR-based interfaces where the robot is able to indicate which item it is going to manipulate next by projecting an holographic sphere on it. In contrast with these strategies, which offered minimalist intention cues, more comprehensive approaches [137, 61, 13] proposed the static, holographic representation of the 3D volume swept by robot's imminent movements, thus yielding more exhaustive information, useful to guarantee operator's safety and to assess whether robot's trajectories are collision-free. While certainly offering a thorough overview of the robot's intended motion, however, the drawback of these approaches lies in such overload of holographic content on the operator's field of view, which is at risk of being counter-effective, negatively impacting on the individual's cognitive load during the collaborative task, and rendering the implementation of these interfaces in real-world scenarios impractical. To this extent, in order to overcome the limits of such state-of-the-art, one of the goals of this Ph.D. work has been attempting to find a trade-off between minimalist interfaces, which may result insufficiently expressive of the robot's intention, and overloaded ones. To fulfill this goal, the formalization of *MR-Space* has been fundamental, in order to identify the optimal trade-off between communicative capabilities and a pragmatic, efficient implementation of the communication interface.

Another limitation of the aforementioned works is that they deal with conveying robot's intentions and planned movements in scenarios where little or no teamwork is required. As such, the communicative effectiveness of MR-based strategies is not proven for situations where explicit interaction between human and robot is needed to achieve the shared task. To this regard, work from Newbury *et al.* [117] attempted to close the gap, introducing a MR-based interface which allows robots to convey, in the form of static holographic representation, their expected end-effector's pose (i.e., position and orientation) during an upcoming handover operation with the human agent. Similarly, research from Arevalo *et al.* [6] presented an interface which can be leveraged by human operators to assess whether upcoming pick-and-place actions performed by the robot teammate will unfold successfully, by means of holographic projections of the robot's gripper. Overall, the user studies conducted by these authors have highlighted a positive effect of MR-mediated communication on the interaction between human and robot, with individuals reporting an improved UX when interfacing with the machine. However, the tasks considered in such experimental validations were rather simple and may not account for the various facets of a complete cooperative process. Therefore, a second important objective of this thesis has been attempting to generalize the aforementioned results to structured collaborative activities, where explicit teamwork is required and where the agents are expected to interact in multiple ways, with the final aim of evaluating the effectiveness of MR-based communication on team efficiency.

As final remark to this paragraph, it is worth mentioning related state-of-the-art approaches where MR-based, RTH communication was used in scenarios involving mobile robots. To this regard, work from Walker *et al.* [166, 154] introduced a MR interface, capable of conveying the navigation intentions of a drone agent as sequence of holographic way-points in the 3D space, useful in scenarios where human operators and flying robots need to coexist. Similarly, more recent work from Gu *et al.* [62] proposed a solution to address situations where direct line of sight between human and robot is not available, enabling the operator to see the robotic teammate's location and its imminent navigation direction as holographic projections through walls. In the context of this Ph.D. work, the safe and efficient communication of robot's navigation intentions to the human teammate has been addressed taking into account such previous studies, but the proposed approach leverages the formalism of the *MR-Space*, which involves animated, holographic cues, rather than simple static representations, providing an additional temporal and direction layer to the communicative act.

2.2.2 HTR Interfaces

When it comes to HTR communication, XR has been extensively explored, with various approaches employing either AR or MR technologies. On the one hand, various solutions have been proposed to simplify the way in which human operators can assign tasks and convey goals or targets to robotic teammates using AR, in combination with hand-held devices. As an example, some studies leveraged tablet-based interfaces [54] or smartphone-based ones

[32, 31], to let operators instruct manipulator robots about items to be fetched, highlighting relevant pick and place locations through digital entities on the device's screen. Similarly, work from Cao *et al.* [25] introduced a smartphone application that operators can use to teach complete action sequences to mobile manipulators, drawing navigation paths on the screen and selecting relevant items to grasp using corresponding virtual overlays. Likewise, the application proposed by Papcun *et al.* [125] allowed operators to convey *no-navigation zones* to a mobile robot, simply drawing geometric, holographic shapes onto the prohibited areas. Nevertheless, all the aforementioned approaches incur in spatial limitations, brought forth by the perspective mismatch between holograms projected onto the device's camera and corresponding 3D locations in the real world. As such, the application of said solutions in real industrial environments would prove scarcely efficient as communication channel [102].

On the contrary, MR-based interfaces have gained interest as promising alternatives, since they overcome the perspective mismatch by letting users experience the holographic layer in first person. To this regard, a plethora of approaches has been proposed in recent years, providing an intuitive HTR communication layer for programming robots' actions. For instance, multiple works proposed holographic interfaces for instructing desired end-effectors' trajectories in manipulator robots, through the definition of way-points in the MR space via a combination of speech and gestural communication [129, 13, 37, 168, 36, 69, 167]. Similar solutions have been proposed also for conveying navigation targets to mobile robots [133, 96].

While programming robots' trajectories in MR has been studied in depth as a topic, it is not the same for KT, which can certainly be regarded, in the context of this thesis, as an instance of HTR communication, aimed at instructing robots' actions through physical guidance. With KT still playing a major role in HRC scenarios [22, 50, 132], it is evident, in light of the aforementioned works, how combining MR and KT could lead to a new form of intuitive teaching process. This could, in turn, open the possibility to perform KT on any robotic platform, including those that were not natively designed for this purpose. As a matter of fact, the literature includes only a few studies who attempted to address this issue. Some solutions still rely on the physical robot for hand guidance and employ the holographic channel only for later visualizing the learned robot action and for adding constraints to the motion [97, 98]. MR-based communication to achieve KT is foreshadowed in [127], where the authors exploit the hand-tracking capabilities of HMD devices to manually guide a holographic industrial manipulator, interacting with its individual links. Similarly, in [126] a system is presented where a tabletop, virtual robot can be taught a simple pick-and-place task via holographic hand guidance. Finally, a recent work from Rivera *et al.* [135] proposed a MR interface for intuitively teaching trajectories to a holographic collaborative manipulator. Nevertheless, all the aforementioned approaches are robot-dependent and only partially address the topic of intuitive, holographic KT, as they lack a structured formalization of the communicative interface. In contrast, during the final part of this Ph.D. work I concentrated my effort on overcoming the limitations of such state-of-the-art, proposing a robust approach at holographic KT, which makes use of the communicative formalism of *MR-Space*. Furthermore, the proposed approach is robot-independent, therefore representing a first attempt towards a framework for universal KT, leveraging intuitive, holographic HTR communication.

2.3 Digital Twins in HRC

As one of the main pillars of Industry 4.0 [63, 79], DTs have recently experienced an exponential growth as research topic, brought forth by the digitalization of companies and widespread adoption of Internet of Things (IoT) devices. They have been successfully employed in multiple manufacturing scenarios, to monitor and optimize production chains [17, 11, 145]. Generally speaking, a DT can be described as the virtual counterpart of a physical system, in which bi-directional, integrated communication occurs between the real and the digital instances [82]. The concept of bi-directional communication is inherent of DTs: one the one hand, various sensors and IoT devices are used to extract relevant data from the physical system, and such information is leveraged to drive the digital model and to maintain a consistent virtual representation of the system's state. On the other hand, the DT can in turn be leveraged to validate or steer look-ahead simulations, aimed at influencing decision-making and control mechanisms in the physical system itself. The terminology and main idea behind DTs were first introduced for product life-cycle management [74]. Later, the same authors expanded the original definition by stating the foundational components of DTs: the physical entity, its virtual representation, and the bi-directional communication between them [60]. Based on the level of data integration, it is possible to distinguish between Digital Models (DMs), with no automatic data exchange, Digital Shadows (DSs), with automated data exchange flowing in one direction only, namely from the physical system to the digital model, and pure DTs, with bi-directional automatic data flow [82].

As with similar simulation technologies, DTs generally undergo several life-cycle phases, including development, operation and support [74]. Regarding the development phase, DTs have been successfully employed for various purposes, either as the optimal design tool for factory layouts and collaborative cells [104, 64], or as instrument to deploy, test and refine path planning algorithms for mobile robots in warehouse environments [153]. Similarly, in support phase DTs can be employed for training operators [173], as well as for reconfiguration of production lines [81]. Nevertheless, the operational phase is the most relevant one in the context of HRC. In this stage, DTs allow active collision monitoring between human and robot [90], workload balance based on agents' individual skills and ergonomic factors [105], and the recognition of human intentions [160, 153]. In the context of this Ph.D. work, the usage of DTs falls exactly within this latter context. Unlike previous works, however, the DT has been explored within the communication framework developed as main objective of this thesis. As such, I used the DT to create a HTR communication layer, which leverages the digital representations of human operators to infer their implicit intention cues, such as gaze and body posture, to achieve a more natural interaction with the robot teammate.

2.4 Evaluating Team Efficiency

The previous sections addressed the most relevant approaches employed in HRC to build communication interfaces. Nevertheless, it is not trivial to assess the effectiveness of such interfaces, since many factors contribute to the overall success of human-robot teams, including coordination, fluency, synchronization, safety, trust and so on, variables which are generally not easy to measure quantitatively. This section therefore provides an overview of strategies adopted to evaluate human-robot teams, distinguishing between approaches aimed at measuring subjective metrics and solutions which instead focus on objective quantities.

When it comes to the evaluation of subjective experience, one of the most prevalent forms of assessment consists in questionnaires. They represent the most straightforward approach for measuring subjective quantities and users' perception. Over the years, questionnaires have been employed to evaluate various subjective aspects of collaborative tasks, including perceived user's safety [140], stress level [169, 57] or workload [95]. To this regard, the most popular instrument is the NASA Task Load Index (NASA-TLX) questionnaire [65], which is able to measure perceived stress and workload levels in operating contexts, and it is therefore applied in multiple domains, such as surgery, aeronautics and HRC.

While the NASA-TLX excels at evaluating physiological quantities, other questionnaires instead focus on assessing the UX of certain interfaces or interactive products. Among these, there is the System Usability Scale (SUS) questionnaire [19]. Such tool provides a standardized approach for evaluating usability of a system or product, offering a quick and reliable measure of users' satisfaction and ease in interacting with a given technology. Aside from the SUS, another widely popular tool is the User Experience Questionnaire (UEQ) [144], which is extensively adopted to evaluate interactive products in terms of their offered UX. With the UEQ, it is possible to evaluate products under six different scales, namely *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation* and *novelty*. Given the focus of this Ph.D. work, namely the development of a valuable communication layer based on MR, the UEQ has been employed multiple times to assess the efficiency and appeal of the proposed holographic interface, in order to find out whether the digital layer offered an intuitive channel for communication, possibly yielding a positive impact on perceived users' trust towards the robot and safety during the collaboration.

Nevertheless, questionnaires alone may not represent a thorough evaluation system, as they may introduce cognitive distortions, especially when the users have to fill them after the experimental trials, which could cause recollection errors that could bias the responses. As suggested in [12], it is crucial to complement subjective evaluations with more objective ones, resulting from the assessment of task performance metrics or behavioral measures, which can provide a fairer estimate of team efficiency. To this extent, work from Hoffman [70] provides a series of objective metrics to evaluate fluency in human-robot teams, including agents' downtime and percentage of concurrent activity. Given their general definition, which is not dependent on the particular HRC scenario, such metrics can be easily employed in any user study where team efficiency is evaluated. Similarly, works from Castro *et al.* [29] and Matstumoto *et al.* [108] present a list of useful metrics, which include success rate of joint actions and completion timings for the collaborative task. The aspect of team coordination is addressed as well in [99], where the authors discuss how efficient communication can improve agents' proactivity, leading to reduced downtime in the task. At the same time, works from Arents *et al.* [5] and Marvel *et al.* [106] address the topic of safe collaboration, discussing the need for efficient interfaces which reduce the hazards of collisions between agents. Overall, the aforementioned metrics represented convenient and reliable quantities to measure team efficiency and have been extensively employed during this Ph.D. work, to find out the impact of holographic communication on human-robot teams.

Chapter 3

Communication Space Formalization and Early Results

Elaboration and integration of two articles
Published in: Proceedings of the 39th IEEE International Conference on Robotics and Automation (ICRA 2022) [100]
Submitted to: Robotics and Autonomous Systems

To achieve efficient HRC in industrial scenarios, effective communication among agents is of utmost importance. This concept has been extensively discussed in previous chapters, highlighting the need to explore novel channels to enrich and simplify interactions between humans and robots. As suggested in Chapter 1 and supported by multiple studies on joint human-human activities [80, 115, 23, 72, 142], team efficiency is significantly influenced by individuals' ability to read each other's intentions during tasks, fostering collaborations with minimal downtime. However, human operators engaged in collaborative tasks with robotic platforms find it challenging to interpret the artificial agent's intentions, particularly in the absence of natural cues such as gaze. To attempt to cope with this aspect, multiple strategies, discussed in Chapter 2, have been proposed in the literature to expand the communicative abilities of cobots. Among these, MR represents a promising, valuable alternative, capable of ensuring straightforward, visual communication. As such, one of the main objectives of this thesis has been the design of an intuitive, expressive interface that collaborative platforms can employ to project their intentions in a meaningful way, by means of holographic cues.

Nevertheless, designing an effective communicative interface is no easy task, regardless of the particular channel employed. Many approaches discussed in Chapter 2 proposed interfaces that were only partially expressive of agents' internal state, resulting in unclear and ambiguous communicative acts. To this extent, a valuable communication layer based on MR can only be designed by having a thorough, analytical formalization of the communicative acts, knowing how information is mediated by the holographic channel and exchanged between agents in a meaningful way. This is only possible if, prior to implementing the holographic interface empirically, a modeling phase occurs. Therefore, before delving into representing robots' intentions via holographic cues, the present Chapter introduces *C-Space*, the first contribution brought forth by this Ph.D. work, namely an analytical framework to model communication acts in collaborative scenarios. Following after that, the Chapter transitions to discuss how robots' intentions can be efficiently represented using holographic cues, leveraging the new formalism. Then, a first practical implementation of such communicative interface is introduced as *MR-HRC-V1*, capable of projecting intention cues for robot manipulators during collaborative task. Finally, the interface is tested in a preliminary user study within a collaborative assembly task, involving a human operator and the robot Baxter, and evaluated in terms of team efficiency, according to some of the metrics discussed in Section 2.4.

3.1 Methods

3.1.1 Communication Space

In the context of this Ph.D. work, communication can be defined as the process by which individuals exchange a *piece of information* (I) through single or multiple communicative channels. The set of all possible channels available in a given interactive context can be denoted as $M = \{m_1, \dots, m_{|M|}\}$, and may account for different forms of vocal, gestural or visual communication, as detailed in Chapter 2. As such, we can define a communicative act C, expressed through a combination of $N \leq |M|$ channels among all the possible ones, and extending over a temporal time span $\boldsymbol{t}$, as

$$C(I, \boldsymbol{t}) = \bigcup_{i=1}^{N} C_{m_i}(I, \boldsymbol{t}_i), \tag{3.1}$$

where $C_{m_i}(I, \boldsymbol{t}_i)$ represents the communicative contribution brought forth by the ith channel m_i in expressing the information I in the time interval $\boldsymbol{t}_i$. It is important to observe that $\boldsymbol{t}_i$ is strictly bounded by the length of the temporal time span $\boldsymbol{t}$ associated with the overall communicative act, whereas for any two different time intervals $\boldsymbol{t}_i$ and $\boldsymbol{t}_j$ all the appropriate relations induced by Allen's interval algebra may hold, in principle.

Since communicative channels may have variable effectiveness depending on the interactive context in which they are applied, combining them usually makes the communication act more straightforward and less ambiguous. This is what happens, for example, in the case of non-verbal communication, which typically involves both gaze and gestures to better convey meaningful intent. For a given channel m, however, there may exist multiple strategies to contribute to the communication of the piece of information I, with different degrees of effectiveness in the communication. For example, humans can naturally express the intention of reaching and picking objects through *legible gestures* using implicitly the upper limbs motion as a communication channel. Conversely, robot motion is generally planned to be *functional*, namely, it prioritizes the successful execution of the action over its legibility by an external observer. While the channel is the same in both cases, the communicative act in the latter case is insufficient to convey the robot's planned action because it accomplishes a different *function*. We can formalize this concept in analytical terms by defining, for any channel m_i in M, a set of channel-specific functions

$$F_{m_i} = \left\{ f_{m_i,1}, \dots, f_{m_i,|F_{m_i}|} \right\}, \tag{3.2}$$

where each $f_{m_i,j}$ specifies one of the possible strategies through which the information I is *rendered* in a communicative act by channel m_i. With this notation, and observing that for a given channel m_i multiple functions may be used to strengthen the communication of a piece of information I, we can formulate the communicative contribution of channel m_i as

$$C_{m_i}(I, \boldsymbol{t}_i) = \bigcup_{j=1}^{L_i} f_{m_i,j}(I, \boldsymbol{t}_{ij}), \tag{3.3}$$

with $L_i \leq |F_{m_i}|$ and where $\boldsymbol{t}_{ij}$ represents the time interval spanning the communicative component associated to function $f_{m_i,j}$. Equations (3.1) and (3.3) serve as the general formulation of the *C-Space*, providing a comprehensive symbolic representation of the communicative act C employed to express information I. In such form, *C-Space* is an abstract representation and can be applied to multiple domains of communication. Nonetheless, the present thesis's work aims to develop a communicative layer based on MR. As such, from now on a specific instance of *C-Space* will be addressed, namely *MR-Space*, which is specifically tailored for intuitive holographic communication. As already mentioned in Chapter 1, the first part of my Ph.D. has been dedicated to the RTH communication aspect, and, as such the *MR-Space* formalism has been first adopted to model and represent robots' intentions by means of holographic cues, with the final aim of improving human operators' awareness and team efficiency during collaborative tasks.

Representing Robot's Intentions

The first step to designing an intuitive human-robot communication layer requires knowing what information to convey through communicative acts. Humans are natural masters in communicating their intent by gaze and motion. However, despite recent advances in social aspects of human-robot communication [87], this is not necessarily true for robots. If we want to provide collaborative robots with the ability to communicate their upcoming intentions, it is first necessary to give a principled yet operational definition of such a qualitative concept. Generally speaking, I argue that robots' intentions can be represented as a series of future robot *states* $\boldsymbol{\tau}$ and *beliefs* $\boldsymbol{\omega}$. In particular, the state at time instant t incorporates information associated with the robot's pose $\boldsymbol{x}(t)$, i.e., position and orientation of the robot in the environment, and its joint positions $\boldsymbol{q}(t)$. Thus, the robot's state can be defined as

$$\boldsymbol{\tau}(t) = \{\boldsymbol{x}(t), \boldsymbol{q}(t)\}\,. \tag{3.4}$$

Beliefs, instead, may refer to external factors related to how the robot intends to interact with objects in its workspace. As an example, in a general HRC context, a belief may represent the state $\boldsymbol{\xi}(t)$ of a particular object that the robot intends to pick up or manipulate in its upcoming action. The term state implies that beliefs can be used, in accordance with the relationship expressed in (3.4), to represent objects' poses and, in case of articulated items, even their internal joint configuration. Throughout this thesis, although this does not pose any limitation to the discussion that follows, it is assumed for the sake of simplicity that robots interact with simple (non-articulated) items and maintain at each time instant a series of single, disjoint beliefs about such objects. Therefore, we define the set of beliefs including poses of all items the robot could interact with as

$$\boldsymbol{\omega}(t) = \left\{\boldsymbol{\xi}_1(t), \ldots, \boldsymbol{\xi}_{|\boldsymbol{\omega}|}(t)\right\}. \tag{3.5}$$

It is noteworthy that the aforementioned distinction between states and beliefs is conceptual. It is aimed at distinguishing between quantities directly associated with the robot, i.e., its configuration and trajectories, and variables depending on how the robot acts within its environment. In *MR-Space* both states and beliefs are represented as holographic objects. As such, we posit that the minimal set I of pieces of information that can be conveyed through communicative acts in a HRC context is given by

$$I = \{\boldsymbol{\tau}, \boldsymbol{\omega}\}, \tag{3.6}$$

where the independence from the time indicates that I includes all possible pieces of information that could be delivered, starting from the sets $\boldsymbol{\tau}$ and $\boldsymbol{\omega}$. With such information, collaborative robots can actively project their future intentions, both in terms of upcoming goal states, in order to represent motions and trajectories, and beliefs, to inform their workmates about which object (or multiple objects) they are going to handle in the upcoming action, therefore enabling a human teammate to properly react. It has to be remarked that the set I is considered *minimal*, in the sense that other qualitatively different pieces of information may be the subject of RTH communication, but either these can be derived from the minimal set, or the minimal set should be communicated in any case.

As discussed in Chapter 2, a typical collaborative robot employed in manufacturing scenarios lacks the communicative channels necessary to project its intentions, and can only convey meaningful intent through its motions during the interaction. We can, therefore, consider robot's motions themselves as a possible communication channel, which we refer to as *mov*, and formalize such motions as trajectories $\mathbf{T}$, expressed in the robot state space $\boldsymbol{\tau}$ and lasting for a time interval $\boldsymbol{t}_{mov}$. Recalling the set of pieces of information I given in (3.6), we can formulate the communicative contribution of robot's motions according to the *C-Space* equations as

$$C_{mov}(I, \boldsymbol{t}_{mov}) = f_{mov}(I, \boldsymbol{t}_{mov}) = \mathbf{T}(\boldsymbol{t}_{mov}), \tag{3.7}$$

where the robot's trajectory is defined as

$$\mathbf{T}(\boldsymbol{t}_{mov}) = \{\boldsymbol{\tau}(t_{mov,s}), \ldots, \boldsymbol{\tau}(t_{mov,e})\}, \tag{3.8}$$

and where $t_{mov,s}$ and $t_{mov,e}$ respectively represent the initial and final time instants of said trajectory, i.e., the endpoints of the time interval $\boldsymbol{t}_{mov}$.

However, as discussed above, robot's motion alone is – typically – not communicative enough for the human operator to predict upcoming robot actions. Therefore, to add meaningfulness to the communication, we posit that it is necessary to complement the communicative act delivered via robot's motions with one coming from an *anticipatory* channel, which may provide intuitive cues about the robot's intention. Specifically, the thesis's focus is on anticipatory communication achieved through MR and, in general terms, we postulate that, by combining robot's motions (the first channel) with appropriate anticipatory holographic cues (the second channel), it may be possible to achieve a more intuitive and natural communication, leading up to an overall better form of collaboration. We can introduce the general formulation of the *MR-Space* by giving a symbolic representation of the whole communicative act, resulting from the interplay between the robot's motion channel *mov* and the holographic channel *mr*, as

$$C^{*}\left(I,\boldsymbol{t}\right)=C_{mov}\left(I,\boldsymbol{t}_{mov}\right)\cup C_{mr}\left(I,\boldsymbol{t}_{mr}\right), \tag{3.9}$$

where C_{mr} represents the communicative contribution obtained via the channel *mr*. Since C_{mr} is a communicative act mediated by a channel aimed at anticipating the actual robot's motions, it holds that the initial time instant of $\boldsymbol{t}$ corresponds to the initial time instant of $\boldsymbol{t}_{mr}$, whereas the final time instant of $\boldsymbol{t}$ corresponds to the final time instant of $\boldsymbol{t}_{mov}$. This relationship is better formalized via (3.10)

$$t_{mov,s}=t_{mr,s}+\Delta t, \tag{3.10}$$

which specifies that, throughout the communication, robot's motion begins Δt seconds late with respect to the corresponding holographic cues in *mr*.

According to the principles conceptualized in (3.2), however, the anticipatory visual feedback can take multiple forms depending on the specific function f_{mr} selected to project the robot's future states and beliefs in the MR channel. Among many possibilities, there may exist particular forms of holographic cues which may be more meaningful and intuitive in conveying the robot's intention. As a result, it is also critical to know which visual information, created via the MR channel, may be best suited to represent robot's states and beliefs in an HRC context. To this regard, a thorough comparison between multiple holographic alternatives is given in Chapter 4. On the other hand, in the present Chapter my novel approach is introduced, which leverages works from Rosen *et al.* [137] and Newbury *et al.* [117] as main references from the state-of-the-art. However, unlike previous works, the proposed approach employs dynamic holographic cues, thus offering a complete, expressive overview of the robot's imminent intentions without incurring in excessively cluttered digital overlays, that could otherwise damage the interaction entirely.

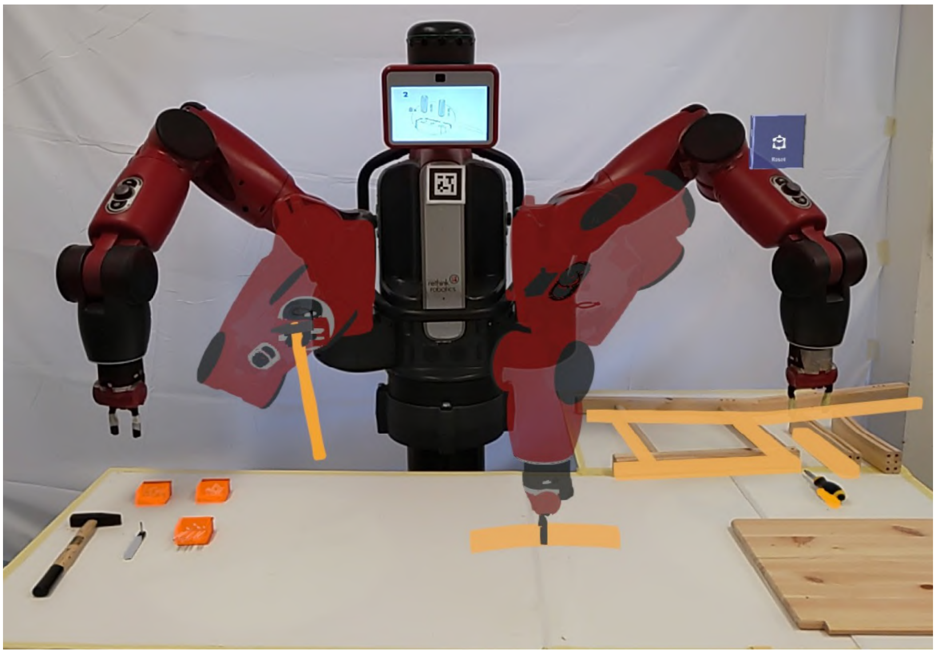


Figure 3.1 First person view of the DHT interface seen through an HMD device. The dynamic holographic cues allow the human operator to see upcoming robot's intentions during a collaborative process.

Dynamic Holographic Trajectory

As anticipated in the previous paragraph, the proposed Dynamic Holographic Trajectory (DHT) approach represents a trade-off, as well as a novel and more effective strategy, compared to the current literature. In work from Rosen *et al.*, and in similar works as well [137, 61, 13], robot's intentions are anticipated as static sequence of holographic states, thus covering the entire 3D volume swept by the imminent trajectory. This strategy, however, produces very cluttered and chaotic visual feedback, resulting in ineffective communicative acts, which do not provide the human operator with temporal and directional cues about the robot's motion. On the other hand, work from Newbury *et al.* [117] provides only static holographic visualization of the robot's final state during an upcoming interaction, thus lacking visual communication about the intermediate robot's states. Furthermore, neither of the aforementioned works deals with projecting robots' beliefs, and they only resort to communicating robot's future states and movements. In contrast, the DHT interface offers such capabilities, thanks to the *MR-Space* formalism. It can visually represent robot's intentions as dynamic holographic trajectories, thus providing thorough, intuitive communication without cluttering the operator's field of view. At the same time, it is able to convey robot's beliefs, thus it offers more expressive communicative capabilities, opening up the possibility to convey more complex intention cues, such as desired handovers with the human teammate and expected positions of items to be manipulated. Overall, the communicative act conveyed by this interface can be modelled as

$$C_{mr}(I, \boldsymbol{t}_{mr}) = \mathbf{T}(\boldsymbol{t}_{mr} + \Delta t) \cup \sum_{k=1}^{D} \Xi_k(\boldsymbol{t}_{mr} + \Delta t), \tag{3.11}$$

where $\boldsymbol{t}_{mr}$ is the time interval associated with the communicative act in mr and where

$$\begin{gathered} \Xi_k(\boldsymbol{t}_{mov}) = \{\boldsymbol{\xi}_k(t_{mov,s}), \dots, \boldsymbol{\xi}_k(t_{mov,e})\} \\ \forall k \in \{1, \dots, D\}. \end{gathered} \tag{3.12}$$

As such, the communicative act C_{mr} can be defined as the combination of the anticipated, holographic robot's state trajectory and the corresponding anticipated, holographic trajectories of the $D \leq |\boldsymbol{\omega}|$ objects manipulated by the upcoming robot action. Although static, a screenshot of this type of anticipatory feedback is provided in Fig. 3.1, but it can be best appreciated in the following video[1].

3.1.2 System's Architecture

This section describes *MR-HRC-V1*, namely the first version of the envisioned software architecture implementing the holographic communication interface discussed in the previous paragraph. The core of *MR-HRC-V1* is shown in Fig. 3.2, enclosed within the dashed lines. The left-hand side of the Figure shows how this core has been integrated in an expanded architecture used in the preliminary experimental scenario, as later discussed in Section 3.1.4. Implementation details are instead provided in Section 3.1.3.

Two are the core components of *MR-HRC-V1.* On the one hand, a module named *Mixed Reality Application* deployed on the HMD renders the MR interface enabling the human teammate to perceive the holograms superimposed to real world objects. Conversely, the high-level component *Robot Action Manager* groups all the modules dealing with robot's motion planning and execution. The two components are interfaced through an appropriate *Communication Adapter*.

Mixed Reality Application acts as virtual counterpart of the HRC scenario. As depicted in Fig. 3.1, this module aims at displaying through the HMD the robot's holographic representation, as well as the holograms of relevant tools and objects involved in the collaborative process, which are later employed to visually project intention cues according to the DHT

---

[1] Video: https://youtu.be/uXiH9ElsiD4

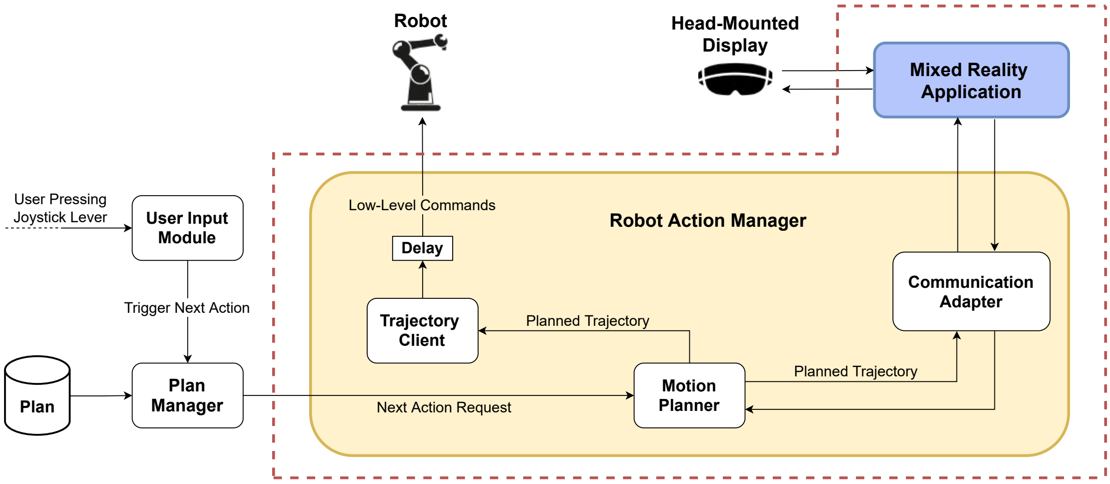


Figure 3.2 The *Mixed Reality Application* and the *Robot Action Manager* are the core components of *MR-HRC-V1*. *Motion Planner* computes robot's movements and dispatches planned trajectories to both *Mixed Reality Application* for the *mr* channel, and *Trajectory Client*, which deals with actual motion execution on the real robot in the *mov* channel after introducing a delay Δt. *Plan Manager* decides the next robot's action and, as such, acts as an external high-level task planner.

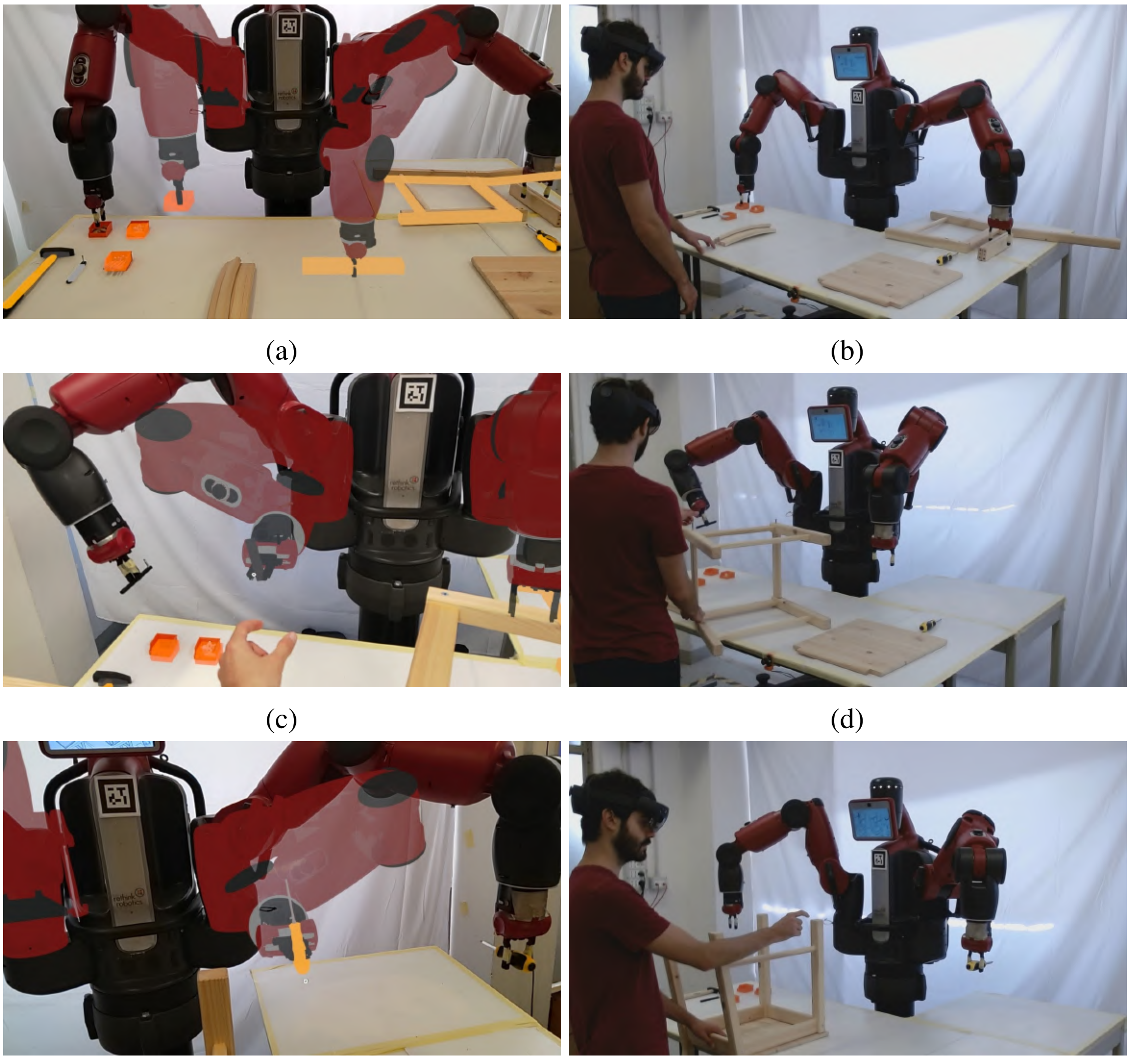
formalism. Focusing on the robot's hologram, whenever a new robot's action is planned, a contrived delay Δt is introduced as foreseen in (3.10), to allow the RTH communication act to take place, with the robot's hologram anticipating its imminent actions as provisioned by the *mr* channel.

In order to obtain a consistent MR experience, the first step involves calibrating the two channels *mov* and *mr*, i.e., the real and the holographic robot's reference frames. For this purpose, we employ a simple 2D bar-code marker attached to the robot, and exploit one of the many available software modules providing marker detection and 3D pose estimation capabilities. Upon detecting the marker, the module returns its 3D coordinates with respect to the HMD's frontal camera and such position is used to spawn and anchor the holograms coherently with the actual scene. As discussed in previous paragraphs, *MR-HRC-V1* deals only with fixed manipulator robots, therefore it is safely assumed that continuous tracking of the marker is not necessary, as the robot's base is fixed throughout the collaboration. As a matter of fact, the 2D marker employed in the experimental scenario is visible in Figure 3.1, attached to the robot's front. Additionally, dealing with fixed manipulator robots introduces, for the present context, a simplification in definition of the robot's intention as presented in Section 3.1.1. In particular, the $\boldsymbol{x}(t)$ component of the robot's state can be ignored without losing generality. Therefore, the present Chapter and the following one will hereafter refer to robot's state to exclusively indicate its joint configuration $\boldsymbol{q}(t)$.

Robot Action Manager is a collection of modules collaboratively enabling robot's actions planning and execution. The functionalities of such modules are detailed hereby, along with brief explanations on how they interact with one another and with the *Mixed Reality Application*.

Communication Adapter establishes a serial connection between *Mixed Reality Application* and *Motion Planner* for data exchange. The module handles all the communication between these two endpoints, including service requests or messages broadcasting, it manages data serialization and deserialization, and forwards all data through the network.

Motion Planner is responsible for planning robot's motions. With the aim of broadening the spectrum of possible applications for *MR-HRC-V1*, we assume the robot to be capable of performing different types of action (i.e., *skills*), each with its own specific motion planning routine. In particular, due to the collaborative scenario considered for this work, which involves an assembly, such skills included *pick-and-place* and *handover* actions, both types of action requiring the robot to fetch objects, but differing in how such items are supplied to the human teammate. In the case of pick-and-place actions, the robot retrieves objects needed for the collaboration process, and places them in a predefined position within the workspace, as in Fig. 3.3a, whereas through handovers it delivers tools in such a way that the human teammate can comfortably grasp them, as shown in Fig. 3.3c or Fig. 3.3e. For this reason, a planning request made to *Motion Planner* by an external high-level task planner, such as *Plan Manager*, must specify two parameters, namely the type of action for which a plan is sought, along with an identifier (ID) representing the requested object or tool to be fetched. Upon receiving a planning request, *Motion Planner* communicates with *Mixed Reality Application* to retrieve the object's pose (i.e., position and orientation) based on its ID, and then it uses such pose to plan the robot's motion for that particular action. Once the motion plan is computed, it is sent back in the form of array of joint configurations which make up the robot's state trajectory $\mathbf{T}(\boldsymbol{t}_{mov})$ to *Mixed Reality Application*, where it is processed and rendered as holographic animation of the robot's arm in the *mr* channel, according to the DHT formalization. As it unfolds, the holographic trajectory may interact with virtual representations of objects in the collaborative workspace, conveying that the robot intends to manipulate certain items or fetch some pieces that are to be delivered to the human operator, thus yielding the holographic component of the RTH communicative act. At the same time, the planned trajectory $\mathbf{T}(\boldsymbol{t}_{mov})$ is forwarded to *Trajectory Client*, which instead deals with execution of the motion on the real robot, i.e., in the *mov* channel. As



(a) (b)

(c) (d)

(e) (f)

Figure 3.3 A collaborative assembly process from two perspectives: left images show the first person view with the anticipatory holograms, whereas right images depict the same instants from an external point of view. In particular, in (a) and (b) Baxter is delivering wooden pieces and a box of dowels, in (c) and (d) Baxter supplies the hex key, and in (e) and (f) Baxter is handing over a screwdriver.

described above, this coordinated execution of trajectories in the robot's motion channel and in the anticipated robot's motion channel constitute the combined RTH communicative act envisioned by the *MR-Space* formalism in (3.9).

Finally, *Trajectory Client* is the module performing the translation from the *Motion Planner*'s output to a series of low-level commands for the joint-level controller, thus enabling the execution of the planned trajectory $\mathbf{T}\left(\boldsymbol{t}_{mov}\right)$ on the real robot. The commands are delayed for Δt seconds (which is a parameter set to 3 seconds in actual experiments), so that the human teammate can visualize the holographic action first. After such interval, the robot's arm begins moving following its virtual counterpart.

3.1.3 Implementation, Frameworks and Equipment

The *Mixed Reality Application* component has been developed using Unity, a worldwide popular game engine, which is supported by most commercial HMDs. The marker detection pipeline is managed by *Vuforia*, a well-known software development kit used to create MR contents easily integrated with Unity. The holographic representation is instead built using the Microsoft *Mixed Reality Toolkit*[2] (MRTK), a collection of tools and libraries specifically developed to design applications for MR-HMDs in Unity. MRTK yields the instruments to transform a standard 2D Unity scene into a 3D MR and provides the building blocks necessary to design augmented user interfaces. In this work, MRTK has been used to create the 3D MR interface by overlaying the virtual models onto the real objects, according to the coordinates returned by the Vuforia's marker detection pipeline. The robot model is incorporated into the Unity scene through the Unified Robot Description Format (URDF) Importer package[3], developed by Unity Technologies, and then rendered as a 3D MR asset through MRTK.

The Robot Operating System (ROS) [128] has been employed to develop the whole *Robot Action Manager* stack. In particular, I used the *MoveIt* [40] framework to implement the *Motion Planner* module. The *Communication Adapter*, instead, has been developed exploiting the ROS-Unity integration package[4], which has been recently published by Unity Technologies.

In order to support other researchers interested in the topic, I have decided to make the code of our architecture publicly available on GitHub[5].

As far as the equipment is concerned, a Microsoft HoloLens 2 headset has been used to run the holographic channel. This powerful HMD natively supports MR applications developed through Unity and possesses a wide range of sensors useful for spatial perception, including an array of four cameras used for head tracking and a time-of-flight (ToF) camera for depth sensing. Finally, the device's screen provides a 52° diagonal field of view for

---

[2] MRTK: https://github.com/microsoft/MixedRealityToolkit-Unity.

[3] Unity URDF Importer: https://github.com/Unity-Technologies/URDF-Importer.

[4] ROS-Unity Integration: https://github.com/Unity-Technologies/ROS-TCP-Endpoint.

[5] GitHub: https://github.com/TheEngineRoom-UniGe/MixedRealityHRC.git.

holograms projection, thus making it suitable to display digital overlays even in case of a close-proximity collaboration process with a robot.

The employed robot platform is Baxter from Rethink Robotics [53], a well-known dual-arm manipulator, which has been previously employed in related research works [139, 137]. In order to ensure that communication could be established between the ROS environment running within the Baxter's embedded computer and the Unity app deployed on HoloLens, the robot and the HMD have been connected to the same local network.

3.1.4 Experimental Setup

Having delineated the *MR-Space* communicative formalism and with a practical implementation of the holographic interface at hand, a preliminary user study has been conducted in a collaborative assembly setup, using the implementation and hardware described in the previous paragraph. This pilot study was aimed at evaluating effectiveness of the DHT communication scheme in terms of team efficiency, thus assessing whether MR-based, RTH communication positively impacted the collaboration with respect to fluency and coordination among agents. In particular, it has been hypothesized that the overall communicative act C^* described in (3.9), resulting from the interplay of *mr* and *mov* channels

- *H1* reduces the number of accidental collisions during the collaboration process and, consequently, (hypothesis *H1.a*) collaboration downtime is reduced as well;
- *H2* increases the human teammate's proactivity, thus improving team coordination.

Collaborative Scenario

The target task considered in this Chapter consists in the collaborative assembly of a wooden chair. The workspace whereby the human-robot collaboration process takes place is visible in Figure 3.3. It includes a table which use is shared by both the human and the robot teammates, on top of which various wooden pieces, components, and such tools as a screwdriver, a hammer, and an hex key are placed. The assembly components, such as screws, dowels and bolts, are stored in orange boxes, which the robot can supply with during the collaborative process.

The task is divided in 10 sequential steps. During each step Baxter performs pick-and-place or handover operations, providing its human teammate with the items necessary to build the piece of furniture step by step, and putting tools and components away when they are not needed anymore. The assembly actions, instead, are strictly left to the human teammate. In order to provide the human teammate with support during the assembly phase, step by step visual instructions are shown on the Baxter's *head* display. It is worth mentioning that the typical trial lasts around 10 minutes, during which the human and the robot teammates keep collaborating on the assembly. An exemplar run of the collaboration process is available in the video referenced in Section 3.1.1.

In order to adapt the system's architecture to this peculiar scenario, two additional software components have been implemented and added, i.e., *Plan Manager* and *User Input Module*. *Plan Manager* handles the overall Baxter's plan, which for the present purposes consists in a sequence of scripted actions that the robot carries out during the collaborative task. The plan is stored as a text file and, due to the sequential nature of the assembly process, the actions are performed one by one in a fixed order. The module waits for an input from the human teammate, then reads the next action to carry out from the sequence, which specifies what the robot has to do next, along with the ID of the required object to fetch, and finally sends a request to *Motion Planner* with such parameters. On the other hand, the *User Input Module* listens to human inputs through an analog joystick module mounted on the human worker's side of the workspace. Whenever human teammates have completed an assembly step and are idle, they can press the joystick lever, thus sending a Boolean command to *Plan Manager*, which in turn publishes the next action message.

User Study

A pilot user study has been conducted with $S = 12$ subjects (10 males and 2 females), aged between 24-38 (*Avg* = 27.08, *StdDev* = 3.62), with little to no prior experience with MR-HMDs. Each participant was requested to complete the collaborative assembly task with Baxter in two distinct conditions, namely:

- *C1* wearing the HMD, and with the *mr* channel activated, i.e., involving the whole communicative act C^*;
- *C2* without wearing the HMD, only with the *mov* channel active, therefore with the communicative act C_{mov} only.

In order to avoid introducing unwanted biases, each participant performed the two trials in reverse order with respect to the predecessor.

Before beginning the trial in condition *C1*, participants were provided with a brief tutorial to understand how to navigate within the HoloLens menus. Once settled and comfortable, they were asked to stand in front of the robot, open the Unity application from within the main menu, and then proceed to look at the 2D marker attached to the Baxter's front, thus enabling the spawning of the various holograms. They were then asked to align the various wooden pieces, tools and components involved in the assembly task with their virtual counterparts visible in the holographic scene. After that, they were able to start their trial by pressing the joystick's lever. During the trial performed in condition *C2*, the various items and components were adjusted and aligned for each participant in advance.

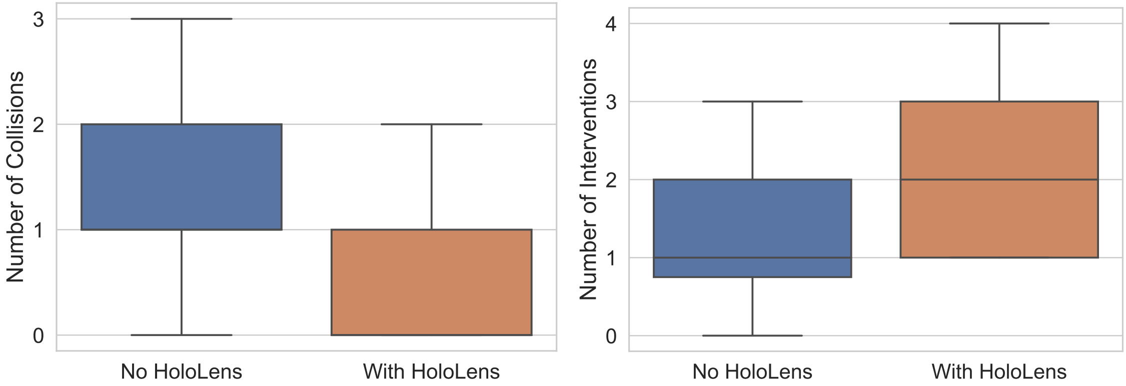
During trials, in order to evaluate hypothesis *H1.a*, the time required to perform each collaborative step, and the total time necessary to complete the assembly task were measured. Moreover, video recordings of the experiments were made, and later used to extract metrics useful to evaluate hypotheses *H1* and *H2*. In particular, within each trial we counted:

- For *H1*, the number of unintentional collisions occurring between Baxter and the participant, or involving the robot and wooden chair, throughout the assembly process. In this scenario, a collision was considered as such only if the robot bumped into the participant, halting its execution, or in case the robot collided with the half-assembled chair, possibly disrupting the state of the assembly process. This metric gave us a measure of the individuals' awareness of the robot's actions and movements, a fundamental aspect to achieve fluent collaboration;
- For *H2*, the number of times the participant proactively intervened to help the robot teammate complete an action successfully. This included situations where the human operator realized that the robot's imminent action was going to fail, and they proactively intervened to correct its course, e.g., by adjusting the pose of an object to facilitate grasping.

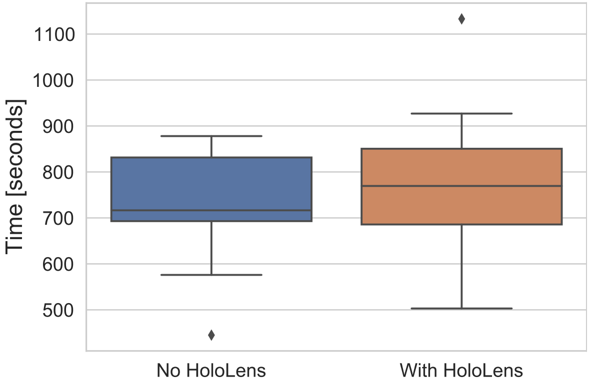
3.2 Results

The results obtained in the user study are summarized in Figure 3.4, where data related to collisions, interventions and time are presented.

The box plots in Figure 3.4a show the measured number of unintentional collisions during each trial. Due to the limited size of the shared workspace, collisions were more likely to happen during later stages of the assembly process, when the table was mostly cluttered by the half-built chair. Participants performing the experiment in condition *C2* were more prone to be caught off-guard by the robot movements, thus resulting in an increased number of collisions. On the contrary, participants performing the trial in condition *C1* were more aware of their robot teammate's presence. Therefore, they knew better how to position themselves



(a) Number of collisions per trial (b) Number of interventions per trial



(c) Time required to complete the task

Figure 3.4 Results of the user study in the two experimental conditions. (a) and (b) show improvements in the trials carried out with the MR equipment, i.e., the HoloLens: at least half the population managed to complete the collaboration without experiencing collisions and at the same time participants felt more inclined to intervene and assist the robot; conversely, (c) shows no significant improvement in the adoption of the MR feedback in terms of task speed execution.

around the workspace while the interaction unfolded. From the box plots in Figure 3.4a, one can notice that the data do not follow a normal distribution, therefore hypothesis *H1* has been evaluated through a non-parametric test, namely the one-tailed Wilcoxon signed-rank test [172]. The test provided a statistic $W = 3.5$ and $p < 0.01$. This result was compared with the critical value W_c extracted from the table in [171] by fixing the significance level $\alpha = 0.05$ and the number of participants S, thus yielding $W_c = 17$. Since $W < W_c$, the test allowed us to reject the null hypothesis and conclude that there is a statistical difference between the two experimental conditions.

Similarly, Figure 3.4b depicts the measured number of proactive interventions per trial. In this case, participants in condition *C1* were able to understand the robot intentions in advance, and thus were more likely to help Baxter grasp an ill-positioned tool or a wooden piece by slightly adjusting its position on the table. Conversely, in condition *C2*, users could not anticipate what the robot was about to do and, as such, felt less prone to intervene and help it. This resulted also in the robot possibly failing a grasping action, consequently damaging the overall quality of the interaction. As before, hypothesis *H2* has been evaluated through the Wilcoxon test, which yielded a statistic $W = 9.5$ with $p < 0.04$. Comparing W with W_c enabled us to reject the null hypothesis and to state that the two experimental conditions are statistically different.

Figure 3.4c, however, shows that the amount of time required to complete the assembly task is not considerably affected by the adoption of the MR setup. This is also confirmed by the test, which does not yield significant results ($W = 27.0$, $p = 0.381$). For this reason, we could not reject the null hypothesis for *H1.a*. This behavior could be explained as a result of the collaborative task's structure, our explanation being as follows. On the one hand, robot actions are scripted, therefore the time required for their completion is unchanged with respect to the experimental condition while, on the other hand, each participant performed their assigned assembly steps in roughly the same amount of time over the course of the two trials. Therefore, although the number of collisions is indeed higher in condition *C2*, the seconds lost by the robot to recover after every accidental bump do not represent a significant deterioration of the overall task execution time over the course of a 10 or more minutes assembly process.

3.3 Discussion

This Chapter introduced *MR-Space*, an holographic communication formalism which represents the common thread of this Ph.D. work. Such formalization could not be achieved without the prior development of an analytical tool, namely the *C-Space*, that enables modeling communication acts in generalized interactive contexts. The proposed holographic formalism, which is exclusively investigated, for the present Chapter, to enable intuitive RTH communication, represents a novel, more expressive and straightforward strategy compared to previous approaches, aimed at conveying robots' intentions to human teammates through holographic cues. In particular, the adoption of dynamic, digital overlays offered a new communicative layer, where robots can project their imminent state trajectories, as well as convey beliefs about objects they intend to manipulate, providing additional insight to the operator interacting with the machine. Such formalism has been translated into a practical implementation of software architecture, named *MR-HRC-V1*, which employs HMD devices to achieve the envisioned holographic RTH communication scheme in scenarios involving a human operator and a fixed manipulator robot. A preliminary used study has been conducted using this implementation in a scenario of collaborative assembly and several objective metrics, among those discussed in Section 2.4, have been evaluated. To this regard, the adoption of the holographic communication scheme highlighted a positive effect in terms of fluency throughout the interaction, with a reduced number of accidental collisions occurring between human and robot. At the same time, the aspect of team coordination has experienced improvements, with the human teammate behaving more proactively to support the robotic counterpart in critical situations, thanks to the intuitive, holographic cues foreshadowing the robot's intentions. Nevertheless, the only aspect which did not register any improvement was the collaboration time, possibly due to the particular nature of the experimental scenario.

Although the user study yielded positive results, its preliminary nature required further understanding to assess the possible beneficial impact of MR-based communication on the collaboration. Additionally, only some task-related metrics were evaluated during the present study, with no assessment whatsoever in terms of users' experience, which could highlight more subjective quantities, such as trust towards the robot or perceived safety. To this extent, a more extensive user study has been conducted under a similar collaborative assembly scenario, with the final aim of generalizing the aforementioned results, in light of a broader population, and in order to evaluate the holographic communication scheme for its UX. Such extensive user study and the corresponding results are the topic of the next Chapter.

Chapter 4

Extensive Evaluation and Comparison with Other Holographic Communication Strategies

Elaboration and integration of an article submitted to:
Robotics and Autonomous Systems

The results presented in the previous Chapter provided a first overview of the potential of MR-based communication in HRC. Nevertheless, the limited population involved in such investigation made it impossible to consider the results as final. As such, a more extensive assessment was necessary, combining objective metrics with subjective ones, to obtain a fairer estimate of the effectiveness of holographic communication, both in terms of team efficiency and perceived UX. To this extent, the present Chapter details an extensive user study carried out with 60 participants in a similar scenario of collaborative assembly, involving a human operator and the robot Baxter. The objective of such extended study is two-fold:

G1 Attempting to generalize the results obtained in Chapter 3, in light of a larger population and more exhaustive objective and subjective metrics;

G2 Evaluating the proposed DHT strategy against other holographic communication strategies from the literature, to provide a thorough comparison, possibly highlighting the advantages of the solution proposed during this thesis's work.

The comparison between holographic communication strategies is justified in light of the principle expressed in (3.2), which states that, given a particular channel (i.e., MR, in the present context), any piece of information I may be rendered in multiple ways, some more effective and expressive than others. As a result, it is necessary to find out whether the proposed DHT approach offers more meaningful and intuitive communicative acts, compared to similar strategies aimed at the same goal. To this regard, two additional communication alternatives have been taken into account from related literature. Then, to achieve a thorough and fair comparison among the three, the Chapter starts by presenting such two alternatives in detail, providing a complete modeling of their communicative acts in light of the *MR-Space* formalism. Following after that, the extended experimental campaign is described, with a particular focus on the in-depth aspects that have been investigated in the present study, with respect to the previous one. Finally, the results are presented and discussed, highlighting consistencies with preliminary results, and remarking the practical limitations, which will be addressed in the next chapters.

4.1 Methods

4.1.1 Formalization of Alternative Communication Strategies

Before introducing the two modalities that have been compared with the DHT approach, a small note is necessary, to avoid confusion in the reader. Throughout the present Chapter, the DHT scheme will be referred to as mr_3, as opposed to the two alternatives, which will respectively be identified as mr_1 and mr_2, and are introduced in the following two paragraphs. Such consistency in the nomenclature ensures more clear reading, while at the same time behaving coherently with the principle (3.2). Specifically, the channel considered for communication is MR in all three cases, but the approaches differ in how information is translated and rendered as anticipatory holographic cues.

Static Trail of Robot States

The first holographic communication modality considered, denoted as mr_1, draws inspiration from [137]. This communicative interface displays, for any particular robot's motion, a static and continuous trail of holograms, representing the sequence of states that the robot will go through while executing the intended trajectory (Fig. 4.1a.). Whenever the upcoming action requires the robot to fetch or manipulate objects in the shared workspace, the interface also depicts the robot's corresponding beliefs, i.e., holographic versions of the objects involved in the action projected at their expected pose. From a formal point of view, we can express the overall communicative act conveyed by this interface as

$$C_{mr_1}\left(I,\boldsymbol{t}_{mr}\right)=\sum_{t=t_{mr,s}}^{t_{mr,e}}\boldsymbol{\tau}\left(t+\Delta t\right)\,\cup\,\sum_{k=1}^{D}\boldsymbol{\xi}_k\left(t_{mr,e}+\Delta t\right), \tag{4.1}$$

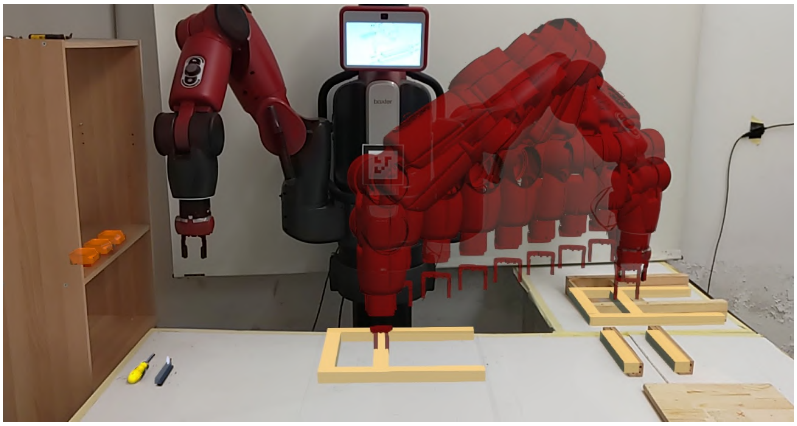
where $t_{mr,s}$ and $t_{mr,e}$ follow from the relationship expressed in (3.10), $\boldsymbol{\tau}$ defines the specific trail associated with the robot's state trajectory (i.e., sequence of joint positions), D specifies how many objects are involved in the robot's action, and each $\boldsymbol{\xi}_k(t_{mr,e}+\Delta t)$ yields the static anticipatory belief describing the expected pose of the kth object being manipulated by the robot. In order to avoid cluttering the user's view, only the beliefs corresponding to the final instant $t_{mr,e}$ of the anticipated robot's trajectory are shown, due to their relevance in the collaborative task, as they forecast how objects are delivered in the working environment by the robot or handed over to the human teammate in case of handover operations. Conversely, the use of the summation in (4.1) stresses the fact that all states $\boldsymbol{\tau}(t_i)$ describing the robot's trajectory are statically rendered at the same time to give the human operator a full picture of the 3D volume swept by the action of their robot counterpart.

Static Final Robot State

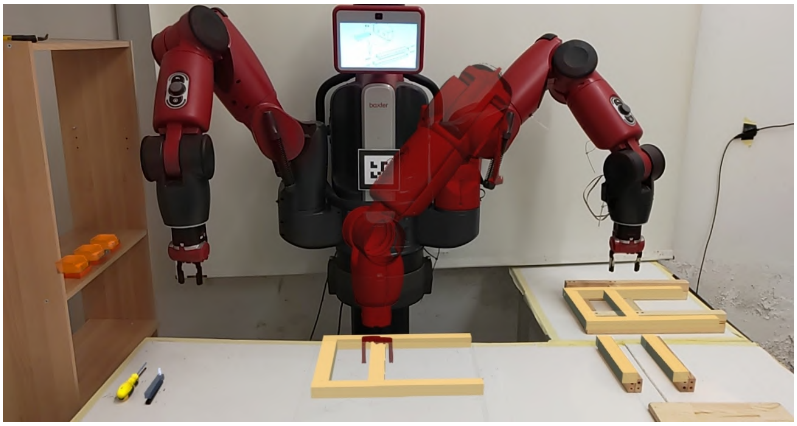
The second communication modality, denoted as mr_2, is instead inspired by [117]. In this case, the holographic cues only statically show the final state $\boldsymbol{\tau}(t_{mov,e})$ of the robot's trajectory $\mathbf{T}(\boldsymbol{t}_{mov})$, as shown in Fig. 4.1b. As an example, if the robot plans to perform a handover operation, this communicative interface only displays the state in which the robot will perform the said operation, without any detail on the intermediate trajectory. Coherently with the previous case, the holograms of the objects involved in the robot's action are displayed at their believed final pose, providing additional insight to the human teammate. Formally, the communicative act conveyed by this second interface can be modelled as

$$C_{mr_2}\left(I,\boldsymbol{t}_{mr}\right)=\boldsymbol{\tau}\left(t_{mr,e}+\Delta t\right)\,\cup\,\sum_{k=1}^{D}\boldsymbol{\xi}_k\left(t_{mr,e}+\Delta t\right). \tag{4.2}$$

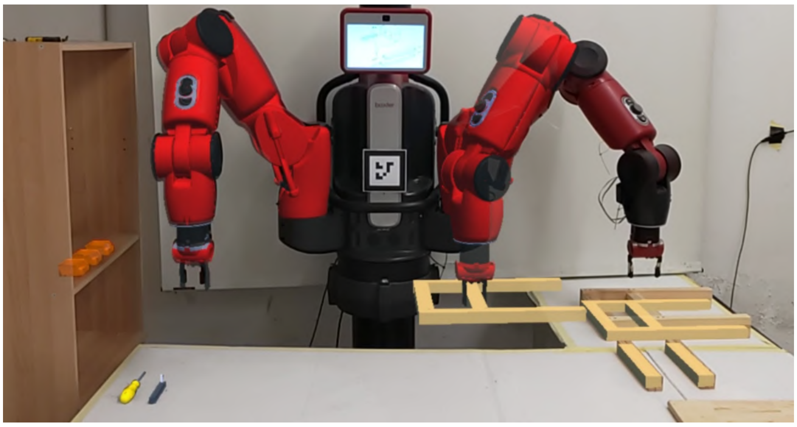
The definition of C_{mr_2} entails that such a communicative act conveys a subset of the anticipatory information generated by C_{mr_1}, providing the human teammate with just minimal cues about the robot's intentions. However, unlike the previous alternative, the current communication modality generates a more straightforward and less cluttered visual feedback and may therefore result in a more intuitive option.



(a) Static trail of robot states



(b) Static final robot state



(c) Dynamic holographic trajectory

Figure 4.1 First-person views depicting the three types of holographic communication modalities explored in this study, as they appear inside the *MR-Space*. The three modalities anticipate the upcoming robot's intentions by overlaying different forms of holographic cues to the scene observed by the human teammate.

Combined Communicative Acts

From the previous paragraphs, it follows that, to maintain consistency in the nomenclature, the holographic, communicative act resulting from the DHT approach has to be renamed C_{mr_3}, while still maintaining the same definition as (3.11).

In light of the *G2* objective, that is assessing which, among the three holographic communication alternatives, yields the most intuitive and expressive intention cues, it is necessary to define the combined communicative acts, resulting from the interplay of *mr* channel, declined through the corresponding rendering function, and *mov* channel. As such, the following relationships hold for the rest of the Chapter:

$$C_i^* = C_{mov} \cup C_{mr_i} \quad \forall i \in \{1, 2, 3\} . \tag{4.3}$$

4.1.2 Experimental Design

In this Section, the experimental hypotheses are presented, along with the collaborative scenario designed to test them. Given the multiple similarities with respect to the previous experimental setup described in Section 3.1.4, only relevant differences will be highlighted in the following paragraphs.

Hypotheses

As stated at the beginning of the Chapter, the experimental phase serves two purposes. On the one hand, *G1* aims to confirm whether MR is a suitable channel for RTH communication in HRC, and whether it improves team efficiency by enabling human operators to intuitively understand robot's intentions, as suggested by previous preliminary results. To this end, two experimental conditions, akin to those considered for the previous study, have been compared, with multiple hypotheses taken into account. As a matter of fact, the first hypothesis has been left unchanged for the present experimental campaign, attempting to generalize and corroborate the results of the previous study. As such, the present *H1* argues that:

H1 The communicative act C^*, resulting from combining *mr* and *mov* channels, lowers the amount of unintentional collisions occurring during the collaboration.

This particular aspect is crucial for a fluent interaction between human and robot, as several works discuss how poor communication can lead to agents hindering one another [58, 73] or to accidents [94]. Similarly, coordination between team members, that is, the ability to synchronize actions and carry out joint activities smoothly [70, 150], has been evaluated as well. However, the second hypothesis for the present study has been adjusted to be more meaningful in a context requiring explicit interaction between human and robot. As such, *H2* states that:

H2 The communicative act C^* reduces communication ambiguities in joint actions, e.g., handovers, thus lowering the rate of *failed* interactions.

Finally, the third hypothesis has again been conserved, in an attempt to evaluate whether MR-based communication can positively influence team efficiency in terms of task completion time [85]. Consequently, *H3* for the present study claims that:

H3 The communicative act C^* speeds up the collaboration process.

Aside from the assessment of task-related aspects, *G2* aims at providing a thorough comparison between the three holographic communication schemes, evaluating their capabilities in light of the aforementioned hypotheses, while at the same time appraising participant's perceived UX when experiencing the various forms of RTH holographic cues during collaboration.

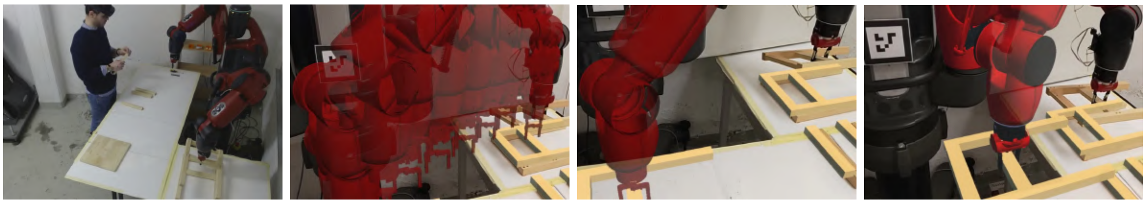
Collaborative Assembly

Consistently with the previous campaign, the current experimental scenario involved a collaborative assembly of a wooden IKEA stool[1]. As before, the human was tasked with the actual assembly process, while the robot teammate collaboratively provided the various tools and components when needed. According to the definition given in [1], such a form of interaction falls under the *cooperation* category (level 3 collaboration), where humans and robots work together on the same goal in a shared workspace, but *concurrent* activities are not strictly required. Differently from the old setup, the shared workspace has been rearranged, as shown in Fig. 4.1, adding a side-way bookshelf where various boxes holding assembly components were stored, in order to leave the majority of the table as free space for the assembly process. Throughout the interaction, the robot employed the bookshelf to retrieve components' boxes when needed, and subsequently put them away, back to their designed spot when no longer necessary.

The task followed the same structure as described in Section 3.1.4, with a fixed sequence of collaborative steps in which the robot performed pick-and-place or handover actions, retrieving the tools or components necessary for the human to proceed with the assembly. However, the role of the button located on the participant's side of the workspace has been

---

[1] IKEA Oddvar stool, https://www.ikea.com/gb/en/p/oddvar-stool-pine-20249330/



(a) External view (b) Static trail (c) Static final state (d) Dynamic trajectory

(e) External view (f) Static trail (g) Static final state (h) Dynamic trajectory

Figure 4.2 Snapshots from the user study depicting subjects involved in the collaborative assembly with the robot Baxter. Specifically, Fig. 4.2a and Fig. 4.2e show different instants of the assembly process, respectively a *delivery* operation performed by the robot to supply a wooden piece and a later *handover* where the robot provides the hex key to the user. For each action, the three pictures on the right represent the holographic projections that the user is experiencing inside the *MR-Space*. In particular, the three pictures depict the three alternative communication strategies introduced in Section 4.1.1, as envisioned by the *MR-Space*. It is possible to note that Fig. 4.2d and Fig. 4.2h, corresponding to the mr_3 strategy, anticipate the robot's imminent intentions as dynamic holographic cues, coherently with the analytical formalization of the communicative act C_{mr_3}.

altered in the current campaign. Since the pace of the collaboration was established by the robot's actions, the button could now be employed by individuals to pause the robot teammate. Specifically, pressing the button ensured that the robot completed its current action, then it idly waited until the user triggered the button again. This allowed humans, if they lagged behind the robot's pace, to pause their teammate for completing the current assembly step before proceeding to the next.

User Study

Within the collaborative scenario described above, an experimental campaign has been conducted involving $S = 60$ participants (47 males and 13 females), aged 20-32 ($Avg = 23.86$, $StdDev = 2.37$), each with very little or no prior experience with wearable MR visors. Authorization to proceed with the user study has been requested and issued by the ethical committee for research at the University of Genoa before the start of the campaign (protocol n. 2021/65 - November 18, 2021). The population size S has been carefully chosen through *statistical power analysis*, considering a significance level $\alpha = 0.05$, a statistical power set to 90% and an expected effect size of 0.8. Particularly, this latter value has been estimated from the preliminary results obtained in the user study described in the previous Chapter.

In order to evaluate the effect of each communicative act C_i^*, and highlight the difference with respect to C_{mov} alone, participants have been randomly divided in three groups of twenty individuals each. Consistently with the previous study, all subjects were tasked with completing the collaborative assembly with Baxter in two separate conditions, namely:

E1 wearing the HMD and with the *mr* channel active, therefore experiencing the whole communicative act C^*, declined in the form C_i^* according to the participant's membership to the *i*th group;

E2 experiencing the communicative act C_{mov} only, thus without wearing the HMD and with just the *mov* channel.

As before, in order to minimize undesirable biases, two consecutive participants from the same group performed their two trials in reverse order. As such, within each group, ten individuals perform their first trial in condition *E1*, while the remaining ten start off with condition *E2*.

Following the methodology of the previous study, participants under condition *E1* were given some time to accustom to the holographic visor before starting the experiment, as well as some tips to interact with the HoloLens menus. Then, they could proceed with their trial, experiencing one of the three holographic communication modalities throughout the collaboration, according to their group membership. Upon concluding their experiment, participants were asked to fill the UEQ to evaluate their subjective UX with that particular form of holographic communication.

Aside from assessing subjective UX of the three communication strategies via UEQ, video recordings of the experiments were made as well, to extract useful objective metrics in the post-campaign analysis. Therefore, unlike the previous study, a combination of self-assessments and task-related metrics has been considered [12] to obtain a fairer and more complete appraisal of RTH holographic communication, declined in its various forms. Regarding the metrics used to test the experimental hypotheses, *H1* has been evaluated coherently with the previous study, as the assessment was aimed at corroborating preliminary results. On the other hand, hypothesis *H2* and team coordination in general have been evaluated by counting the failed interactions between agents, due to the individual misreading robot's intentions. In the current context, two types of failed interactions were considered, that could occur during the assembly process:

- failed handovers, resulting from the participant being unable to retrieve the supplied item from the robot's hand, causing it to fall on the table or on the ground;
- failed retrieval of assembly components, e.g., dowels and screws, from the small box before the robot brought it away from human reach. This error could happen since the robot's actions had a fixed duration, and the participant was expected to retrieve the components timely.

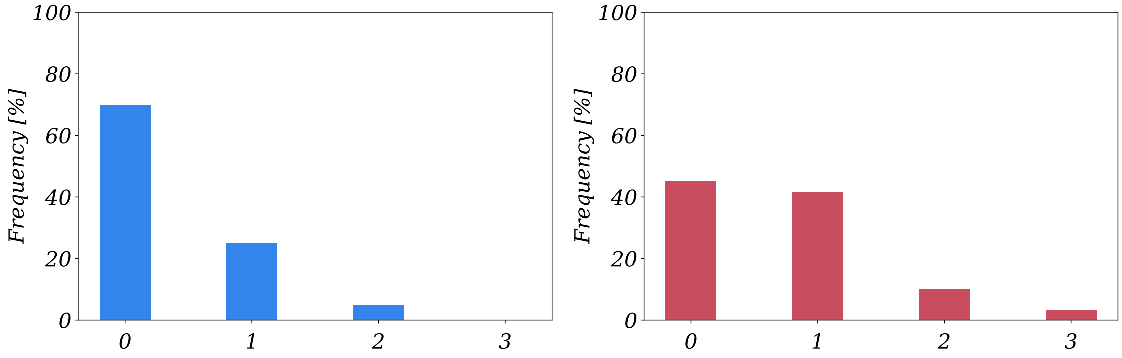
Finally, hypothesis *H3* has been evaluated consistently with the previous study, by measuring the overall time required to complete the collaborative process. In particular, the time was measured as the interval between the first robot's action, triggered by the user pressing the button to signal the start of the interaction, and the completion of the assembly process, again signaled by the user pressing the button one last time.

4.2 Results

This Section illustrates the experimental results obtained from the user study described in the previous paragraphs. In a preliminary phase, the main focus of the analysis is comparing the combined communicative act C^* with respect to the sole act C_{mov}, in terms of the three hypotheses *H1*, *H2* and *H3*, regardless of the holographic communication modality experienced by the single participant. Then, the three communication modalities are evaluated separately in terms of the aforementioned hypotheses and the perceived experience of the three groups of participants as measured by the UEQ.

4.2.1 Communicative Act C^* vs C_{mov}

Fig. 4.3 depicts the results obtained by the post-campaign analysis of the videos, in terms of unintentional collisions happening during the experiments. For the current investigation, conditions *E1* and *E2* are compared ignoring the different types of holographic communication perceived in the three groups, to evaluate whether the presence of the *mr* channel was effectively beneficial for the collaboration. The histograms in the Figure represent the percentage of individuals who completed the collaborative assembly experiencing a particular number of accidental collisions with the robot, in the two experimental conditions. It can be observed that around 70% of the participants subjected to the communicative act C^* managed to complete the collaboration with no collision at all, compared to around 40% of the individuals in case of the communicative act C_{mov} alone. From a practical point of



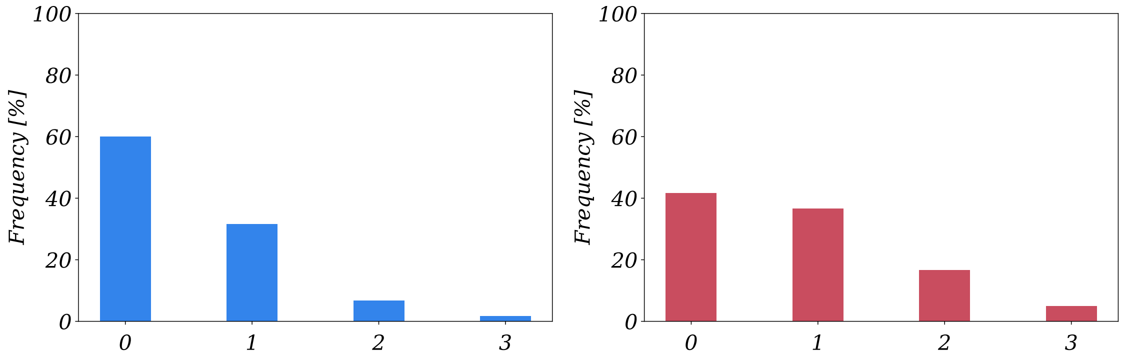
(a) Collisions with communicative act C^* (b) Collisions with communicative act C_{mov}

Figure 4.3 Histograms depicting the number of accidental collisions occurring with the two different communicative acts.

view, the presence of holographic RTH communication made participants more aware of the upcoming robot's actions, generally improving their reaction time and limiting the times in which they were caught off-guard by robot's movements, which could end up in a possible collision between the two.

Since data in Fig. 4.3 are not normally distributed, and following the methodology of the previous study, the one-tailed Wilcoxon signed-rank test has been employed to evaluate hypothesis *H1*. However, due to the large population size S, we had to refer to the asymptotic properties of the Wilcoxon distribution to extract a meaningful interpretation of the test statistics. The W distribution tends to a normal one with well-known mean and variance for large populations [171], therefore the resulting statistics ($W = 570$) was converted into the corresponding z-score $= -2.54$, which, in turn, yielded a p-value < 0.01. As such, it was possible to conclude that the two distributions were statistically different. An additional post-hoc power analysis carried out on these results yielded an experimental power of 93%, providing further support to the validation of hypothesis *H1*.

Similarly, Fig. 4.4 shows the results related to the failed joint actions during the experiments. The histograms represent the percentage of participants who experienced a particular number of failures during the collaboration. Again, it can be noted that the percentage of individuals who managed to complete the collaboration without experiencing failures is greater when the communicative act C^* is employed. In this case, the presence of the RTH holographic communication enabled the participants to infer which component or tool Baxter was going to deliver and how the robot intended to perform the interaction, increasing the likelihood that the joint action unfolded smoothly. On the contrary, the lack of visual communication led participants not to understanding the robot's intention, resulting, for



(a) Failures with the communicative act C^* (b) Failures with the communicative act C_{mov}

Figure 4.4 Histograms depicting the number of failed interactions happening with the two different communicative acts.

example, in them unintentionally dropping the tool which was offered by Baxter during a handover.

As for the previous case, the assumption of normal distribution could not be made for the data depicted in Fig. 4.4. Consequently, to evaluate hypothesis *H2*, the one-tailed Wilcoxon signed-rank test has been used, yielding a statistics $W = 584$, equivalent to a z-score $= -2.43$ and with a corresponding p-value < 0.01. This result implied the statistically significant difference between the two distributions. A supplemental post-hoc power analysis corroborated the results, yielding an experimental power of 89%, which is comparable to the 90% threshold established through the *a-priori* assumptions.

Finally, Fig. 4.5 depicts the results related to the total time needed to complete the collaborative assembly in the two experimental conditions. From the box plots, it is possible to note that the presence of the holographic channel *mr* was not effective in speeding up the collaboration process. To confirm this hunch, a T-test was performed on the two distributions, which were found to be normal through Shapiro-Wilk test [146] (p-value > 0.5 for both distributions). The T-test returned a p-value > 0.1, thus indicating no significant difference between the two experimental conditions and therefore it was not possible to confirm hypothesis *H3*. The ineffectiveness of the *mr* channel in reducing the overall execution time can be explained by the experimental scenario itself, in which collisions between humans and robots, and possible failed interactions, while worsening the quality of the collaboration, did not necessarily slow down the assembly process significantly in a trial lasting ten or more minutes. This current limitation has been addressed in the upcoming chapters, hypothesizing that the holographic communication could play a more effective role in speeding up the collaboration in those scenarios of concurrent interaction in which the

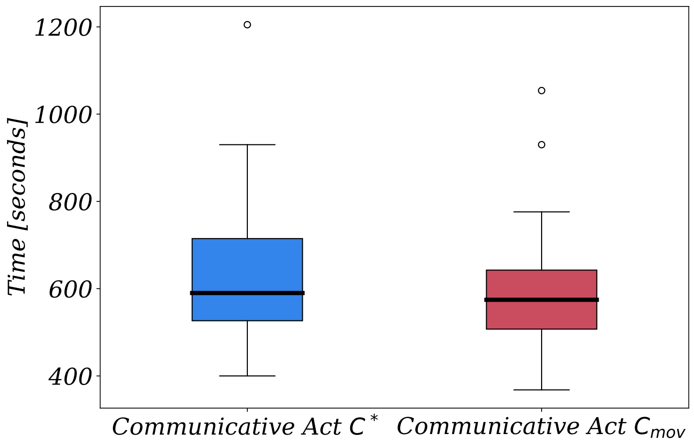
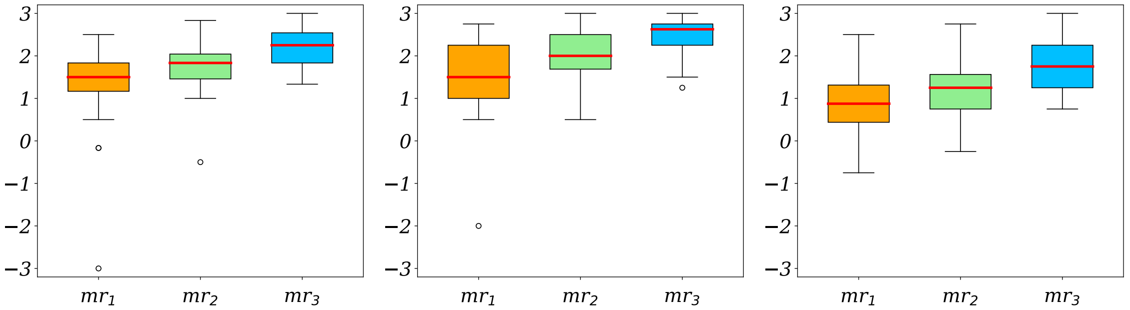


Figure 4.5 Comparison of the time required to complete the collaboration with the two communicative acts. The median value for each distribution is plotted as a thick black line.

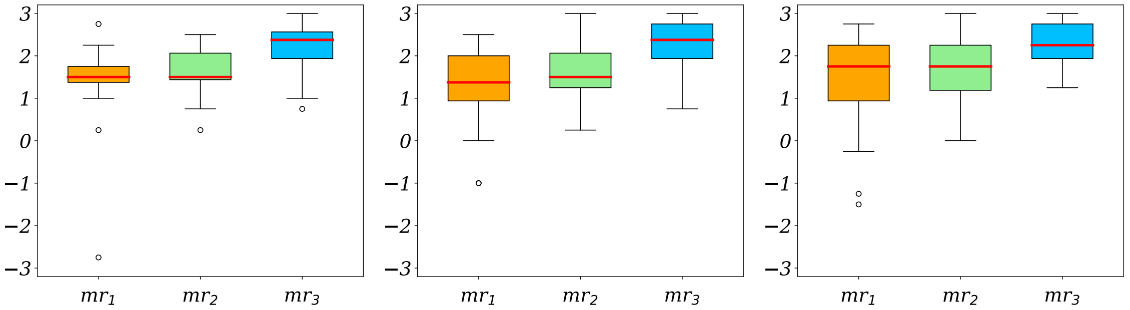
robot can modify its decision process at run time, and carry out independent, simultaneous actions with respect to the human teammate.

4.2.2 Comparison Between Communication Modalities

According to the initial findings, the communication act C^* provided a generally more seamless interaction between participants and the robot, therefore the next step in the analysis consisted in evaluating the UEQ results to identify the communicative interface yielding the most natural and intuitive UX. Participants responded to each of the twenty-six questions proposed in the UEQ with an integer score ranging from -3, corresponding to an *extremely negative* evaluation, to $+3$, which was instead associated with an *extremely positive* score. Results were then rearranged according to the scales proposed in [144], averaging the twenty-six scores of each participant into six global marks, reflecting the evaluation scales of the questionnaire. Fig. 4.6 reports the questionnaires' results, grouped by evaluation scale and communication modality. By inspecting the plots, it is possible to note that the communication modality mr_3 received the best scores on all scales. In order to test the significance of the results, analysis of variance (ANOVA) [143] has been employed on the six triplets. Such a test, however, requires that all distributions involved are assumed normal, therefore a normality check through Shapiro-Wilk test has been conducted first, obtaining p-values > 0.1 in all cases. Subsequently, the ANOVA yielded p-values < 0.01 on each evaluation scale. Moreover, post-hoc analyses carried out via T-test between the pairs mr_1-mr_3 and mr_2-mr_3 yielded that the scores gained by mr_3 were statistically different with respect to the other two modalities in each of the six scales, confirming what could be



(a) Attractiveness (b) Perspicuity (c) Efficiency

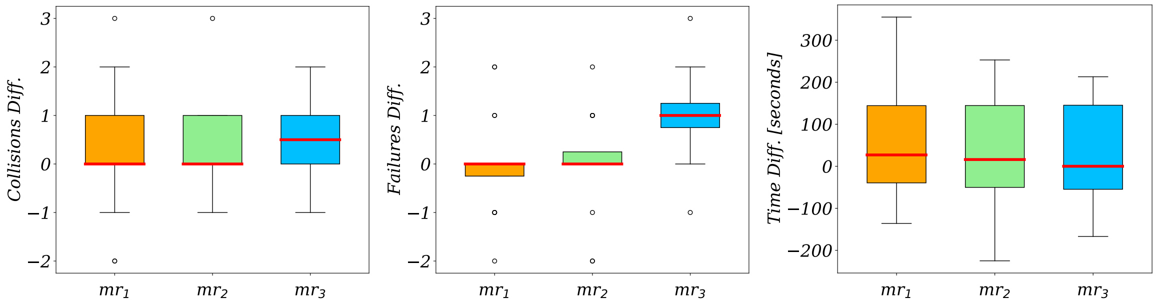


(d) Dependability (e) Stimulation (f) Novelty

Figure 4.6 Measured UEQ scores on the six evaluation scales. The median value for each distribution is plotted as a red line. The DHT modality (mr_3) received overall better scores on every scale, with particularly high mean values in the *perspicuity* and *dependability* scales, compared to the other two modalities.

observed in the plots. Additionally, a supplemental post-hoc power analysis performed on the results of the ANOVA tests returned an experimental power strictly greater than 90% on each evaluation scale, in accordance with what had been established during the experimental design.

With these results in mind, we could conclude that participants preferred the DHT modality, both in terms of appeal of the holographic cues (higher scores in *attractiveness* and *stimulation* scales) and usability of the communicative interface (higher grades in the *efficiency* scale). Particularly high median values have been achieved in the *perspicuity* ($Med = 2.6$) and *dependability* ($Med = 2.4$) scales, suggesting that subjects found the dynamic holographic cues generated by the communicative act C_{mr_3} more intuitive and dependable, and thus more meaningful throughout the collaborative process. It is noteworthy that mr_1 and mr_2 received comparable results with equivalent median values in the *dependability*, *stimulation* and *novelty* scales, indicating no overall preference between the two. This result could be a hint that the static holographic communication is perceived as not particularly



(a) Diff. in number of collisions (b) Diff. in number of failures (c) Time difference

Figure 4.7 Results of the differential analysis performed on the quantitative metrics used to evaluate *H1*, *H2* and *H3*. The box plots depict the differential distribution computed between condition *E2* and *E1* for each participant, and are grouped per type of communication modality. The median value for each distribution is plotted as a red line.

pragmatic and expressive of the robot's intention. Furthermore, mr_1 has been generally perceived as a less efficient and intuitive communication modality with respect to the other two, due to the cluttered visual feedback provided by the communicative act C_{mr_1}, as well as less attractive.

In order to corroborate such findings and understand whether the preferred UX yielded by mr_3 also resulted in the most seamless and natural form of collaboration between humans and robots, we proceeded with a differential study of the quantitative metrics extracted while analyzing the videos of the experiments. For each participant, we computed the difference, respectively in terms of the number of collisions, failures and time taken to complete the assembly, between condition *E2* and *E1*, to highlight which, among the three holographic communication modalities, ensured the most noticeable improvement in the collaboration with reference to the three hypotheses *H1*, *H2* and *H3*. The results of this investigation are reported in Fig. 4.7. In particular, Fig. 4.7a shows that no significant disparity could be observed in terms of collisions difference between the three communication modalities. This result implies that the communicative acts C_{mr_1}, C_{mr_2}, and C_{mr_3} behaved comparably, on average, with the three different modalities equally anticipating the robot's imminent trajectories and improving participants' reaction time, resulting in a reduced number of unintentional collisions between humans and robots in condition *E1*. To confirm such a result, a normality check on the three distributions has been performed through the Shapiro-Wilk test. Since at least the first two distributions could not be assumed normal (p-values < 0.05), we opted to use the Kruskal-Wallis test [83], a non-parametric version of the ANOVA. The test yielded a p-value > 0.2, as such we were able to confirm the non-significant difference between the three distributions.

Similarly, Fig. 4.7c shows that no significant variation could be observed in terms of the difference in task completion time when adopting one holographic interface or the others. This outcome is essentially coherent with what has already been discussed when comparing the communicative acts C^* and C_{mov}, indicating that the adoption of the *mr* channel in the current experimental scenario was not meaningful in speeding up the collaborative pace.

Conversely, the box plots in Fig. 4.7b reveal a variation in the differential distribution of failed joint actions when the holographic communication modality mr_3 was adopted, compared to employing the other two. A Kruskal-Wallis test carried out on the triplet of distributions highlighted a statistical difference between them, returning a p-value < 0.01. Further analysis performed through a Wilcoxon test between pairs mr_1-mr_3 and mr_2-mr_3 yielded similar p-values < 0.01, confirming that mr_3 was statistically different with respect to the other two distributions.

4.3 Discussion

Overall, the combination of subjective results derived from questionnaires and task-related metrics extracted from videos enabled concluding that the communicative act C^*, resulting from joining the *mr* channel and robot's motion one, proved more meaningful and informative of robot's intention than *mov* alone. This, in general, led to interactions between human and machine less subject to unintentional collisions and failed interactions. However, only the differential study allowed us to highlight which, among the three communication schemes, ensured the most noticeable improvement in terms of team efficiency between condition *E1* and *E2*. On the one hand, all three holographic communication alternatives proved equally useful in improving participants' reaction times and awareness, reducing the chance of unintentional collisions occurring in condition *E1*, compared to condition *E2*. This result suggests that, overall, the RTH holographic communication improves team fluency, but its effect can greatly vary depending on how the holographic cues are declined. As a matter of fact, the results in Fig. 4.7b suggested that the DHT approach, brought forth by the communicative act C_{mr_3}, was perceived as significantly more expressive in conveying the robot's intention to individuals and, as such, the corresponding joint actions were more likely to unfold flawlessly in condition *E1* than in condition *E2*. Specifically, the usage of dynamic holographic cues to preview how upcoming RTH handovers would take place made participants more responsive and ready to proactively carry out their part in the interaction, thus positively impacting on team coordination.

In line with these findings, the questionnaire results indicated that the communication scheme mr_3 was consistently considered the most suitable option for the specific experimental context. It received the highest scores across all six scales of the UEQ. Notably, the high scores for the *perspicuity* and *dependability* scales, measuring the scheme's intuitiveness and predictability, suggest a potential enhancement in the operator's trust towards the robot during interaction. This improvement might justify, in turn, the increased proactivity observed in participants when using the DHT scheme. However, future research could delve deeper into psychological aspects to precisely examine the impact of RTH communication on perceived user safety and trust, employing specialized assessment tools.

In conclusion, the extensive user study discussed in this Chapter reported findings which are consistent with those of the preliminary investigation. In particular, the goal *G1* could be achieved, as both hypotheses *H1* and *H2* could be confirmed and generalized in light of the larger population. Specifically, it has been assessed that RTH holographic communication acts as meaningful and effective channel to convey robot's intentions, while at the same time the DHT approach proposed in the previous Chapter has received better evaluation in terms of UX, compared to related approaches from the state-of-the-art. As such, the secondary objective *G2* has been accomplished as well. Nevertheless, one hypothesis could not be verified in the present study, coherently with preliminary results. The holographic communication scheme proved not effective in speeding up the collaborative pace, possibly due to the fixed configuration of the experiments, where a static sequence of actions was required to complete the task. To this regard, efforts have been made, which are discussed in upcoming Chapter, to design a more suitable experimental scenario, where robot's sequence of actions was not fixed *a-priori*, and where concurrency between humans and robots played an important factor for the overall objective of the collaboration. In such a setting, it is argued that holographic RTH communication could have a larger impact in terms of task completion time. As a matter of fact, the next Chapter will focus on extending the experimental validations to broader domains of HRC, that is when mobile collaboration among agents is required, thus relaxing the constraint of fixed manipulators adopted so far. Finally, as the potential of intuitive, holographic RTH communication has been mostly addressed in the current chapters, the discussion will proceed towards the introduction of a bi-directional communication interface, where the MR layer is complemented by a DT to achieve intuitive and efficient HTR communication as well.

Chapter 5

Mixed Reality and Digital Twin: Towards a Bi-directional Communication Layer

Elaboration and integration of an article submitted to:
33rd IEEE International Conference on Robot and Human Interactive Communication
(RO-MAN 2024)

The extensive user study described in the previous Chapter yielded promising findings, suggesting that holographic RTH communication can play an important role in improving team efficiency, in terms of fluency and coordination between agents. Additionally, it has been appraised that the DHT strategy provides more intuitive and meaningful intention cues, while at the same time being preferred by users as more engaging and efficient compared to other approaches. Nevertheless, it's crucial to acknowledge that these results were obtained within static collaborative scenarios, and the effectiveness of holographic communication was primarily evaluated for conveying intention cues in fixed manipulator robots. Additionally, a specific hypothesis could not be validated in the course of the two user studies conducted so far, possibly indicating that certain constraints introduced up to now had to be relaxed to observe a more meaningful impact of MR-based communication.

With these results in mind, the next phase of the Ph.D. work focused on precisely relaxing these constraints to observe the impact of RTH holographic communication in a more generalized domain of HRC. This Chapter introduces an experimental scenario where humans and robots are no longer confined to static workstations. Instead, they navigate a dynamic environment, concurrently executing parallel and independent tasks while collaborating and interacting under specific circumstances. This scenario allows for the relaxation of the constraint outlined in Section 3.1.2. Notably, having a collaborative robot capable of moving

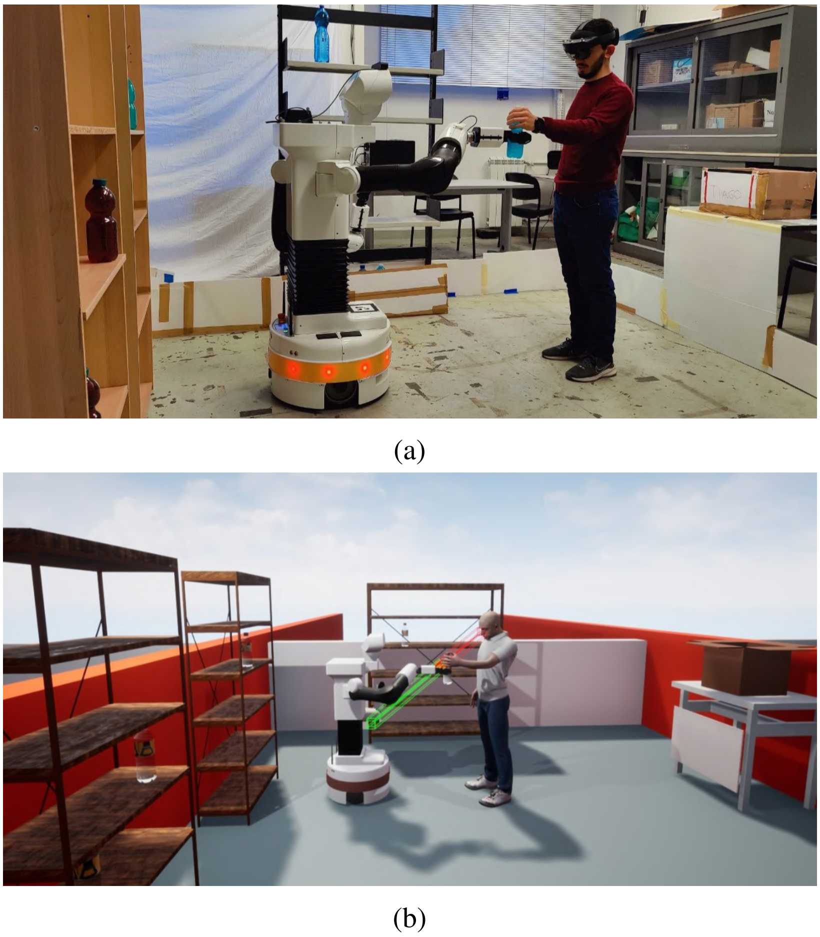


Figure 5.1 The HRC scenario is shown in Fig. 5.1a, providing context for Fig. 5.1b, where the DT system monitors the collaboration, replicates agents' state and infers handover intentions from a combination of gaze and postures.

throughout the environment, potentially utilizing holographic communication to convey imminent navigation trajectories, means that the $\boldsymbol{x}(t)$ component of the robot's state can no longer be ignored. Therefore, this Chapter refers to the robot's state as the complete state $\boldsymbol{\tau}(t)$ expressed in (3.4). Moreover, another constraint relaxed in this investigation is the fixed sequence of robot actions. In this study, the robot can now choose among multiple actions that can be executed without a specific order. As such, the role of RTH communication becomes even more relevant in foreshadowing robot's intentions.

Aside from the generalization of the collaborative scenario, which required an update of the software architecture to comply with the relaxed constraints, this Chapter marks a significant stride toward realizing the envisioned bi-directional communicative layer. It introduces a DT into the overall system, complementing the MR framework to facilitate implicit HTR communication throughout the collaborative process. To this regard, the DT is employed to replicate online the HRC process, monitoring agents in real-time, through integration of various sources of information and sensory data. Over the course of the collaboration, the DT continuously tracks agents, and extrapolates intention cues from the implicit behavior of the human operator (i.e., from a combination of gaze and body posture), activating certain interaction logic in the robot's routine to promptly react to the communication act of the human counterpart. In summary, the integration of MR for conveying robot's intentions to the human, along with a DT system tracking agents' states in real-time and utilizing this information to guide the robot's behavior, represents a first effort to incorporate these technologies into a comprehensive communication framework for HRC, where both RTH and HTR communicative aspects are taken into account.

Given these premises, the Chapter starts by detailing the updated software architecture, named *MR-HRC-V2*, which integrates the aforementioned additions, thus expanding its communicative capabilities. Then, the logistics-like collaborative environment is presented, along with its virtual representation inside the DT. The discussion proceeds by detailing the user study carried out to evaluate the expanded architecture and the communication formalism. Finally, results are presented distinguishing between the effect of MR for RTH communication and DT for HTR one.

5.1 Methods

The first update to the software architecture follows from the relaxation of the static workspace constraint. Given the robot's ability to move throughout the environment, it was necessary to provide artificial agents with a mean to convey their imminent navigation intentions, following the formalization of the *MR-Space*. As discussed in Chapter 2, the current state-of-the-art in this terms consists, on the one hand, in work from Walker *et al.* [166], where the MR layer is employed to provide a static sequence of holographic spheres representing the planned trajectory for an unmanned aerial vehicle. On the other hand, work from Gu *et al.* [62] proposed a holographic interface which users can employ to locate the robot teammate even through walls, thanks to virtual overlays which disclose the artificial agent's position and its imminent navigation direction in the environment. Nevertheless, the aforementioned approaches suffer from the same limitations discussed when DHT strategy has been proposed, as opposed to previous works. As a matter of fact, the holographic communication in these approaches is either chaotic and cluttered, or too minimal to be effectively informative of the robot's navigation intentions. As such, the strategy proposed in this work, following the

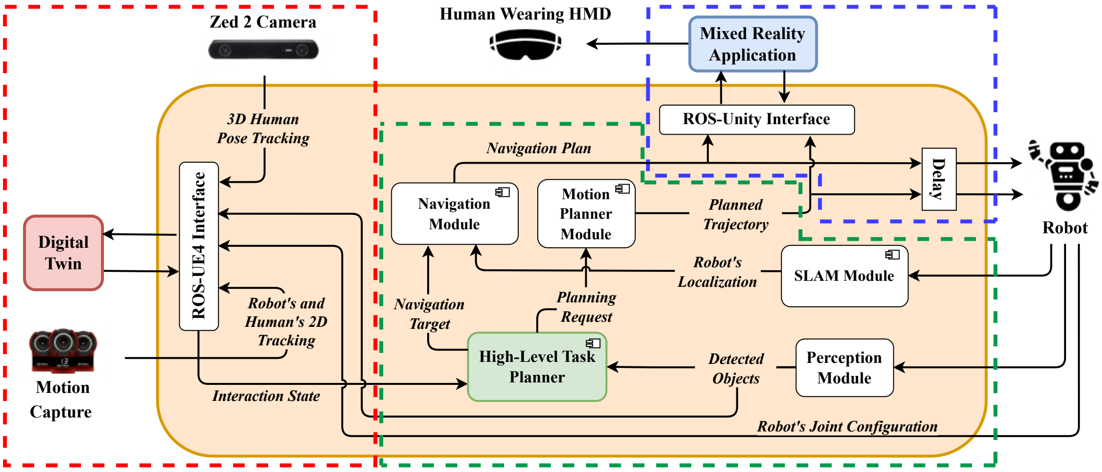


Figure 5.2 A detailed overview of *MR-HRC-V2*, with the three main blocks highlighted by the corresponding colors.

DHT paradigm, offers a more comprehensive and direct communication act, presenting the imminent trajectory of the robot through dynamic holographic cues. Specifically, considering that the robot's state now encompasses both its spatial pose $\boldsymbol{x}(t)$ and joint configuration $\boldsymbol{q}(t)$, the communication act, conveying the robot's navigation intentions, adheres to the same definition as in (3.11). Here however, the constraint $D = 0$ is set to indicate that, in general, such an act does not involve any robot's belief. The result is a more flexible solution, where human operators can easily anticipate the entire navigation trajectory and how it will unfold over time, providing additional temporal and directional information to the RTH communicative act. Although static, a screenshot of this type of holographic communication is given in Fig. 5.3a, where the robot is conveying its intentions to navigate towards the bookshelf to possibly grasp an item.

5.1.1 System's Architecture

The *MR-HRC-V2* architecture, depicted in Fig. 5.2, consists of three building blocks, that is, the *Digital Twin* component, outlined with the red dashed line, the *Mixed Reality Application*, in blue, and the core of the robot application, in green.

The part of the architecture handling robot's operations follows the classical sense-reason-act paradigm. In this context, localization and recognition of objects (see Fig. 5.3c) and robot self-localization in the environment are handled online by the *Perception* and *SLAM* modules using onboard sensors. Then, the *High-Level Task Planner* module plans the sequence of actions (e.g., *pick a bottle* or *navigate to shelf*) based on the perceived objects, the final goal, and the interaction state from *Digital Twin*, discussed later on. Actions resulting from the *High-Level Task Planner* are handled by the *Motion Planner* when they involve manipulations or by the *Navigation Planner* when they are navigation tasks.

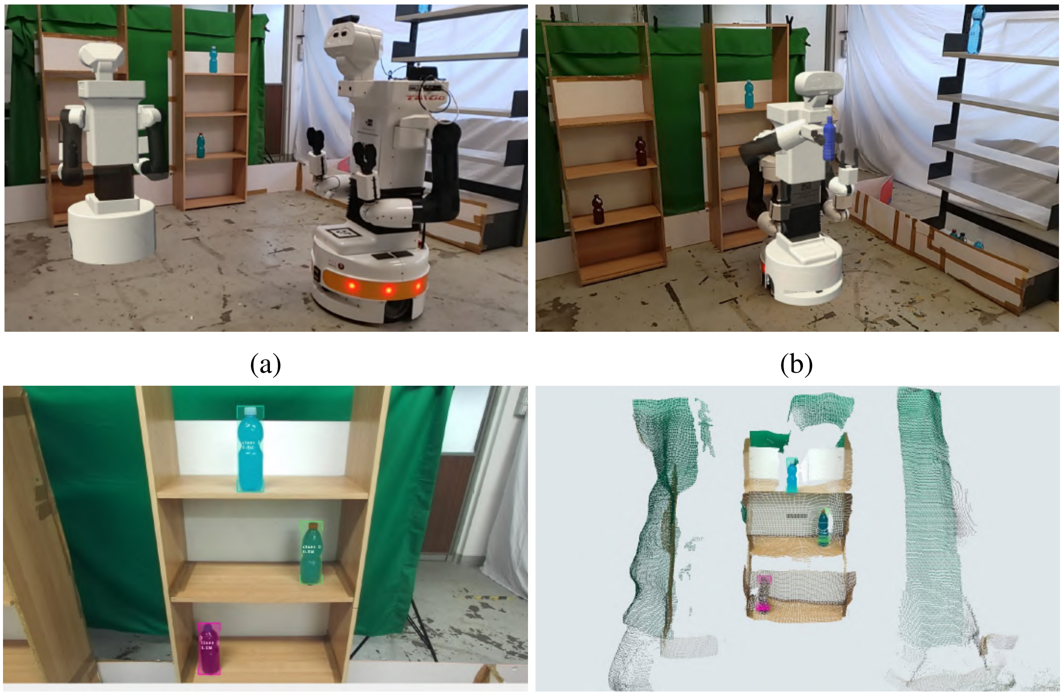
The *Mixed Reality Application* communicates the robot's intentions to the human teammate using the DHT paradigm (see Fig. 5.3a and 5.3b), with both navigation and manipulation actions that can be rendered as holographic cues.

Conversely, *Digital Twin* creates and maintains a virtual replica of the collaborative scenario, that is, robot, human, and objects, using sensory information collected by the robot and the sensors distributed in the environment, see Fig. 5.1. Furthermore, the DT can process collected data to recognize occurrences of interaction states, such as HTR handovers, which are used by *High-Level Task Planner* to reason accordingly.

5.1.2 Implementation, Frameworks, and Equipment

The *MR-HRC-V2* architecture, presented in the last section, has been developed in continuity with its previous counterpart by embracing the open-source paradigm and making tailored contributions to relevant projects as needed. The ROS framework has been used as implementation platform and, for validation purposes, integrated with TIAGo++ [121], a well-known mobile manipulator robot from Pal Robotics. Here, implementation details for each module in the architecture are provided, highlighting significant contributions to open-source projects.

Robot operations are controlled with already available software modules. The robot's arms motion planning is performed via *Moveit*, whereas autonomous localization and navigation are handled by the *ROS Navigation Stack*. TIAGo's original perception capabilities have been extended to detect and localize objects in the environment. On top of the robot's head, a ZED2 camera has been mounted, calibrated with respect to the robot's reference frame and providing a wide-angle field of view (FOV). Whenever a new frame from the ZED2's left camera is captured, it is processed with YOLOv5 to recognize and localize objects. The output is an array of 2D bounding boxes $O = \{(x_0, y_0, l_0, w_0), (x_1, y_1, l_1, w_1) \dots, (x_{|O|}, y_{|O|}, l_{|O|}, w_{|O|})\}$ where x and y are the coordinates of the bounding box's upper left corner, and l and w are the bounding box's length and width. The 3D pose of an object is estimated by projecting its 2D bounding box on the corresponding depth map, see Fig. 5.3d. The result is an array of 3D bounding boxes $B = \{(x_0, y_0, z_0), (x_1, y_2, z_3), \dots, (x_{|B|}, y_{|B|}, z_{|B|})\}$. The *Perception Module* runs on an NVIDIA JETSON TX2, with 256 CUDA cores, mounted on the robot. At this stage, the *High-Level Task Planner* holds a predefined sequential set of actions for the robot



(a) (b)

(c) (d)

Figure 5.3 Fig. 5.3a and 5.3b show the holographic communication from the human perspective, respectively, for robot navigation and handover. Fig. 5.3c and 5.3d present the results of the perception module, with the bottles detected by YOLOv5m and the corresponding point cloud acquired by the ZED2 camera mounted on the robot.

to perform, i.e., *pick*, *transport*, and *place* actions. At run time, the predefined actions are grounded with values from the perception system, i.e., the object positions. When a task is unfeasible, the robot waits for human assistance. For example, when an object is out of the robot's reachable workspace, the *High-Level Task Planner* activates the *handover mode*: Tiago reaches a predefined handover pose (see Fig. 5.1) and waits for the human to hand over the desired object.

The *Mixed Reality Application*, developed using Unity and deployed to Hololens 2, runs at a 30Hz refresh rate with the native Hololens resolution. On top of the 3D engine, two SDKs were employed, namely *Vuforia*, which is used to extract the robot's position in the MR device's reference frame using a 25h9 April tag attached to the robot's base, and the already mentioned MRTK, responsible for overlaying 3D holograms on top of the real world. In this implementation of the MR-based, RTH communication, the delays Δt between holographic cues and subsequent robot actions have been set to 5 seconds, a value empirically determined to allow a reasonable separation between the two without significantly affecting task pace.

In continuation with the old *MR-HRC-V1* architecture, the already mentioned ROS-Unity interface has been employed for integration, thus providing proper communication with the rest of the modules. Additionally, the MR application's source code is openly available to other researchers through GitHub[1].

The *Digital Twin* module is developed and populated with objects offline using Unreal Engine 4 (UE4), a cutting-edge 3D engine with state-of-the-art visual and physics capabilities. UE4 and ROS have been interfaced through a URoboSim[2] plugin version we contributed to. Through this interface, the virtual instance received all the information necessary to represent the state of the collaboration in real-time, see Fig. 5.1. The DT run on a standalone machine equipped with an NVIDIA RTX A4000, effectively mitigating possible system bottlenecks. The variables tracked and continuously updated by the DT included the positions of objects, human operator, and robot, both of which represented with their corresponding real-world posture. In particular, the human's and robot's positions in 2D were tracked using a motion capture (MoCap) system composed of eight OptiTrack Flex 13 cameras, driven by a separate PC at 100Hz. The reflective markers for the tracking were positioned on the robot's top side and on the MR visor for maximum visibility. Instead, the robot's posture was replicated by directly reading its joint state at a frequency of 50Hz, while the human's one was perceived using a second ZED2 stereo camera. The ZED2 skeleton tracking functionality has been adopted, and the resulting skeleton was broadcast to UE4 using the Live Link plugin[3] at 15 Hz. This secondary ZED2 has been positioned in the top right corner of the map depicted in Fig. 5.4, and a preliminary calibration had to be performed prior to each experimental validation by having the individual assume a T-pose in front of it.

The fusion of high-quality motion capture technology alongside the ZED2 cameras provided a precise perception of the environment, sufficient, in our scenario, for the DT to drive the HRC without having to compensate for uncertainties. All the sub-systems, including the MR application, DT, and motion capture, communicated synchronously over TCP sockets with an average latency of 50 to 60 ms through a 50 Mbits/s access point. DT estimated the human gaze focus by casting a ray from human eyes with the head's inclination and determining its intersections with objects. This information, combined with the human's posture analysis, has been used to recognizing the individual's intent to perform an handover with the robot teammate, thus providing a reliable solution to employ implicit cues for HTR communication. The system waited for four concurrent conditions to determine the handover intention: (i) the robot should be in the *handover mode*, (ii) the human hand should be close

---

[1] GitHub: github.com/TheEngineRoom-UniGe/MR-Tiago

[2] URoboSim plugin: github.com/TheEngineRoom-UniGe/URoboSim

[3] Live Link plugin: docs.unrealengine.com/5.1/en-US/live-link-in-unreal-engine/

to the robot gripper, (iii) the gaze focus should be on the gripper, and (iv) the elbow angle should be close to ninety degrees. Once recognized, the handover intention was notified to the *High-Level Task Planner*, which planned the robot's gripper closure, exiting the *handover mode*. Again, the DT's source code is available on GitHub[4].

5.1.3 Experimental Setup

The experiments carried out and described in this Section had two objectives, which are hereafter denoted as *G3* and *G4* to avoid confusion with those of the previous Chapter. In particular, the present study was aimed at:

- *G3* Assessing the possible contributions of the DT to achieve intuitive, seamless HTR communication, employing cues derived from implicit human behavior;
- *G4* Evaluate the efficiency of MR-based, RTH communication in light of the more complex experimental scenario.

Taking such objectives into account, and in continuity with experimental validations carried out in previous chapters, the following hypotheses have been considered for the present investigation, namely that the comprehensive communication framework, combining MR and DT, could:

- *H1* Decrease the overall time to complete the collaborative task.
- *H2* Improve the efficiency of the human-robot team.

Collaborative Scenario

To test *MR-HRC-V2* and evaluate the hypotheses, a logistics-like, warehouse setting has been designed, where the human operator shares their working environment with the Tiago robot. The workspace consists in a room with three shelves and two crates, arranged to force human and robot to cross paths unintentionally throughout the experiment, see Fig. 5.4. The human teammate is supposed to restock *Shelf 3* with bottles taken from *Crate A* while the robot prepares an order by picking objects from *Shelf 1* and *Shelf 2* and placing them into *Crate B*.

For the human's task, twelve bottles with random numeric labels are placed inside *Crate A*, shuffled before the beginning of each experiment. The human operator should pick one item at a time from the crate and place it on *Shelf 3*, rearranging the bottles by labels in

---

[4] GitHub: github.com/TheEngineRoom-UniGe/UE-DTForHRC

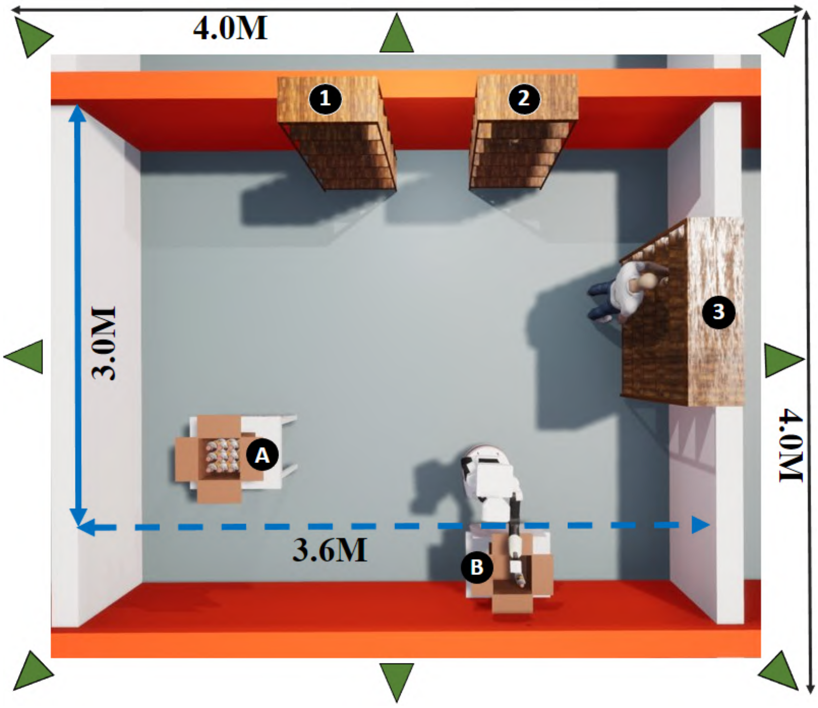


Figure 5.4 Bird-eye view of the collaborative workspace as it appears inside the DT. Various points of interest are labeled, along with the motion capture camera distribution (triangles) and workspace dimensions.

descending order. Conversely, four bottles are distributed between *Shelf 1* and *2*, with one bottle intentionally placed out of the robot's reach. Bottles' positions on the shelves are not predetermined, and the robot casually chooses which bottle to reach and pick next, providing relaxation to the second constraint mentioned at the beginning of the Chapter, thus removing the fixed sequence of robot actions. It is to be noted that robot actions may fail, that is, if an object slips from the gripper, or may be unfeasible, that is, when the bottle is outside the robot's workspace. In these cases, the human should understand the occurrence of the critical situation and supervise the behavior of the robotic teammate, possibly assisting it by correcting its grasp or collecting the unreachable bottle and handing it over. The human and the robot proceed in parallel until the circumstances force them to interact. However, given the compactness of the workspace, they frequently find themselves occluding their respective paths, potentially hindering each other's tasks. It is also worth noting that, given the presence of one bottle outside the robot's workspace, the human is forced to perform a handover with the teammate once during the experiment. In these cases, the robot is instructed to reach a predefined handover pose (see Fig. 5.1), waiting for the human to bring the object. The experiment is complete when the human has arranged all bottles on *Shelf 3*, and the robot has filled *Crate B* with the other four bottles.

User Study

An experimental campaign has been conducted, with $K = 20$ volunteers (15 males and 5 females) all aged between 21-33 ($Avg = 23.7$, $StdDev = 2.49$). All volunteers had little or no prior experience in terms of MR-HMDs and interaction with robotic platforms. The experiment has been approved by the Ethical Committee for research at the University of Genoa through protocol n. 2021/65, issued on November 18, 2021.

Unlike previous methodology, where participants performed the experiment twice, in two separate conditions, here they were randomly assigned to either one or the other. As such, subjects were asked to complete the joint activity with TIAGo++:

C1 Without holographic RTH communication, that is without *mr* channel;

C2 With *mr* channel active, thus experiencing the whole communicative act C^*, declinated in the DHT paradigm.

In both conditions, subjects worn the HMD since it hosted the MoCap markers, but in *C1*, the HMD was turned off. Before the experiment, candidates were instructed on their tasks and how to interact with the robot. For participants performing under *C2*, a very brief overview of how to navigate the holographic menus of the HMD was provided, along with instructions for the initial calibration of the MR application. After that, the experiment could start.

Given the preliminary nature of the study and the complexity of the experimental scenario, we chose to focus our analysis exclusively on the evaluation of *H1* and *H2*, thus only considering objective metrics associated with team efficiency. The two current hypotheses have been tested independently, using quantities measured during the experimental campaign. In particular, for *H1*, each trial was timed, measuring how long it took for the human operator and the robot to complete their respective tasks. As usual, video recordings of the experiments were made, and later analyzed in the post-campaign phase to extract as many relevant metrics as possible to assess hypothesis *H2*. Specifically, multiple quantities have been measured, in accordance with the principles expressed in Section 2.4:

- the number of times a participant proactively assisted the robot, leading to the robot's action unfolding correctly, for example, correcting a bottle position while being picked by the robot;

- the number of failed interactions between agents – the definition of this parameter is broad and includes both instances that hinder the interaction, for example, human and robot crossing their paths, and events that lead to the robot failing the bottle delivery, for example, non-intervention of human to adjust grasping position;
- the number of times a participant interrupted their task to investigate the robot's intention and next move.

All of these metrics, while being consistent with those measured in previous studies, are related to team coordination and mutual awareness, aspects which are nonetheless very important for fluent interaction. Overall, they have been employed to assess the effectiveness of RTH communication in such a complex experimental scenario, where a variety of critical situations could arise, depending on how human and robot interacted.

In contrast, the performances of the DT module have been assessed through two different metrics. In particular, we measured:

- the DT's accuracy in predicting human intention cues from implicit signals;
- the corresponding responsiveness, that is, the time needed by the digital model to recognize handover intentions.

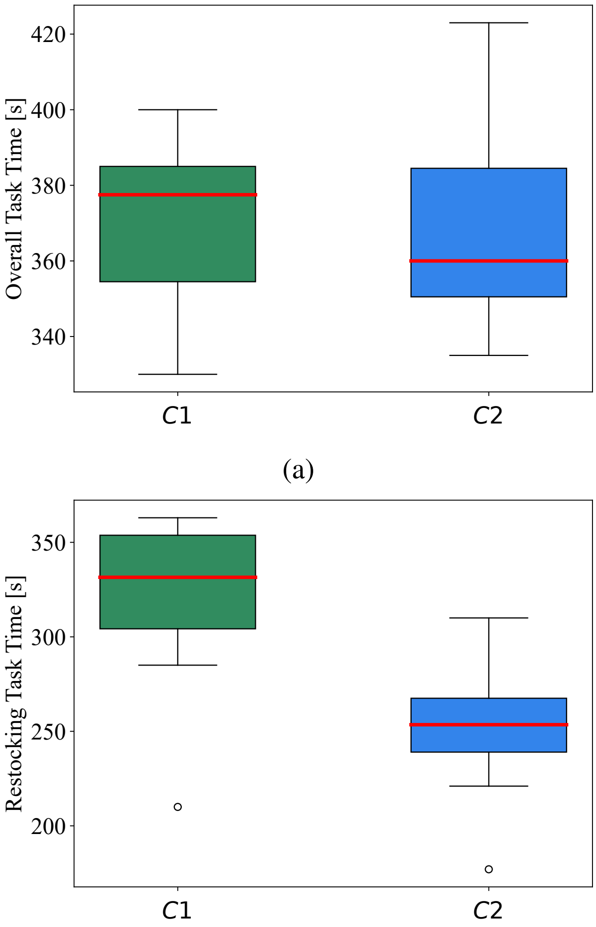
In this way, it could be possible to provide a preliminary estimate of the DT's capabilities when used as tool for intuitive, seamless HTR communication.

5.2 Results

5.2.1 MR for Mobile Collaboration

The results of the user study are hereby reported, with a focus on the two aforementioned hypotheses.

As for *H1*, Fig. 5.5a and 5.5b, respectively, report the results related to completion time for the overall collaboration and completion time for the human restocking task only. It is possible to note in Fig. 5.5a that the time required to complete the collaborative task remained comparable in both experimental conditions. However, Fig. 5.5b shows that participants under *C2* completed their restocking task, on average, in around 250 seconds, whereas the average measured time in condition *C1* was around 330 seconds. While these numbers may appear quite large for such a simple restocking task, it is to be noted that subjects were also simultaneously required to supervise the robot's actions and intervene

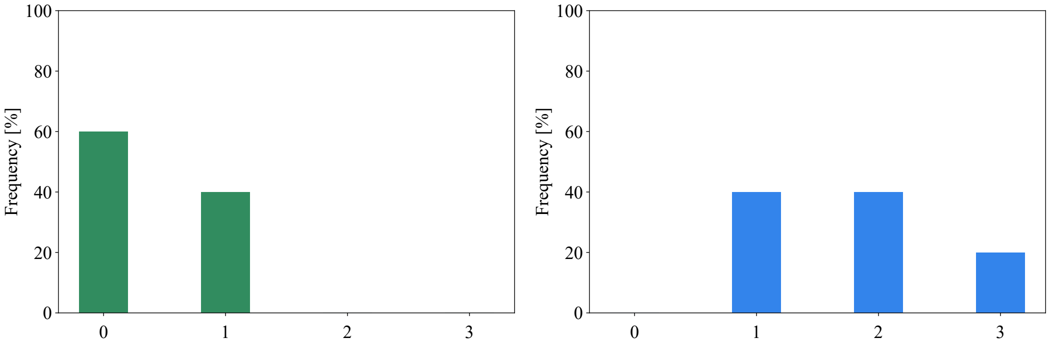


(a)

(b)

Figure 5.5 Time metrics (in seconds) observed during trials, in the two experimental conditions. Fig. 5.5a depicts the total time needed to complete the collaboration, measured once the human and the robot had both completed their tasks. Conversely, Fig. 5.5b depicts the time taken by participants to complete their restocking task, measured once the human had put all twelve bottles on *shelf 3* in the correct order. The small circles depicted in Fig. 5.5b highlight an outlier in the distribution.

when necessary. As such, an interpretation of the results conveyed by Fig. 5.5a and 5.5b is given as follows: participants always completed their task before the robot, and the total completion time depended only on the robot's performance, which was comparable in the two conditions. Nevertheless, participants in condition *C2*, aware of the robot's upcoming intentions thanks to the RTH holographic interface, managed to plan their movements and actions synchronously with those of their robot teammate, resulting in fewer mutual obstructions and faster completion times on the human's side. To further validate such results, we performed a T-test on the distributions depicted in Fig. 5.5a and 5.5b, which could be



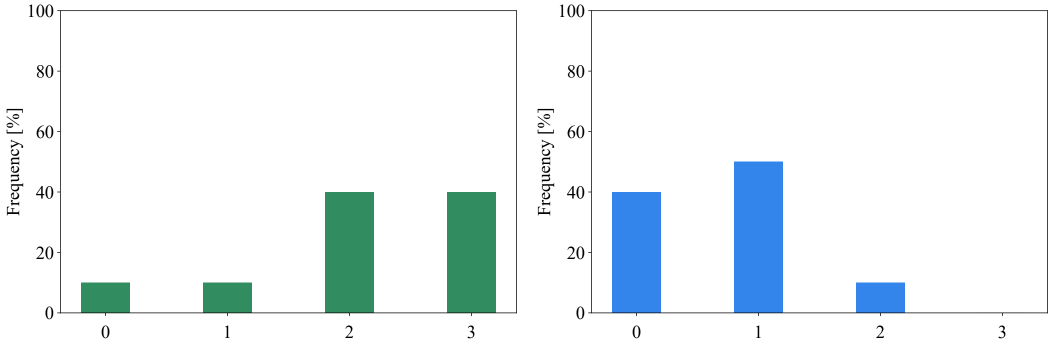
(a) Human assistive interventions under *C1*. (b) Human assistive interventions under *C2*.

Figure 5.6 Histograms depicting the number of human proactive interventions per trial.

assumed normal through the Shapiro-Wilk test (p-value > 0.05 for all distributions). For the data in Fig. 5.5a, the T-test returned p-value > 0.4, confirming that no significant difference could be observed in the total completion time in conditions *C1* or *C2*. Conversely, the T-test performed on the distributions in Fig. 5.5b yielded p-value < 0.01, thus corroborating the significant difference between times measured on completion of the restocking task under *C1* or *C2*.

On the other hand, the following discussion is related to the various metrics employed to evaluate *H2*. Fig. 5.6 reports the results related to the amount of human assistance offered to the robot. In particular, the histograms depict the percentage of participants who completed the collaboration by performing a number of proactive interventions to assist the robot. For example, in condition *C1*, 60% of subjects concluded their experiment without helping the robot. On the contrary, participants in condition *C2* were more proactive due to the RTH communication, which allowed them to understand the situation in advance and proceed to aid the robot in completing an action. Since data distribution in Fig. 5.6 could not be assumed normal, the Wilcoxon signed-rank test was adopted [172] for statistical evaluation. The test yielded a statistic $W = 36$, which have been compared with the critical value $W_c = 60$ extracted from [171], fixing the significance level $\alpha = 0.05$ and the sample size $K = 20$. Since $W < W_c$, the null hypothesis could be rejected, confirming the significant difference between the degree of human assistance in conditions *C1* and *C2*.

Similar considerations can be made by observing Fig. 5.7, which depicts data related to failed interactions during the experiments. Under condition *C1*, participants lacked an intuitive communication channel with the robot. Therefore, they were more easily caught off-guard by the robot's actions, hindering each other's task or failing to interact when needed. In condition *C2*, on the other hand, participants were more aware and responsive, increasing



(a) Failed interactions under *C1*. (b) Failed interactions under *C2*.

Figure 5.7 Histograms depicting the number of failed interactions per trial.

the likelihood of detecting and intervening in potentially hazardous situations. Again, the Wilcoxon test was used to evaluate the significance of such results. In this case, the test yielded statistic $W = 4$, and by comparing this value with the critical one (W_c) mentioned before, we could confirm the statistical difference between the two experimental conditions.

Finally, Fig. 5.8 illustrates the number of times participants interrupted their restocking task to observe the robot. In condition *C1*, without a holographic communication channel, it was more challenging for subjects to infer the robot's intention. Therefore, they often needed to stop their current task to observe the robot's actions before they could pick their next move. Participants under *C2* could see the robot's following actions intuitively and in advance, allowing them to plan their moves without disrupting their pace, resulting in fewer mutual obstructions and less time required to complete the restocking task. We assessed the significance of these results through a T-test, carried out after ensuring that distributions could be assumed normal (Shapiro-Wilk test yielded p-values > 0.2 for both cases). The T-test returned a p-value < 0.01, corroborating the statistical difference between *C1* and *C2*.

5.2.2 DT's Performances

Throughout the 20 experiments, the capabilities and responsiveness of the DT model developed during this work were assessed. In particular, we focused on measuring the inference time, that is the interval required by DT to properly recognize handover intentions. Such time has been measured from video recordings of the experiments, starting from the instant in which the human reached the handover pose, to the subsequent moment when the robot proceeded to close the gripper. DT was always successful in recognizing handovers during the 20 trials, suggesting that it can estimate collaborative intentions online from the combination of agents' postures and gazes. The time taken by the system to recognize the handover

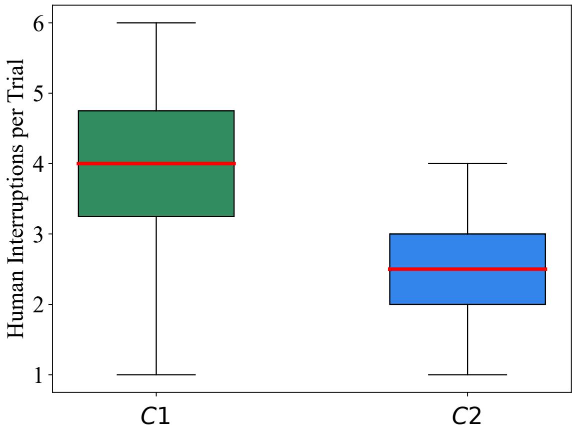


Figure 5.8 Number of times per trial in which participants had to interrupt their task to observe the robot, infer its upcoming intentions, and subsequently plan their following action.

was short (*Avg* inference time $t_{inf} = 2.3\,seconds$, $StdDev = 0.33\,seconds$), and in only three cases the process required more than 5 seconds.

5.3 Discussion

This Chapter presented a first attempt at integrating MR and DT into a comprehensive framework covering both directions of communication in HRC. On the one hand, the findings observed in terms of RTH communication are consistent with previous chapters, thus corroborating the effectiveness of MR for intuitively conveying robot's intention cues using the DHT approach. Specifically, the ability to represent not only manipulation actions through dynamic, animated overlays, but navigation intentions as well, resulted in more fluent and seamless collaborations, where the human operator was more aware of the teammate's movements, and more proactive in responding to robot's actions, intervening to aid when necessary. The increased rate of successful interactions observed in condition *C2*, as well as the enhanced degree of human assistance offered to the robot, proved how crucial the communication aspect is for a more efficient collaboration. In addition, unlike previous investigations, the current findings highlighted the positive effect of holographic communication in speeding up a portion of the overall task, namely the human restocking activity. To this regard, it is worth noting how the ability to preview imminent robot's movements in condition *C2* enabled individuals to plan their actions ahead, carefully choosing their steps to avoid obstructing the robot's pace, as opposed to participants in condition *C1*, who appeared more hesitant in taking action due to the lack of feedback. In light of these findings, it was possible to conclude that the goal *G4* had been achieved, providing multiple evidence of how MR can represent an effective RTH communication strategy in various domains of HRC.

On the other hand, the adoption of the DT for the HTR layer provided an added value to the *MR-HRC-V2* architecture. Although such digital model has been employed to recognize human's non-verbal cues in a simple handover scenario, promising results in terms of recognition accuracy and responsiveness have been observed, highlighting how DT can be employed to estimate HTR implicit communicative acts online. Thus, the objective *G3* could be considered as accomplished. Nonetheless, further research efforts could be undertaken to study the effectiveness of DT when employed to extrapolate intention cues under more complex interactions between human and robot.

With these results in mind, the next, natural step of the research involved the creation of a generalized, modular architecture, which could encompass the features of *MR-HRC-V1* and *MR-HRC-V2*, without depending on a particular HRC context, so that it could be easily utilized by other researchers and companies to take advantage of the proposed DHT communication scheme. Nevertheless, at this point during my Ph.D. journey I realized how the software architectures developed so far were platform-centric and constrained to the particular HRC scenario they had been created for, thus it would have been impossible to build a modular, robot-independent framework from such implementations. As such, a great effort has been made to re-write the code from scratch, using a different game engine for improved holographic realism, in an attempt to build the envisioned open-source and reusable framework for bi-directional, holographic communication. This new and improved version of the architecture is the topic of the next Chapter.

Chapter 6

A Modular, Open-Source Architecture for Holographic Human-Robot Communication

Elaboration and integration of an article published in:
Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023) [101]

The experimental results discussed so far have shown the positive impact of holographic, RTH communication applied in HRC settings. As a matter of fact, given the little hardware and limited calibration procedures required, such a MR-based framework could be easily adapted for a multitude of operating scenarios, either research-oriented or manufacturing-centered, providing a useful layer for human-robot communication. Nevertheless, the value offered by this holographic framework is significant only if other practitioners can re-use the developed software components with minimum effort and maximum portability. Unfortunately, the implementations developed up to this point of the Ph.D., namely the two versions of the *MR-HRC* architecture, while adhering to the open-source paradigm, could be re-used by others only after a non-negligible setup phase, comprising various configuration steps, to make sure the code could be compiled, executed and successfully deployed on the HMD. At the same time, the implementations were platform-centric, that is they were intended to work only for the specified robot, in a particular HRC scenario. In other words, switching Baxter's or Tiago's holographic representations, by incorporating another URDF model in the Unity scene, was not straightforward and multiple modifications were required in the code to ensure a fully operational holographic experience.

In light of these issues, I decided to re-write the implementation from scratch, in an attempt to build the envisioned open-source, modular architecture, which could be re-used by other practitioners to taken advantage of the DHT communication scheme, with minimum effort and in generalized contexts of HRC, thus with no constraint on the particular robot platform. Such complete rework involved major overhauls in the software, including the adoption of a different game engine for more realistic and immersive holographic experiences, and the employment of a cloud-based data exchange platform to overcome some inherent limitations of ROS, thus opening up the possibility to use the proposed MR-based communication scheme even in real-world scenarios where ROS may not be available.

Given these premises, the present Chapter provides a thorough description of the resulting architecture, denoted *RICO-MR*, as already mentioned in Chapter 1. First, an in-depth description of the various components is given, then a simple use case example is provided to showcase how other practitioners can use the application for any given collaborative context. Finally, the Chapter provides a brief digression on potential, additional applications of the architecture, thanks to the new and improved physics engine. In particular, the possibility to use *RICO-MR* as full-fledged, holographic DT is discussed and some examples showcased.

6.1 Methods

6.1.1 Software Architecture

The proposed *RICO-MR* software architecture is illustrated in Fig. 6.1. It comprises two macro components: the *Mixed Reality Application* (on the right), built with Unreal Engine 4.27 (UE4), and the *System*'s architecture (on the left). The communication between the two components operates through *Apache Kafka*, an open-source, high-performance distributed streaming platform.

Mixed Reality Application

As already stated multiple times throughout this thesis, *Mixed Reality Application* plays a central role in the architecture and is deployed on the HMD device worn by a human teammate. It is responsible for rendering the holographic layer used by the robot to convey its imminent intentions, according to the DHT paradigm. Within its UE4 implementation, the holographic layer is built using Microsoft's *Mixed Reality UX Tools* plugin, a popular framework providing building blocks and functionalities to develop 3D virtual experiences and targeting a specific family of HMDs, namely Microsoft HoloLens, and HoloLens2. The

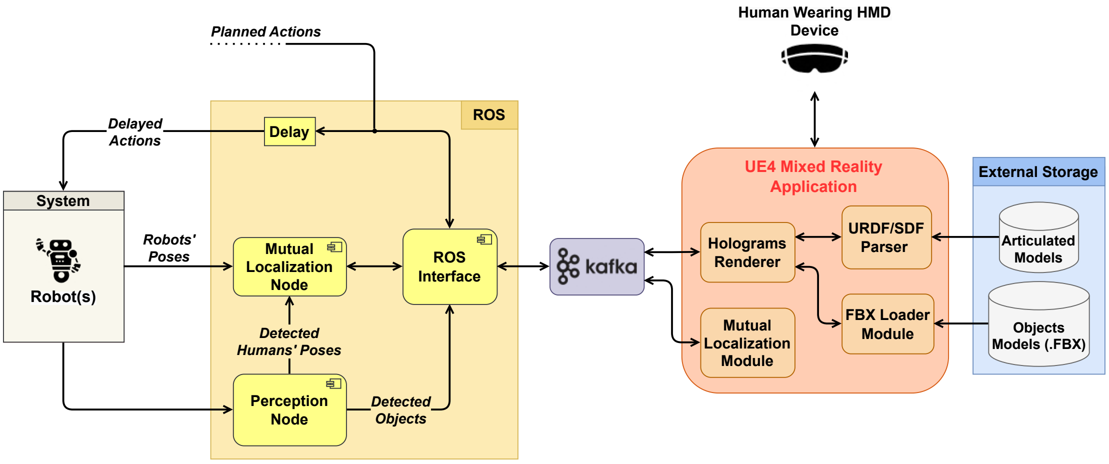


Figure 6.1 Overview of the *RICO-MR* architecture.

primary distinction from previous iterations of the architecture lies in the UE4 application's capability to dynamically parse and load 3D models at runtime. This includes both simple and articulated models, allowing for the generation of holograms representing robots, items, and relevant tools within the MR layer. In contrast to earlier implementations, where the holographic model of the robot was embedded into the Unity scene, which, in turn, was compiled, packaged, and deployed onto the HMD device, here the use of UE4 introduces greater flexibility. With UE4, the robot's assets can be loaded at runtime, enabling swift changes and loading of robotic holograms without incurring additional compilation times. Nonetheless, the communicative capabilities offered by the MR layer remain consistent, as the possibility to load holographic robots, as well as items and other assets relevant to the HRC scenario, guarantees that both robot's states and beliefs can be conveyed. More in detail, a simple side menu within the application allows users to select which robot model to load inside the MR scene. As better highlighted in Section 6.2.1, *Mixed Reality Application* ships with several pre-loaded models. However, the list can be extended by specifying an appropriate remote storage repository in the application settings. Such a repository can be employed to store relevant robot's resources (i.e., URDF and SDF files) and is used at runtime to refresh the list of models ready to be spawned as holographic assets.

Aside from the runtime generation of holographic content, the new implementation of *Mixed Reality Application* provides tools for mutual localization between the HMD and the rest of the system. Specifically, a built-in functionality of Microsoft's MR plugin has been used, that is its QR code tracking capability. To this extent, a simple QR code is employed to establish the initial mutual localization between the HMD and the robot in the real world, ensuring that the holographic model is spawned consistently with its counterpart. Upon completing this initial phase, which is unchanged with respect to previous iterations of the architecture, the HMD continuously publishes its updated pose (i.e., position and orientation in space), thus ensuring consistency between the mutual localization of the human and robot, even if the user moves around the environment. Although the employed plugin can only track one QR code at a time, the application can easily accommodate multi-robot scenarios. The underlying implementation stores the data payload of each tracked marker. As such, it is possible to spawn several robot models inside the MR layer by simply generating as many QR codes with different textual content. However, users need to keep in mind that the mutual localization phase is carried out only once, with the first QR code tracked, which effectively plays the role of spatial anchor between HMD and surrounding environment. Therefore, regardless of the number of markers employed in the particular scenario, the HMD's pose is always updated and published with respect to such anchor. Nevertheless, it is always possible to query the system to find out relative localization between the first QR code and subsequent ones, ensuring that consistent spatial relationship between HMD and markers can be computed at any time.

Finally, coherently with how custom robot's resources are handled, *Mixed Reality Application* also ships with a link to a repository containing FBX files of objects of common use which can be loaded as holograms on *System*'s request. The link can be modified in the application settings, enabling users to customize the holographic layer according to their particular HRC scenario. This feature also ensures that users can effectively adopt the application off-the-shelf, with no need to compile the project and manually load their FBX files inside the UE editor. Nevertheless, to provide support to researchers and companies interested in expanding the capabilities of the architecture or adding modules to it, we decided to make the UE4 project publicly available under MIT license[1].

System's Architecture

The *System* is purposely denoted by general terms to indicate that *RICO-MR* is independent of the adopted robotic platform. Its intrinsic modularity, combined with the QR code tracking capabilities, allows scaling up to accommodate multiple robots simultaneously, thus broadening the range of HRC scenarios in which the architecture can be employed. Moreover, the *System* can account for external sources of data (e.g., motion capture systems and external depth cameras), which can be integrated into ad hoc applications. In continuity with previous implementations, the *System*'s architecture is developed using ROS, as it allows straightforward integration with ROS-based robotic platforms, such as Baxter or Tiago.

---

[1] GitHub: https://github.com/TheEngineRoom-UniGe/RICO-MR

However, as later discussed when detailing the role of *Kafka* in the architecture, the ROS adoption is optional and can be partially or fully replaced.

In general, the *System* provides *perception* and *localization* capabilities. Both can originate directly from the robot(s) or external sources. Through perception, the robot(s) perceive the tools and objects inside the collaborative space, recognize them through appropriate object detection models, and track their poses. This information flows through *Kafka* to *Mixed Reality Application*, which loads the corresponding resources (i.e., FBX files) and spawns the objects' holograms consistently with the real world. Once objects are detected by the perception and their holograms spawned in the MR layer, such holographic counterparts can be used by the robot to project its intended beliefs, alongside its upcoming states, whenever a new MR-based communication act is issued. Again, this aspect greatly enhances the level of re-usability of the architecture, as the holographic scene no longer needs to be populated off-line with virtual objects, but rather all assets can be loaded at runtime, in response to *System*'s perception. To this regard, it is worth mentioning here how the adoption of UE4 plays a major role in boosting the level of realism offered by RTH holographic communicative acts. The high-quality physics simulation ensures realistic collisions and forces between virtual objects, thus making it possible for holographic robots to manipulate and interact with holographic items as if they were real, further improving the effectiveness of the RTH communication aspect. This newfound realism is particularly relevant if the robot has to issue communication acts aimed at projecting manipulations of asymmetric or skewed objects.

Moving back to the *System*, perception can also account for detecting and tracking the human teammate's pose, which is fed to the localization node. This component maintains a coherent estimate of the mutual localization between agents by merging the perception data with the localization data from the HMD. This mechanism ensures that the holographic representations are consistent even in scenarios of mobile collaboration, where both human and robotic agents move throughout the shared workspace.

Conveying Robot's Intentions

Regarding the robot's actions to be displayed as holographic intentions, the proposed architecture is modular, enabling developers to integrate their custom action planners. As such, it is possible to deal with manipulation and/or navigation actions, adapting the architecture to any collaborative scenario. As an example, for ROS-based solutions, the architecture can easily be interfaced respectively with *MoveIt* for motion planning and manipulations and with *Navigation Stack* for path planning purposes.

More in detail, a list of *Kafka* topics exposed by the architecture is presented in Table 6.1. For each robot model tracked within the holographic layer, two topics are available, respectively the */robot/{id}/navigation_plan* for previewing navigation actions and */robot/{id}/joint_trajectory* for manipulation actions. The *id* parameter is automatically assigned by the architecture once a new robot model is identified through the QR code and its holographic representation spawned in the MR scene. This mechanism allows the architecture to manage multiple robots at once. Overall, the flow of information can be described as follows. Planned robot actions published on the respective topics are dispatched through *Kafka* and received by *Mixed Reality Application*, which proceeds to render them as holographic animations, according to the DHT paradigm. At the same time, the small, fully parameterizable *delay* Δt is introduced to the real action executed on the robot to ensure that the human teammate can experience the holographic intentions first.

Table 6.1 List of *Kafka* topics available to interact with the architecture. We report, for each topic, the corresponding ROS message type to ensure support for ROS-based solutions. In particular, conversion from the ROS message to its equivalent representation in *Kafka* could be achieved by filling the *Kafka* message's payload with the JSON-serialized content of the ROS message.

Published topic	
Topic name	**ROS message type**
/hmd/{id}/pose	*geometry_msgs/PoseStamped*
Subscribed topics	
Topic name	**ROS message type**
/robot/{id}/navigation_plan	*nav_msgs/Path*
/robot/{id}/joint_trajectory	*trajectory_msgs/JointTrajectory*
/object/{id}/state	*rico_msgs/State*

Table 6.1 also shows a */object/{id}/state* topic, which is exposed by the architecture and purposely intended for perception components being used within the particular HRC scenario. This topic makes it possible to integrate external perception pipelines, enabling robot(s) to detect and track objects in the collaborative space and spawn the corresponding holographic representations accordingly. To this extent, the *id* parameter can be configured so that each tracked object has its own *Kafka* topic. Additionally, the topic's type is custom-defined, enabling developers to specify the object category (as a string field), its pose, and its joint state in the case of articulated objects. Such information is received by *Mixed Reality Application*, which proceeds to load the resources and spawn the holographic model

consistently. The ROS package, including definitions of custom message types, is provided in the accompanying repository.

Finally, as mentioned in the previous paragraphs, the architecture publishes the HMD's pose tracking in a topic named */hmd/{id}/pose*. This information is made available to the external world for third-party applications needing localization information to develop multiuser or shared MR experiences. As before, the *id* parameter is automatically assigned to each HMD connecting to the architecture, thus ensuring that the system can scale up to accommodate multiple users, broadening the possible applications in multi-human-robot interaction scenarios.

Kafka Component

Apache Kafka is an open-source, high-performant, and distributed platform for data streaming based on the publish-subscribe paradigm. In the context of this work, it enables communication between the various components of the *RICO-MR* architecture, while ensuring integration with other solutions, regardless of their usage of ROS.

The adoption of *Kafka* aims to overcome some inherent limitations of the publish-subscribe system for ROS and ROS2 architectures. On the one hand, ROS networking is based on the concept of a single master, with multiple nodes acting as producers and consumers. This configuration, however, has intrinsic scalability issues, as the master node can become the system's bottleneck in case of large throughput of data. In addition, such a scheme does not allow real-time constraints to be met since swarms of robots or highly complex robotic platforms may render the system unstable. On the other hand, ROS2 has been developed to solve the single master issue by adopting Data Distribution Service (DDS) middleware. Although this solution scales horizontally, providing a proper architecture for multiple robots, it still suffers from the limitation of a push-based architecture; a single server keeps track of all the subscribers and consumers and delivers the messages accordingly. For instance, in a robotics application where a robot is pushing information at a high frequency, this type of architecture is prone to bottlenecks because the server does not keep up with high messaging rates.

Conversely, by adopting *Kafka* the server is not responsible for sending messages to all consumers. Instead, it maintains an offset record, which consumers can request. Such a push/pull system provides horizontal scalability with a message replication feature, i.e., in the case of a faulty server, the messages will not be lost, and the performance will not be affected. Nevertheless, to still guarantee legacy interaction with ROS-based solutions, a *ROS-Kafka interface* has been designed to extend the publish-subscribe capability of ROS with the distributed streaming service of *Kafka*. This additional interface makes sure that each ROS node is treated as a master that can publish and subscribe to any topic, even those topics streamed by other master nodes, by the moment *Kafka* handles message sharing between the nodes. This approach not only provides a robust horizontally scalable solution, but also opens the possibility of integration with different robotics frameworks other than ROS, either in cooperation with ROS or as a full replacement.

Finally, it is worth noting how the adoption of a cloud-native protocol such as *Kafka* facilitates the integration of *RICO-MR* with external DTs, to build comprehensive human-robot communication layers, similarly to what has been discussed in Chapter 5. This ease of integration follows from the abstract nature of *Kafka*'s communication protocol, which is unrelated to the programming language or hardware used, thus rendering the system extendable. For instance, a DT running on a full-fledged PC or cloud server can perform heavy computations or look-ahead simulations and communicate visual feedback about the ongoing collaboration directly to *Mixed Reality Application* running on the embedded HMD device.

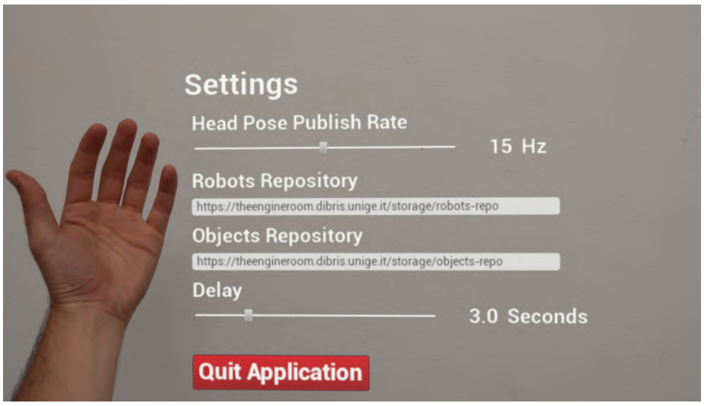
6.2 Results

6.2.1 Example of Usage

As already mentioned, *RICO-MR*'s holographic application can be employed off the shelf without further customization. It is sufficient to follow the steps in the GitHub repository to download its packaged version and install it on the HMD device, ideally a Microsoft HoloLens2, via a USB connection.

Application Settings

Upon launching the application, users can change settings by accessing the corresponding hand-attached menu, which is programmed to pop up whenever they look at their left hand. To avoid involuntary triggers, this action requires the user to gaze at their hand, which is laid flat and with inward palm. Fig. 6.2a depicts such settings menu as it appears to the user inside the holographic layer. As previously stated, the application ships with default links referencing public repositories made available with the application itself. Such repositories host, respectively, robots' and objects' resources. The former comprises assets (i.e., URDF and SDF files) associated with robot models commonly employed in research applications, including *Baxter*, *TIAGo++*, as well as *Panda* from Franka Emika and *UR5* from Universal



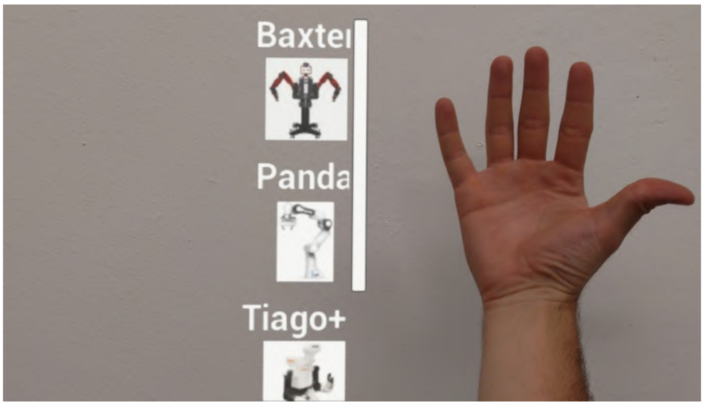
Head Pose Publish Rate 15 Hz

Robots Repository https://theengineroom.dibris.unige.it/storage/robots-repo

Objects Repository https://theengineroom.dibris.unige.it/storage/objects-repo

Delay 3.0 Seconds

(a) Settings menu.



(b) Model selection menu.

Figure 6.2 Overview of the holographic menus inside the MR layer, enabling users to customize the application's behavior.

Robots. On the other hand, the second repository stores models (i.e., FBX files) of simple and common items which can appear in HRC scenarios, including screwdrivers, hammers, and water bottles. Nevertheless, by simply modifying such links in the settings, users can point to their repositories, ensuring that other robot models and objects can be employed and loaded as holographic assets. It is important to note that the application can also deal with articulated objects, i.e., objects endowed with internal degrees of freedom. Such items, assuming they are properly described by a URDF file, can easily be spawned as holographic assets in the same way as robot models do.

Additionally, users can interact with the settings menu to customize the publication rate for the HMD's pose, which is triggered once mutual localization between the visor and QR code is established. Allowed rates range from a minimum frequency of 1 *Hz* up to a maximum of 30 samples per second.

Finally, users can adjust the delay Δt controlling the elapsed time between holographic and subsequent robot actions. This feature makes it possible to experiment and find the optimal temporal distance between holographic intentions and actions, ensuring smooth interaction between agents and thus maximizing team efficiency.

Spawning Robot Holograms

To properly spawn robots inside the holographic scene, users should select the corresponding models via the menu attached to their right hand, which works in the same fashion as the settings. This menu, depicted in Fig. 6.2b, enables users to choose robots from a list of available resources. In particular, Fig. 6.2b shows the list with the four models described above, which ship with the default robots' remote repository. Once the custom repository is configured in the settings, on every startup the application will connect to it, checking if new robot models have been uploaded. In such cases, the corresponding resources are downloaded, thus making them available in the model selection menu for the user. Upon selecting the desired model, the user can close the menu, turn to the QR code and scan it with the HMD's camera, causing the robot hologram to spawn at the marker's estimated location. Subsequently, the user can manually adjust the holographic model's position by simply interacting with it through hand gestures if the robot's virtual replica and its physical counterpart are misaligned. From that moment on, the architecture exposes the topics associated with the newly spawned robot model, enabling the MR application to receive messages and issue the corresponding RTH holographic communication acts.

6.2.2 *RICO-MR* as Holographic DT

As previously highlighted, the adoption of UE4 as game engine has brought forth a key advantage, that is the ability to leverage its superior physics for constructing more realistic holographic experiences. In the earlier Unity-based implementation, collisions between virtual entities were imprecise, as the physical simulations occurred at a kinematic level. Consequently, interactions between holographic objects, such as the robot manipulating an item to anticipate a pick-and-place action, achieved the intended effect but appeared rigid and unrealistic. On the contrary, the current solution relies on the dynamic simulation of collisions and forces, resulting in more realistic RTH communicative acts, especially when the robot needs to convey complex bi-manual manipulation actions. Simultaneously, the enhanced physics enables the utilization of a fundamental feature of MR-HMD devices, namely their ability to map the surrounding environment and create a digital representation of it. With a

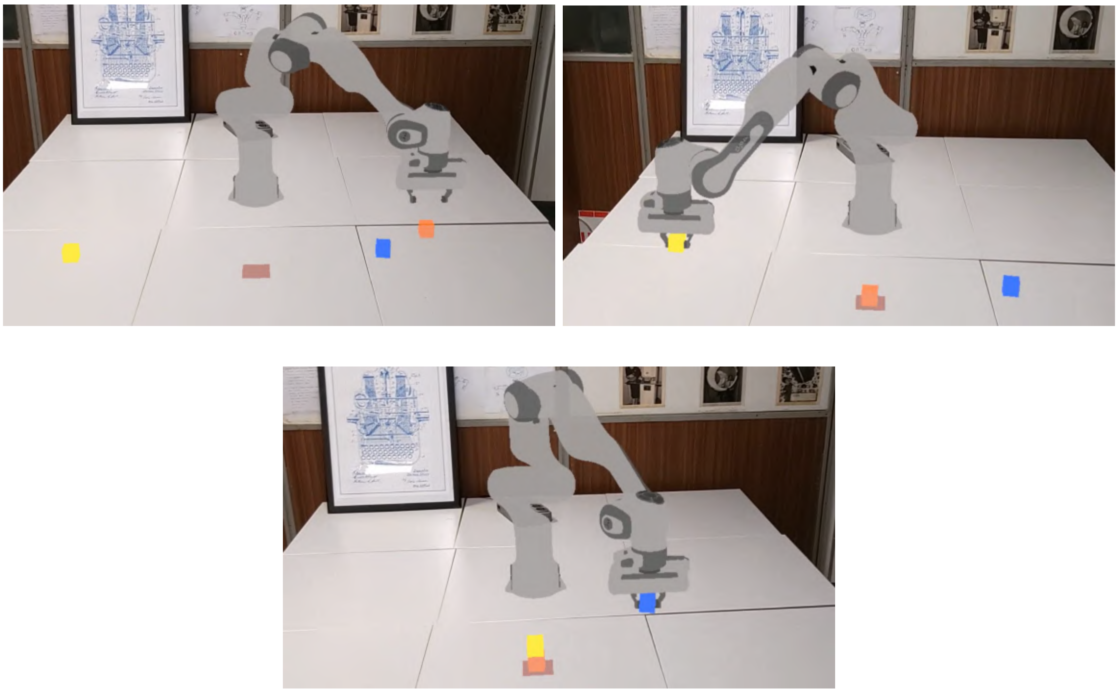


Figure 6.3 Panda robot collecting and stacking cubes in MR. It is noteworthy how all virtual entities are subject to gravity and physically respond to the interaction with the table.

virtual map of the workspace, it becomes possible to simulate interactions between real and virtual objects, constructing immersive hybrid experiences. In such scenarios, holographic items fall realistically, subject to gravity, and adhere consistently to tables and surfaces as if they were tangible entities.

Given these premises, it is worth noting how the holographic application devised in the context of *RICO-MR*'s architecture could potentially play a larger role than mere MR-based communication layer. As a matter of fact, the digital layer could effectively be employed as DT in training scenarios or in all those circumstances where interaction with a real artificial agent is not possible. To this regard, some example applications are depicted in Fig. 6.3 and Fig. 6.4. In particular, Fig. 6.3 showcases a scenario where a virtual Panda robot performs a sequence of pick-and-place actions to collect and stack a series of holographic cubes, which have been randomly displaced by the user. Although the virtual robot is spawned using a combination of model selection and QR code, as described in the previous Section, all holographic entities are physics-enabled and interact with the surrounding workspace consistently. Thus, the operator can manipulate and release the cubes in the air, just to watch them fall on the table realistically. Then, through a simple vocal command, the user can issue the *collection* routine and have Panda pick and stack the various cubes, provided they

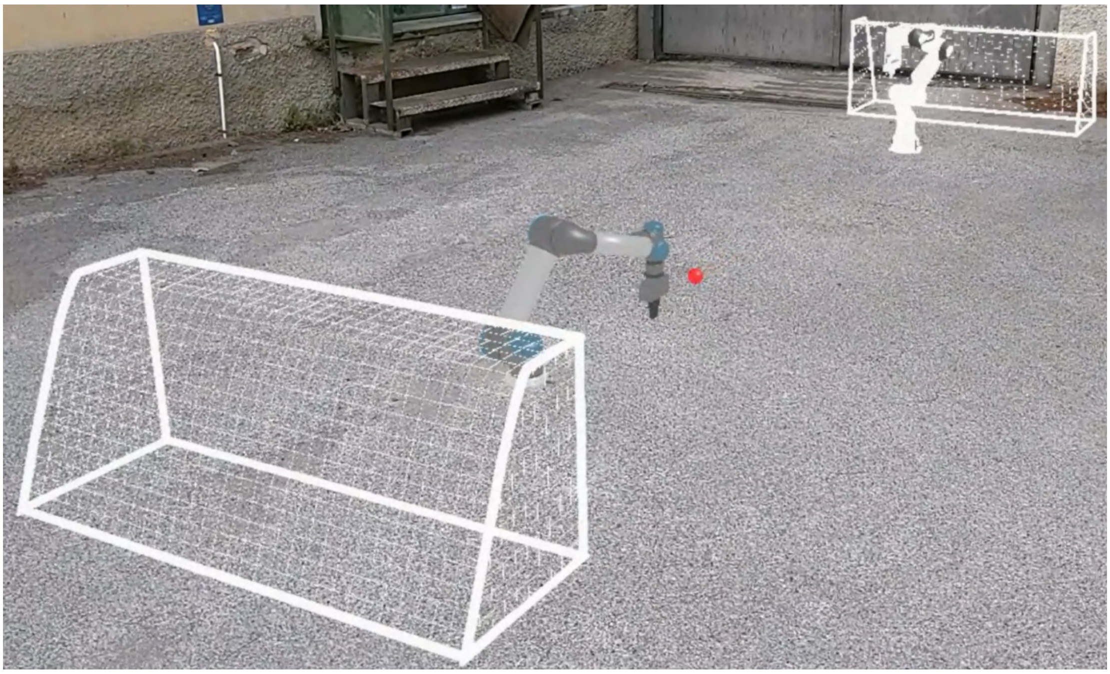


Figure 6.4 Panda and UR5 playing soccer in MR. Here, the ball is physically simulated and responds to collisions with the robots' grippers who move to catch it. Based on the impact force and direction, the ball is pushed towards the opponent.

are within reachable limits. Similar solutions could find applications in HRC for various purposes. On the one hand, interacting with holographic robots could provide proper training and acclimatization for operators, before exposing them to actual interaction with the physical machine. At the same time, the holographic layer could serve as mean to design and preview collaborative cells or to study and investigate novel interactive and cooperative scenarios for human-robot teams, all without requiring the employment of actual robotic hardware.

In a similar fashion, Fig. 6.4 showcases a scenario where two holographic robots, namely Panda and UR5, play soccer in the MR layer. In particular, this example provides an overview of the capabilities of *RICO-MR*, which is able to handle multi-robot contexts such as this. In this scenario, the robots are controlled through ROS and programmed to move their end-effector towards the incoming ball, aiming to hit it and send it in the opposite direction while simultaneously protecting their goal. The physics-enabled simulation ensures that the ball responds realistically to collisions with the robots' grippers, bouncing back in the opposite direction with a speed and trajectory consistent with the impact. This capability allows, for example, the design of complex motion planning strategies, with the goal of hitting the ball at a specific speed and angle to create a trajectory that can outsmart the opponent.

Overall, the ability to transform the holographic communicative layer into a physical simulation adds value to the *RICO-MR* architecture. This enhancement could significantly broaden its range of applications in real-world scenarios, serving as a tool for training and education, as well as for research and development purposes, enabling the design and study of new collaborative environments.

6.3 Discussion

This Chapter has presented *RICO-MR*, the latest implementation developed throughout this Ph.D. work to fulfill the role of envisioned holographic, intuitive communication layer for HRC. *RICO-MR* has been designed to overcome the various limitations of previous installments, providing a generalized, re-usable framework that other practitioners can easily employ in their HRC studies on in manufacturing settings, to take advantage of the communication scheme formalized by the *MR-Space*. In particular, the code has been publicly released, making both binary and source code available. This ensures support for end-users of the holographic application, while facilitating researchers and developers interested in expanding and customizing the software components.

RICO-MR's strengths include the capability to load holographic assets at runtime, allowing for use in any scenario, regardless of the specific robot platform employed. Additionally, the adoption of a cloud-based data streaming infrastructure such as *Kafka* ensures that *RICO-MR* can scale up to accommodate multi-robot and multi-user experiences, all the while providing support and integration for both ROS-based and ROS-independent frameworks. The result is a modular and scalable architecture, which leverage UE's physics to ensure more realistic interaction between holographic entities, thus achieving superior RTH communicative capabilities, while also potentially playing a significant role as MR-based simulation tool.

With such an architecture at hand, the final step in this Ph.D. work consisted in investigating the holographic, HTR communication aspect, to achieve the envisioned bi-directional communicative interface. Unlike the solution discussed in Chapter 5, however, which relied on an external DT to capture human non-verbal cues, the following Chapter will explicitly address holographic-based HTR communication, achieved by extending the functionalities of the MR application without relying on supplementary equipment or infrastructure. In particular, the topic of KT will addressed, given how relevant it is in collaborative settings, and how such form of teaching can effectively be regarded as HTR communication aimed at conveying robot's tasks and actions. KT will be formalized within the *MR-Space* communication framework and a practical *RICO-MR*-based implementation will be detailed, enabling holographic KT on any robotic platform that is URDF compliant.

Chapter 7

Holographic Kinesthetic Teaching as Human-to-Robot Communication

Elaboration and integration of an article submitted to:
33rd IEEE International Conference on Robot and Human Interactive Communication
(RO-MAN 2024)

Up to now, the dissertation has mostly regarded one-directional holographic communication, namely from the robot to the human teammate. However, in light of the results discussed in the previous Chapter, that is with a modular and robust architecture at hand, it is now possible to tackle the aspect to HTR communication, aimed at conveying relevant pieces of information to the robot through intuitive holographic cues in the digital layer.

Following on the discussion of Section 2.2.2, the present Chapter deals with KT and treats such form of teaching process as a proper HTR communication act, in which human operators instruct robot teammates about upcoming tasks and actions they need to carry out. Although KT is crucial for most collaborative settings, as it allows easy transfer of skills to the robot, its performances heavily depend on a combination of control scheme used for its implementation and complexity of the robot's kinematic chain [124]. As a result, most practical KT applications offer less than optimal experiences, where human operators are required to manually move stiff robot links to achieve the intended effect. A promising solution to this drawback could come from using the MR layer as channel to achieve KT, subsequently transferring the learned skill to the real robot for execution. This could potentially make operators' life easier, while offering them a intuitive interaction alternative to achieve the same goal. To this regard, the *RICO-MR* architecture could serve as mean to this end, given that it is robot-independent and offers high-level physical interaction,

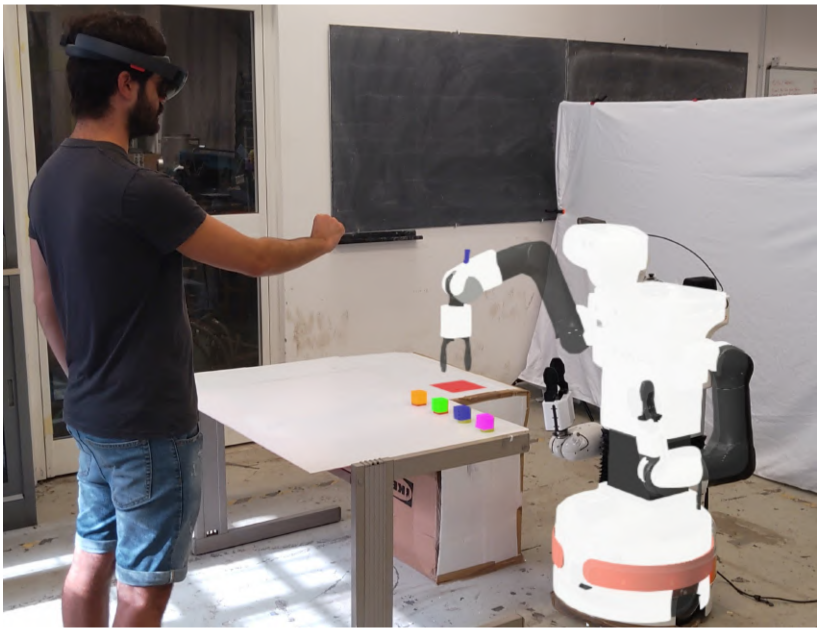


Figure 7.1 The user in the middle of a holographic KT session with the Tiago++ robot. Through a combination of gestural and vocal interaction, the operator can use the MR interface to teach the robot actions intuitively.

thus able to simulate the whole process of KT consistently. As such, the Chapter introduces a holographic-based KT approach, which leverages MR for intuitive and straightforward HTR communication, enhancing the interaction layer by letting users instruct robot's tasks in the digital space. Among other advantages, the adoption of MR-based KT, while facilitating teaching process in those scenarios where robots already possess means for traditional KT, could represent an added values for all those robotic platforms that instead do not support it, thus extending the possibility of interaction in HRC contexts.

Given these premises, the Chapter starts by framing the KT process into the communication space formalized in Section 3.1.1, thus contextualizing the proposed approach within this thesis's work. Then, the software implementation is discussed, which extends the functionalities of *RICO-MR* to achieve holographic-based KT, thus realizing the envisioned bi-directional communication framework where both RTH and HTR aspects are accounted for. The architecture is then tested through an experimental campaign, aimed at assessing whether holographic-based KT is as effective and intuitive as traditional, manual KT. Both self-assessments and task-related metrics are extracted and a thorough comparison is reported, both in terms of UX and objective metrics.

7.1 Methods

7.1.1 Formalization

The first step towards properly framing KT into the holographic communication space developed so far involves identifying the relevant information I exchanged during such teaching sessions. In particular, it is argued that the act of KT implies teaching robots about their future states $\boldsymbol{\tau}(t)$, defined in accordance to (3.4). In turn, by teaching sequences of future states, operators achieve the intended effect of conveying whole trajectories to their robot teammate, either for manipulation or navigation purposes, thus maintaining consistency with the relationship expressed in (3.8). As such, it follows that the set of information I which can be conveyed during KT sessions can be modeled as $I = \{\boldsymbol{\tau}\}$.

Having defined the set I, we observe that KT is achieved by hand-guiding the robot's wrist or end-effector. According to the *C-Space* formalism expressed in (3.1), this act involves a gesture-mediated communication that enables users to teach robots about their future states in a simple way and can be described as follows

$$C_{gest}(I, \boldsymbol{t}_{gest}) = \mathbf{T}(\boldsymbol{t}_{gest}), \tag{7.1}$$

where $\mathbf{T}(\boldsymbol{t}_{gest})$ describes the robot's trajectory that is conveyed via gestural guidance during the interval $\boldsymbol{t}_{gest}$ spanning the KT session and whose definition follows from (3.8).

With this formalization in mind, it is argued that KT can be translated and framed into *MR-Space* by letting users convey robots' trajectories via gestural guidance on a virtual counterpart of the robot. Since a holographic version of the robot does not need ad-hoc sensors for KT, this would broaden the possibility of applying such a technique on every robotic platform. To further strengthen the communicative framework and ensure a more natural interaction, it is hypothesized that adding the vocal channel would improve users' experience, enabling them to control more detailed aspects of the KT session, including the *start* and *stop* on the taught robot trajectory, or the possibility to *open* and *close* the robot's gripper for teaching pick-and-place actions. According to such modeling, the holographic-based KT process is translated into a HTR communication act combining gestural and vocal interaction and, as such, can be formalized as follows

$$C^{KT}(I, \boldsymbol{t}) = C_{gest}(I, \boldsymbol{t}_{gest}) \,\cup\, C_{voc}(I, \boldsymbol{t}_{voc}). \tag{7.2}$$

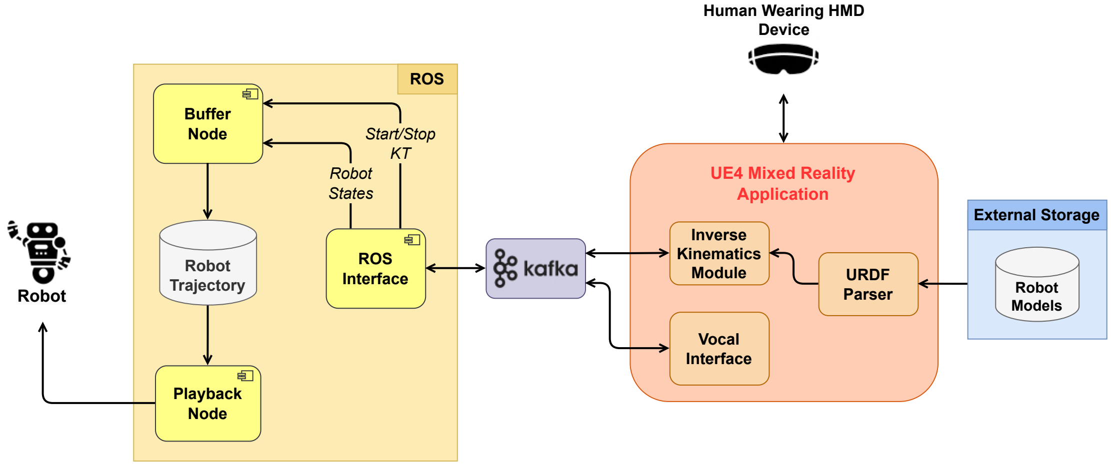


Figure 7.2 Overview of the proposed architecture implementing holographic KT, extending the framework detailed in Chapter 6

7.1.2 Software Architecture

The software components developed in the context of this work constitute a modular extension of *RICO-MR* and are publicly available under MIT licence in a separate branch of the main repository[1]. The proposed architecture exploits functionalities developed for *RICO-MR* to achieve the holographic KT envisioned in Section 7.1.1. However, currently, the architecture allows holographic KT with fixed manipulators only. As such, a simplification in the formalization is provided for the time being, and we hereafter refer to the notion of robot's state to indicate its joint configuration $\boldsymbol{q}(t)$ only.

Mixed Reality Application

Mixed Reality Application is directly mediated from the *RICO-MR* version. Whenever users load the application in their HMD device, they step into an empty holographic scene and employ the hand-attached menu to select the robot they wish to spawn for KT purposes. Aside from the already mentioned pre-loaded models that ship with the current version of the application, users can extend the list of supported robots by uploading corresponding URDF files to the customizable remote repository. As such, it is possible to employ the proposed application to carry out KT with any URDF-compliant robot.

Upon selecting the robot model, users can spawn it in the environment using the usual QR code as spatial anchor, exploiting *RICO-MR*'s already discussed features. Along with the robot model, a grey holographic sphere, visible in Fig. 7.1, is spawned and superimposed

---

[1] GitHub: https://github.com/TheEngineRoom-UniGe/RICO-MR/tree/kt

on the robot's wrist. This sphere serves as a point of interaction between the human and the robot. Using the hand-tracking capabilities of the HMD, the operator can directly manipulate the sphere by controlling its rotation and translation in space. The robot, in turn, follows the sphere and aligns its wrist's pose with it by solving the Inverse Kinematics (IK). To this extent, the Denavit-Hartenberg (DH) parameters necessary for the computation of the IK are extracted from the robot model's URDF and fed to the *IK Module*, which continuously computes the joint configuration needed to achieve the desired pose of the wrist. Specifically, the IK computation occurs with a rate of $30Hz$. As such, by interacting with the grey sphere and hand-guiding it, users can communicate future robot's states and, consequently, teach trajectories and actions to the robot teammate.

Consistently with the formalization given in Section 7.1.1, a voice interface is also active inside *Mixed Reality Application*. Four basic commands are available, ensuring that the user can control the *start* / *stop* of the KT session and the *open* / *closed* state of the robot's gripper, offering the possibility to teach more complex motions such as pick-and-place or handover actions.

Recording and Playback

While *Mixed Reality Application* provides the holographic interface to perform KT, recording and subsequent playback of the robot's actions take place respectively through *Kafka* and ROS. On the one hand, *Kafka* is employed for input / output data exchange with *Mixed Reality Application*. In particular, it streams the sequence of robot's states at a rate of $20Hz$, beginning as soon as the user signals the start of the KT session through vocal command.

On the other hand, two ROS nodes act respectively as *Buffer* for the robot's trajectory streamed through *Kafka* and *Playback* of the recorded motion. The *Buffer Node* subscribes to the *Kafka* topic to access the robot's states, and it saves them to file for later execution. To this end, the *ROS-Kafka Interface* is employed to convert incoming *Kafka* messages into their equivalent ROS representation. Finally, the *Playback Node* forwards state commands to the internal low-level controller of the robot at the same rate as the recording to reproduce the desired motion.

7.1.3 Experimental Validation

Hypotheses and Experimental Scenario

The experimental campaign discussed in this Chapter aims to determine if the proposed holographic KT approach can act as efficient HTR communication layer, as well as a valid

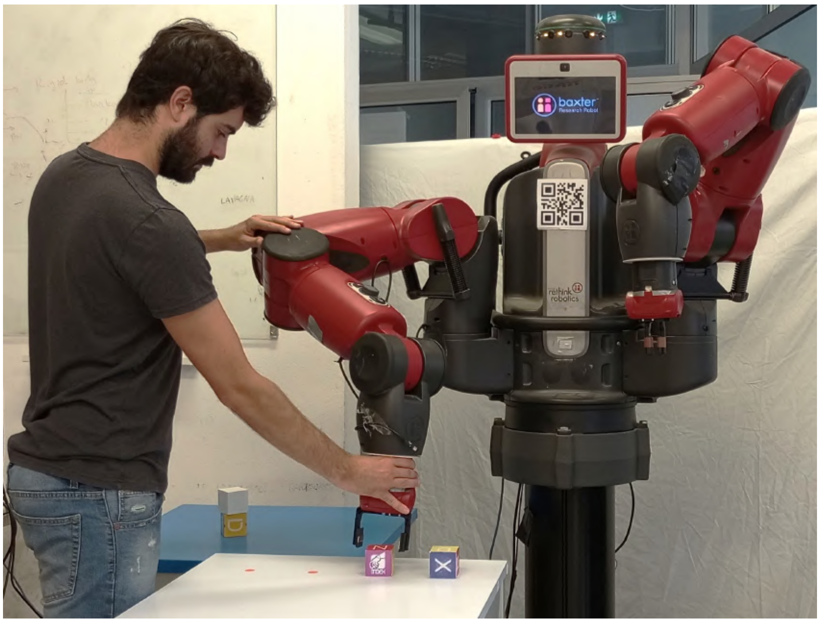


Figure 7.3 User interacting with Baxter during physical KT session. The operator drives the robot's arm through gestural interaction, teaching the sequence of pick-and-place actions needed to complete the stacking task.

alternative to standard KT for robots that do not possess the software and hardware means for physical hand guidance. To achieve such a result, a suitable human-robot interactive scenario has been devised, enabling comparison between traditional, physical KT and the novel holographic approach, with two different robot models involved for more generalized results. In particular, both Baxter and Tiago++ have been employed, given that both platforms natively support physical KT. Similarly, a Microsoft HoloLens 2 has been adopted, since it provides state-of-the-art hand tracking and voice interaction.

From a formal point of view, to provide a thorough comparison between physical KT and holographic KT, a series of hypotheses has been formulated, and evaluated through preliminary user study:

H1 Both approaches have equivalent HTR communicative power, that is, actions conveyed via holographic KT ensure similar playback outcomes as those taught through physical KT;

H2 The two KT techniques provide comparable UX results.

Regarding the interactive task employed to evaluate the two KT alternatives, a simple *stacking task* has been devised. Specifically, the human should use KT to teach a sequence of pick-and-place actions aimed at stacking four cubes on top of each other according to a predefined order. Fig. 7.3 depicts the experimental scenario, showing a user in the middle of a physical KT session with the Baxter robot.

User Study

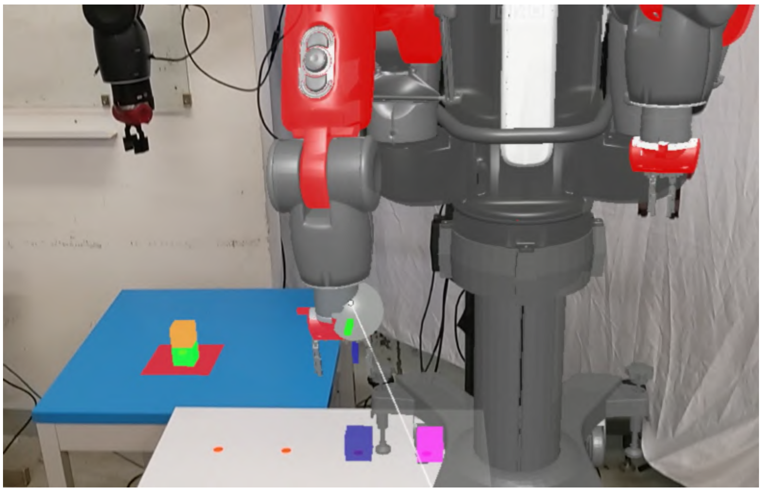
An experimental campaign has been carried out with $K = 12$ volunteers (9 males and 3 females), all aged between 21-32 ($Avg = 26.3$, $StdDev = 3.07$) and having limited or null experience with MR and HMD devices. The subjects were divided into two groups. The first group performed the experiment with Tiago++, while the second group used Baxter. In both groups, subjects were asked to perform the KT session in two different experimental conditions, namely

C1 Without wearing the HMD and performing traditional, physical KT;

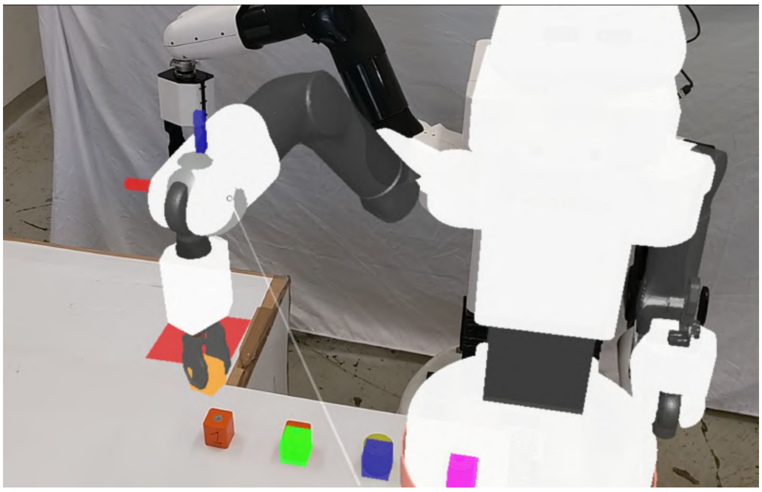
C2 Wearing the HMD and performing holographic KT.

Participants were initially instructed on the stacking task and assigned a randomized order for the cubes to be collected. Then, they performed their first trial, in condition *C1*. Subsequently, before beginning the experiment with HMD, subjects were briefly instructed on how to interact with the HoloLens holographic menus and interface. Then, once accustomed, they carried out their second trial in condition *C2*. To achieve a consistent KT experience, the holographic interface in condition *C2* included also four virtual cubes placed coherently with their real-world counterparts, as shown in Fig. 7.4. Such virtual cubes were physics-enabled and behaved like the real ones, aiding the participant in recording the holographic KT session. In both cases, the voice interface was active for controlling the *start / stop* of the KT session and the *open / closed* state of the robot's gripper. However, while in condition *C2* the vocal interface was embedded into the holographic application running on the HoloLens 2, in condition *C1* it was simulated via *Wizard of Oz* device.

After successfully completing each KT session, the playback phase was manually triggered, causing the robot to reproduce the taught action. This phase allowed us to rank the KT session quantitatively through a combination of two distinct variables, useful in the evaluation of *H1*. On the one hand, we counted the number of cubes successfully stacked by the robot during playback. As such, it was possible to evaluate the communicative capabilities of each KT alternative, assessing how well the combination of vocal and gestural HTR communication translated into the corresponding robot action. On the other hand, we recorded the duration of each KT session and employed such quantity to compare the two techniques in terms of time necessary to teach the desired action.



(a) Holographic KT with Baxter



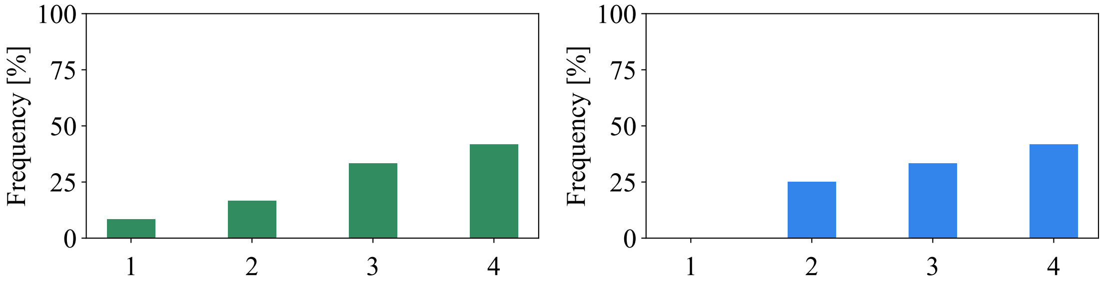
(b) Holographic KT with Tiago

Figure 7.4 Screenshots taken from within the holographic interface experienced by participants. The MR layer takes advantage of UE4’s physics to achieve a realistic KT experience in the digital space, with users being able to teach the full sequence of pick-and-place actions on the holographic entities.

Finally, after completing their trials, each participant was required to fill out the UEQ, to find out and compare the UX of the two different KT strategies. In accordance with hypothesis *H2*, to provide a consistent comparison between the two KT techniques, each participant compiled the UEQ twice, thus evaluating both physical and holographic KT sessions from a UX point of view.

7.2 Results

The present Section reports and discusses the results obtained from the preliminary user study. In particular, it has been observed that, regardless of the robot, the two groups of subjects achieved comparable results when teaching the stacking task in both experimental conditions. As such, Fig. 7.5 reports only the aggregated results, comparing conditions *C1* and *C2* without discerning the interactions occurred with Tiago++ or Baxter. The histograms show the percentage of playback sessions where the robot successfully stacked a certain number of cubes. For example, in both experimental conditions, around 40% of the subjects achieved a flawless KT, resulting in the robot successfully stacking all four cubes while replaying the taught trajectory.



(a) Cubes stacked in condition *C1*.

(b) Cubes stacked in condition *C2*.

Figure 7.5 Histograms depicting the number of cubes successfully stacked by the robots during the playback phase, in the two experimental conditions.

By observing the plots of Fig. 7.5, it is possible to note how physical and holographic KT yielded comparable results. Keeping into account that such distributions could not be assumed normal, a statistical evaluation of the two conditions has been performed via the non parametric one-tailed Wilcoxon signed-rank test. The test provided a statistic $W = 20$, with p-value> 0.3. Such result was compared with the critical one $W_c = 17$, obtained by fixing the population size K and the significance level $\alpha = 0.05$. By observing that $W > W_c$, the null hypothesis could not be rejected and therefore it was possible to conclude that no significant difference was highlighted between *C1* and *C2*. This result, in turn, allowed us to confirm the initial hypothesis *H1*, possibly indicating that the two communicative interfaces (i.e., physical and holographic) ensure consistent performances while executing KT.

Regarding the overall time needed to perform KT, it has been observed that in condition *C2* participants were always slower because of their limited expertise with MR devices. As such, a differential analysis has been carried out by computing, for each participant, the difference in terms of time taken to complete the KT session between condition *C2*

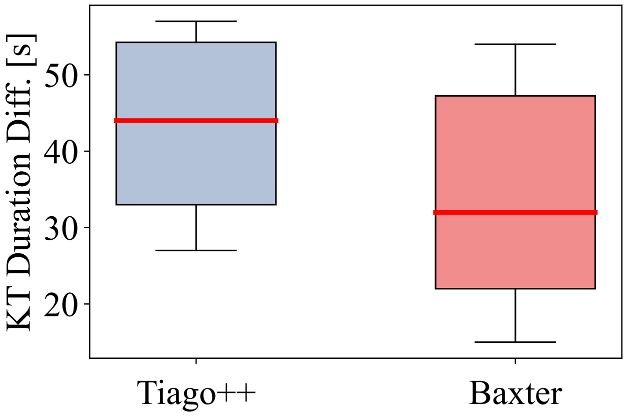
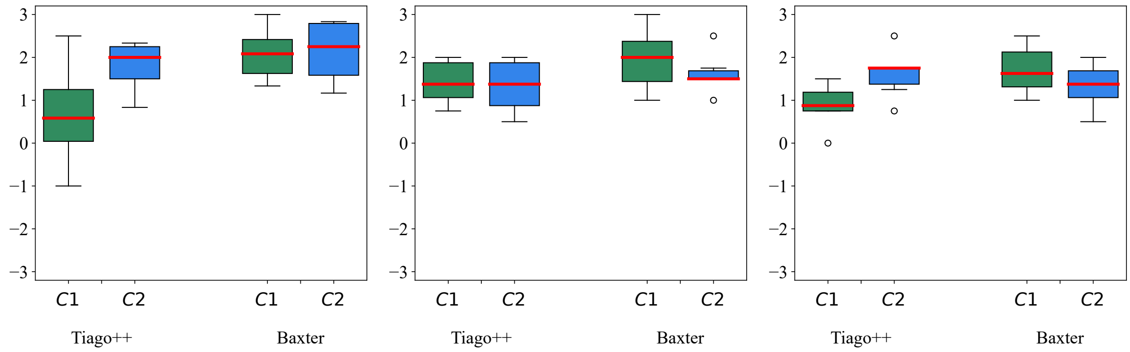


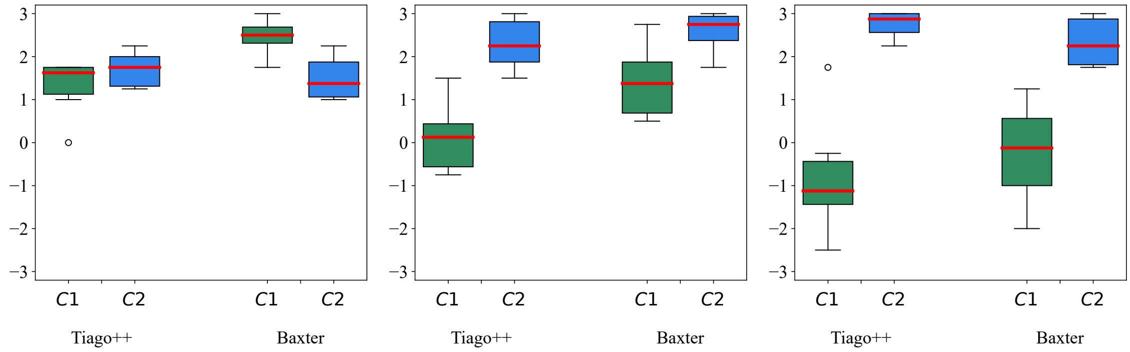
Figure 7.6 Differential distributions depicting the temporal overhead introduced by the MR channel when performing KT under *C2*.

and *C1*. These results are reported in Fig. 7.6. The boxplots highlight that, on average, holographic KT lasted, respectively for Tiago++ and Baxter, 44 and 32 seconds longer than the corresponding physical sessions. Compared with the average times measured to complete the physical KT sessions with the two robots, the MR-based approach introduced, respectively, a mean temporal overhead of 37% and 33%. However, Fig. 7.6 shows no significant difference between temporal overheads when using one robot or the other. This result is also confirmed by the T-test on the two differential distributions, which yielded a p-value>0.2. In other words, the overhead introduced by the MR channel was consistent among the two robots. As such, it is hereby argued that, while slowing down the experience when interacting with robots who natively possess the hardware-software components needed for physical KT, the proposed approach would potentially provide an added value in all those contexts where KT is otherwise not available and robot tasks are still manually programmed.

Finally, Fig. 7.7 reports the results obtained from the UEQ questionnaires, grouped per evaluation scale and robot type. Here, scores range in the interval $[-3,3]$, with positive values indicating features that users appreciated given a particular interface. Specifically, Fig. 7.7c and 7.7b highlight that both KT approaches provided comparable results in terms of *efficiency* and *perspicuity* (i.e., how intuitive and pragmatic the interface appeared to users), regardless of the robot employed. Such results are corroborated by statistical analysis performed through Kruskal-Wallis test. The test yielded, for both scales, p-values >0.05, indicating no significant difference between the distributions. Furthermore, holographic KT scored particularly well in terms of *attractiveness*, *stimulation* and *novelty*, suggesting that participants found the interaction with the holographic environment more engaging and original compared to the physical one. The only scale where holographic KT did a slightly



(a) Attractiveness (b) Perspicuity (c) Efficiency



(d) Dependability (e) Stimulation (f) Novelty

Figure 7.7 Measured UEQ scores on the six evaluation scales, grouped by robot type and experimental conditions. The median value for each distribution is plotted as a red line.

worse job is *dependability*, which measures how safe and predictable the users perceive a given interface. In this case, physical KT was still perceived as more predictable, particularly with the robot Baxter, compared to the MR-based approach, which nonetheless obtained positive scores with both robots.

7.3 Discussion

This Chapter marked a first attempt at tackling the aspect of efficient, holographic HTR communication, introducing a novel approach at KT in collaborative settings. As a matter of fact, KT is considered, in the context of this thesis, as a form of communication where the human operator instructs the robot teammate about upcoming tasks and actions through physical, hand-guided demonstration. In accordance with this definition, KT has been modeled as an HTR communicative act in the *C-Space* formalism, aimed at conveying future robot's state trajectories **T** via gestural interaction.

Recognizing the significance of KT in manufacturing settings and acknowledging the challenges in achieving efficient, platform-independent implementations, this Chapter introduces a universal framework for MR-based KT. The framework leverages the holographic layer and the potential of *RICO-MR* to facilitate such teaching process nearly on any robotic platform. To this regard, the holographic-based KT act is first described in the context of *MR-Space* as a combination of vocal and gestural interaction applied to the holographic robot, to effectively convey its future state trajectory as previously formalized. Then, a practical implementation of this framework is presented, which builds upon the existing *RICO-MR* architecture and expands its holographic capabilities to account for such form of HTR communication. The interface is then tested in a user study involving two different robot models and multiple users, in an attempt to compare holographic-based KT with traditional, physical KT. Specifically, both objective and subjective evaluations have been collected and the findings suggest that the communicative capabilities of holographic KT match those of hand-guided demonstrations, with similar results on UX as well. This could, in turn, suggest that MR-based KT is an effective HTR communication act and could serve as a valid alternative to physical KT in all those scenarios where the robotic platform employed does not natively support such teaching mechanism, thus widening the spectrum of interaction between human and robot.

This latest implementation marks the last step of this Ph.D. work. Overall, the final MR-based architecture supports both aspects of holographic communication, thus providing an integrated and, possibly, valuable layer to facilitate collaboration between human and robot. On the one hand, the RTH aspect ensures that robot's intentions can be conveyed throughout collaborative processes using the DHT paradigm, which offers dynamic, expressive intention cues in both fixed and mobile interactive scenarios. On the other hand, the HTR aspect adds a convenient, platform-independent interaction layer, where human operators can communicate robot's state trajectories and teach tasks to their teammate using intuitive, holographic-based KT. In-between these features, the *RICO-MR* architecture stands providing modular, scalable, cloud-based integration for additional frameworks and components, such as external DTs, that could be complemented to build even more comprehensive and sophisticated communication layers.

Chapter 8

Conclusions

8.1 Overview

In this thesis, the importance of effective communication to foster more efficient collaboration in human-robot teams has been addressed. This aspect is particularly relevant in manufacturing scenarios, where collaborative platforms and individuals work together in close-proximity and where important factors such as team coordination and synchronization could benefit from a robust, intuitive communication layer involving the agents. In particular, this thesis's work has been inspired by relevant studies on human-human collaboration, whose findings suggest how individuals employ a whole layer of implicit communication to convey and infer each other's intentions, thus naturally maximizing coordination and efficiency. As such, the present thesis has postulated how a similar approach could be undertaken to achieve a more natural and efficient collaboration in hybrid human-robot teams, introducing an intuitive channel capable of ensuring straightforward communication between agents, enabling them to exchange meaningful intention cues throughout the interaction.

Such an objective, however, required as first step the identification of an expressive communication channel, suitable enough for unstructured industrial scenarios. After a thorough analysis of relevant literature, the choice fell on MR, a promising, emerging technology which blends together real and virtual world to create hybrid experiences where holographic entities and real objects interact. In this context, MR, perceived by human operators thanks to compact, wearable HMD devices, could enable the design of expressive, meaningful interfaces, ensuring effective and straightforward communication between human and robot. With such a technology at disposal, the main challenge of this Ph.D. work consisted in the construction of a bi-directional communication layer, leveraging holographic content to tackle both aspects of RTH and HTR communication during collaborative activities, with the final aim of improving team efficiency and teamwork.

In light of these premises, the following sections provide a summary of the main contributions of this thesis, detailing how the holographic channel has been leveraged to represent and communicate robots' intentions intuitively, to ensure simplified and more straightforward KT sessions, aimed at instructing robots' tasks within the MR layer, and, finally, to build a modular, generalized software architecture easily re-usable by other practitioners and companies, interested in taking advantage of said communication scheme.

8.1.1 RTH Communication

The first part of this Ph.D. journey consisted in the development of a structured, RTH communicative framework aimed at conveying robot's intentions to the human teammate throughout collaborative processes, leveraging the holographic layer. In this context, a preliminary step undertaken towards such goal has been the introduction of an analytical framework, denoted *C-Space*, that facilitates modeling communication acts in generalized HRC contexts. Such communicative space, while potentially applicable in various interactive domains, served as mean to define and analytically represent robot's intentions, simplifying the subsequent translation process from theoretical formalism to practical, expressive holographic cues at implementation-level. More in detail, the ability to model robots' intentions as series of future states $\boldsymbol{\tau}$ and beliefs $\boldsymbol{\omega}$ opened up the possibility to express both simple and complex intentions cues, ranging from the communication of simple motions and trajectories, to more sophisticated conveyance of actions involving objects in the collaborative workspace, such as handovers. At software-level, a first implementation of the holographic interface, named *MR-HRC-V1*, has been achieved following the proposed DHT paradigm, which ensures dynamic communication acts, where animated, digital overlays are leveraged to convey robot's intentions in a meaningful way without cluttering the users' field of view, as it occurred in previous, related research studies. To this regard, a preliminary user study has been conducted in a collaborative assembly scenario, where a human operator was required to cooperate with the robot Baxter while wearing the HMD to experience the aforementioned holographic communication scheme. Preliminary findings observed suggested that the possibility to experience robot's intentions as digital overlays improved team coordination and contributed towards a more fluent collaboration.

In light of these results, a more comprehensive user study has been undertaken, with the aim of generalizing the aforementioned results, while at the same time comparing the DHT communication scheme with other relevant approaches from the state-of-the-art, both from objective and subjective perspectives. A similar collaborative assembly scenario has therefore been devised, and and extensive user study with 60 participants has been conducted, comparing three holographic communication schemes, which differed in how robot's intentions were represented as digital overlays to the human operator during collaboration. Relevant task-related metrics have been measured, and subjective participants' experience has been appraised through the UEQ, a popular self-assessment tool aimed at estimating the UX of interactive products. The overall results highlighted the positive impact of MR-based, RTH communication on the collaboration, with participants showcasing improved coordination with the robot teammate, and an increased rate of successful joint actions (e.g., handovers). Additionally, the DHT scheme reported higher results in all scales of the UEQ, suggesting how such form of holographic communication was appraised as more efficient, straightforward and intuitive, given its dynamic and expressive nature, compared to related approaches.

Following on these findings, the subsequent step of the work involved generalization of the communication scheme to broader domains of HRC, relaxing the constraint of collaboration under fixed workstations. An updated version of the software architecture, denoted *MR-HRC-V2* has been designed, capable of projecting holographic intention cues not only for fixed manipulator robots, but also for mobile platforms, with the possibility, for the user, to preview upcoming navigation trajectories as animated, digital overlays, thus offering a new RTH communicative layer in logistics or warehouse scenarios where operators and moving artificial agents coexist and interact. To evaluate the effectiveness of holographic communication in such contexts, a warehouse-like experimental scenario has been devised, and the updated MR interface has been tested with a third user study, involving 20 subjects and the mobile manipulator robot Tiago++. In such simulated setting, user and robot were required to carry out parallel, independent tasks, moving around the dynamic environment, and were supposed to cooperate and interact when certain conditions arose. Throughout this study, only objective, task-related metrics were assessed and the subsequent results highlighted how MR-based communication contributed to a more fluent interaction between agents, with individuals more aware of the robot's actions thanks to the holographic cues, which improved team synchronization and led to faster task completion on the human's side.

Throughout this first part of the research, evidences were collected on the effectiveness of holographic communication in fostering fluent interactions and more coordinated collaborations, with human operators taking advantage of the RTH communicative acts to synchronize with the teammate and behave more proactively towards it. While the possibilities of applications in real-world manufacturing contexts are promising, further studies could be undertaken before transferring such technology to the industrial world, to evaluate more psychological quantities bearing on the user, such as perceived safety and cognitive load, and how the additional holographic layer impacts on them.

8.1.2 HTR Communication

The HTR aspect has been tackled with two different technologies during this Ph.D. work. On the one hand, the *MR-HRC-V2* architecture marked a first attempt at integrating DT and MR in a comprehensive communicative framework for HRC, where the digital model played a crucial role for HTR communication. In particular, the DT served as a virtual replica of the collaborative scenario, monitoring the agents' state in real-time thanks to a combination of sensory information being acquired in the real experimental setting. In this regard, the human operator was continuously tracked, and implicit, non-verbal intention cues were extracted from a combination of their gaze and posture. Once detected by the DT, such intentionality was signaled to the robot teammate, which proceeded to trigger certain behavioral logic in response. Throughout the experimental campaign in the warehouse-like scenario, the capabilities and responsiveness of the DT have been put to test, assessing its ability to infer human intention cues. The subsequent results highlighted how such digital models can effectively be leveraged for online recognizing non-verbal, HTR communication in HRC contexts.

On the other hand, the aspect of holographic, HTR communication has been tackled in the last part of this Ph.D journey. In contrast with the plethora of approaches that leverage MR to program robots' behaviors, this thesis focused instead on a more specific approach, represented by KT. In particular, the KT paradigm has been framed as HTR communicative technique aimed at instructing robot's state trajectories, and, as such, a corresponding formalization within the *C-Space* has been issued. Based on such formalism, a novel approach at KT has been introduced, leveraging the MR layer to teach trajectories and tasks to a holographic counterpart of the robot, using a combination of operator's voice and gestures, subsequently transferring the learned skill to the real robot for execution. Such an approach, while offering an intuitive, holographic, universal methodology for KT, ensures that such teaching paradigm could be applied to any URDF-compliant robotic platform, regardless of their underlying control schemes and complex kinematic chain. A practical implementation of this MR-based KT interface has been developed, and put to test against traditional, physical KT to assess its communicative capabilities and UX. A user study has therefore been conducted, utilizing Baxter and Tiago as real robotic platforms for both hand-guided and holographic KT. Overall, the results demonstrated how holographic KT behaved comparably to physical KT, both in terms of teaching effectiveness and UX, highlighting how such novel techniques acts as efficient HTR communication.

8.1.3 A Modular Architecture

The latest installment of the software architecture, that is the *RICO-MR* version, provides a generalized, modular implementation of the envisioned communication framework, while adhering to the open-source paradigm. To this regard, the *generalizability* aspect of the architecture derives from it being independent of the particular HRC scenario and robotic platform adopted. Specifically, the possibility to load robots' holograms at runtime ensures that *RICO-MR* can be used as RTH communicative layer in any collaborative context, even in settings where multiple robots are involved. Additionally, the ability to upload and use custom URDF files without manually compiling the code further strengthens the re-usability aspect, granting off-the-shelf access to end-users of the MR application.

On the other hand, the *modularity* aspect derives from the extendable nature of the architecture itself. The usage of *Apache Kafka* as main data exchange infrastructure ensures that the architecture can easily accommodate multi-robot and multi-user experiences, while also offering easy integration with ROS-based applications and ROS-independent ones, thanks to the plethora of open-source software libraries that interface with *Kafka*. As an example, such modularity has been taken advantage of when the HTR communication layer for KT has been developed, adding the IK component to the MR application and leveraging *Kafka* to stream and record robot's state trajectories during teaching sessions. In addition, the abstract nature of *Kafka*'s communication protocol ensures straightforward integration with external software applications, and even with cloud-based solutions. In this context, future research efforts could be oriented towards building a more comprehensive communicative framework combining *RICO-MR* and a cloud-hosted DT, leveraging the potential of *Kafka* for reliable, low-latency data throughput.

Finally, it is worth mentioning a collateral aspect of *RICO-MR*, that is the possibility to act not only as pure MR-based communication layer, but also to play the role of holographic simulation tool, thanks to the adoption of UE4 as game engine for the MR application. To this regard, the realistic physics layer ensures consistent interactions between holographic objects and real environment, paving the way for future research on MR-based simulations as educational or training tools.

8.2 Final Considerations, Limitations, Future Works

The work presented in this thesis aimed to study the role of effective communication in industrial HRC, and how MR could play a significant part in rendering human-robot teams more efficient. Although the final result of this thesis comprises of a robust software architecture, implementing the envisioned bi-directional, holographic communication interface, several additional studies could be undertaken to provide more generalized evidence of how MR influences collaboration and whether this technology can be successfully transferred to real-world industrial settings. This final paragraph summarizes limitations and possible future research directions.

On the one hand, one recurring hypothesis that has only been partially proved throughout this thesis is whether MR-based communication can increase collaborative pace and reduce task completion times. To effectively test such hypothesis, we should reference work from Hoffman [70] and evaluate if holographic communication improves the overall percentage of concurrent activity among agents. Such a validation, however, would require a proper bi-directional communication layer in place to effectively assess the impact of MR on the collaboration. Given that this Ph.D. work has been dedicated to building such holographic scheme, it is therefore postulated that future research efforts could be undertaken to validate the combined communication of robot's intentions and MR-based KT in a suitable collaborative scenario, in an attempt to appraise how they impact task pace and completion times.

On the other hand, while RTH holographic communication has demonstrated, under multiple circumstances, positive effects in terms of human teammate's awareness and proactivity, with consequent improvements on team coordination as well, it could be crucial to evaluate the psychological aspects derived from the adoption of such MR-based scheme. In particular, this could be of significant aid in appraising operators' stress levels and cognitive loads when subject to such form of holographic communication, in order to find out the minimal subset of digital overlays to project that is equally expressive of robots' intentions, but minimally invasive on the operator's field of view.

In addition, future research works could also extend and adapt the capabilities of holographic KT to account for mobile robot platforms as well, in order to broaden the range of tasks and actions that can be conveyed via HTR digital interaction.

Publications

A list of publications, either published or in the publication process, that relate to the thesis's contribution or were made during the Ph.D. period is given below.

Published

- **Macciò S.**, Carfì A., Mastrogiovanni F. (2021) A Mixed Reality Architecture for Human-Robot Collaboration. In Proceedings of the 3rd Italian Conference on Robotics and Intelligent Machines (I-RIM 2021). DOI 10.5281/zenodo.5900561
- **Macciò S.** , Carfì A., Mastrogiovanni F. (2022) Mixed Reality as Communication Medium for Human-Robot Collaboration. In Proceedings of the 39th IEEE International Conference on Robotics and Automation (ICRA 2022). DOI 10.1109/ICRA46639.2022.9812233
- Bongiovanni A., De Luca A., Gava L., Grassi L., Lagomarsino M., Lapolla M., Marino A., Roncagliolo P., **Macciò S.**, Carfì A., Mastrogiovanni F. (2022) Gestural and touchscreen interaction for human-robot collaboration: a comparative study. In Proceedings of the 17th International Conference on Intelligent Autonomous Systems (IAS-17). DOI 10.1007/978-3-031-22216-0_9
- Bonzini A., Seminara L., **Macciò S.**, Carfì A., Jamone L. (2022) Leveraging Symmetry Detection to Speed up Haptic Object Exploration in Robots. In Proceedings of the IEEE International Conference on Development and Learning (ICDL 2022). DOI 10.1109/ICDL53763.2022.9962206
- Shaaban M., **Macciò S.**, Carfì A., Mastrogiovanni F. (2022) Integrating Digital Twin and Mixed Reality in Human-Robot Collaboration. In Proceedings of the 4th Italian Conference on Robotics and Intelligent Machines (I-RIM 2022). DOI 10.5281/zenodo.7531302

- Lastrico R., **Macciò S.**, Carfì A., Traverso P., Mastrogiovanni F. (2023) Estimation of Kidney's Blood Vessels Deformations for Robot-Assisted Surgery. In Proceedings of the 18th International Conference on Intelligent Autonomous Systems (IAS-18).

- **Macciò S.**, Shaaban M., Carfì A., Zaccaria R., Mastrogiovanni F. (2023) RICO-MR: An open-source architecture for robot intent communication through mixed reality. In Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023). DOI 10.1109/RO-MAN57019.2023.10309471

Submitted or in preparation

- **Macciò S.**, Carfì A., Mastrogiovanni F. (2022) Communicating Robot's Intentions Through Mixed Reality. **Submitted to:** Robotics and Autonomous Systems (RAS) journal

- Shaaban M., **Macciò S.**, Carfì A., Mastrogiovanni F. (2024) Investigating Mixed Reality for Communication Between Humans and Mobile Manipulators. **Submitted to:** 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2024)

- **Macciò S.**, Capitanelli A., Carfì A., Picardi M., Tropea P. (2023) Machine learning based risk of fall assessment in post-rehabilitation patients using mobility tests outcomes. **Manuscript in preparation** for the Artificial Intelligence in Medicine journal

- **Macciò S.**, Carfì A., Mastrogiovanni F. (2024) Kinesthetic Teaching in Robotics: a Mixed Reality Approach. **Submitted to:** 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2024)

References

[1] Aaltonen, I., Salmi, T., and Marstio, I. (2018). Refining levels of collaboration to support the design and evaluation of human-robot interaction in the manufacturing industry. *Procedia CIRP*, 72:93–98.

[2] Ajoudani, A., Zanchettin, A. M., Ivaldi, S., Albu-Schäffer, A., Kosuge, K., and Khatib, O. (2018). Progress and prospects of the human–robot collaboration. *Autonomous Robots*, 42(5):957–975.

[3] Akkaladevi, S. C., Plasch, M., Maddukuri, S., Eitzinger, C., Pichler, A., and Rinner, B. (2018). Toward an interactive reinforcement based learning framework for human robot collaborative assembly processes. *Frontiers in Robotics and AI*, 5:126.

[4] Andersen, R. S., Madsen, O., Moeslund, T. B., and Amor, H. B. (2016). Projecting robot intentions into human environments. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 294–301, New York, USA.

[5] Arents, J., Abolins, V., Judvaitis, J., Vismanis, O., Oraby, A., and Ozols, K. (2021). Human–robot collaboration trends and safety aspects: A systematic review. *Journal of Sensor and Actuator Networks*, 10(3):48.

[6] Arévalo, S., Dierks, T., Rücker, F., and Gerken, J. (2020). There's more than meets the eye: Enhancing robot control through augmented visual cues. In *Companion of the 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 104–106, Cambridge United, Kingdom.

[7] Arkin, J., Paul, R., Park, D., Roy, S., Roy, N., and Howard, T. M. (2020). Real-time human-robot communication for manipulation tasks in partially observed environments. In *Proceedings of the 2018 International Symposium on Experimental Robotics*, pages 448–460. Springer.

[8] Arntz, A., Eimler, S. C., and Hoppe, H. U. (2020). "the robot-arm talks back to me"-human perception of augmented human-robot collaboration in virtual reality. In *Proceedings of the 3rd IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 307–312.

[9] Ateş, G. and Kyrkjebø, E. (2022). Design of a gamified training system for human-robot cooperation. In *Proceedings of the 2nd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–7, Maldives.

[10] Baraglia, J., Cakmak, M., Nagai, Y., Rao, R. P., and Asada, M. (2017). Efficient human-robot collaboration: when should a robot take initiative? *The International Journal of Robotics Research*, 36(5-7):563–579.

[11] Baratta, A., Cimino, A., Longo, F., and Nicoletti, L. (2023). Digital twin for human-robot collaboration enhancement in manufacturing systems: literature review and direction for future developments. *Computers & Industrial Engineering*, page 109764.

[12] Bethel, C. L. and Murphy, R. R. (2010). Review of human studies methods in HRI and recommendations. *International Journal of Social Robotics*, 2(4):347–359.

[13] Bolano, G., Fu, Y., Roennau, A., and Dillmann, R. (2021a). Deploying multi-modal communication using augmented reality in a shared workspace. In *Proceedings of the 18th International Conference on Ubiquitous Robots (UR)*, pages 302–307. IEEE.

[14] Bolano, G., Iviani, L., Roennau, A., and Dillmann, R. (2021b). Design and evaluation of a framework for reciprocal speech interaction in human-robot collaboration. In *Proceedings of the 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 806–812, Vancouver, Canada.

[15] Bolano, G., Juelg, C., Roennau, A., and Dillmann, R. (2019). Transparent robot behavior using augmented reality in close human-robot interaction. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7, New Delhi, India.

[16] Bongiovanni, A., De Luca, A., Gava, L., Grassi, L., Lagomarsino, M., Lapolla, M., Marino, A., Roncagliolo, P., Macciò, S., Carfì, A., and Mastrogiovanni, F. (2022). Gestural and touchscreen interaction for human-robot collaboration: a comparative study. *Proceedings of the 17th International Conference on Intelligent Autonomous Systems (IAS-17)*.

[17] Böttjer, T., Tola, D., Kakavandi, F., Wewer, C. R., Ramanujan, D., Gomes, C., Larsen, P. G., and Iosifidis, A. (2023). A review of unit level digital twin applications in the manufacturing industry. *CIRP Journal of Manufacturing Science and Technology*, 45:162–189.

[18] Bright, T., Adali, S., and Bright, G. (2021). Close human robot collaboration by means of a low-cost sensory glove for advanced manufacturing systems. In *Proceedings of the 1st International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–8, Mauritius.

[19] Brooke, J. (2013). Sus: a retrospective. *Journal of usability studies*, 8(2):29–40.

[20] Buisan, G., Sarthou, G., and Alami, R. (2020). Human aware task planning using verbal communication feasibility and costs. In *Proceedings of the 12th International Conference on Social Robotics (ICSR)*, pages 554–565, Golden, CO, USA.

[21] Burgoon, J. K., Guerrero, L. K., and Manusov, V. (2011). Nonverbal signals. *Handbook of interpersonal communication*, pages 239–280.

[22] Calinon, S. (2018). Learning from demonstration (programming by demonstration). *Encyclopedia of robotics*, pages 1–8.

[23] Calisgan, E., Haddadi, A., Van der Loos, H. M., Alcazar, J. A., and Croft, E. A. (2012). Identifying nonverbal cues for automated human-robot turn-taking. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 418–423, Paris, France.

[24] Cao, H.-L., Scholz, C., De Winter, J., Makrini, I. E., and Vanderborght, B. (2023). Investigating the role of multi-modal social cues in human-robot collaboration in industrial settings. *International Journal of Social Robotics*, pages 1–11.

[25] Cao, Y., Xu, Z., Li, F., Zhong, W., Huo, K., and Ramani, K. (2019). V. ra: An in-situ visual authoring system for robot-iot task planning with augmented reality. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS)*, pages 1059–1070, San Diego, CA, USA.

[26] Cardenas-Robledo, L. A., Hernández-Uribe, Ó., Reta, C., and Cantoral-Ceballos, J. A. (2022). Extended reality applications in industry 4.0.-a systematic literature review. *Telematics and Informatics*, page 101863.

[27] Carfì, A. and Mastrogiovanni, F. (2021). Gesture-based human-machine interaction: Taxonomy, problem definition, and analysis. *IEEE Transactions on Cybernetics*, 53(1).

[28] Carfi, A., Motolese, C., Bruno, B., and Mastrogiovanni, F. (2018). Online human gesture recognition using recurrent neural networks and wearable sensors. In *Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 188–195, Nanjing and Tai'an, China.

[29] Castro, A., Silva, F., and Santos, V. (2021). Trends of human-robot collaboration in industry contexts: Handover, learning, and metrics. *Sensors*, 21(12):4113.

[30] Cha, E., Matarić, M., and Fong, T. (2016). Nonverbal signaling for non-humanoid robots during human-robot collaboration. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 601–602, Christchurch, New Zealand.

[31] Chacko, S. M. and Kapila, V. (2019a). Augmented reality as a medium for human-robot collaborative tasks. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–8, New Delhi, India.

[32] Chacko, S. M. and Kapila, V. (2019b). An augmented reality interface for human-robot interaction in unconstrained environments. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3222–3228, Macau.

[33] Chadalavada, R. T., Andreasson, H., Krug, R., and Lilienthal, A. J. (2015). That's on my mind! robot to human intention communication through on-board projection on shared floor space. In *Proceedings of the 7th European Conference on Mobile Robots (ECMR)*, pages 1–6, Lincoln, United Kingdom.

[34] Chadalavada, R. T., Andreasson, H., Schindler, M., Palm, R., and Lilienthal, A. J. (2020). Bi-directional navigation intent communication using spatial augmented reality and eye-tracking glasses for improved safety in human–robot interaction. *Robotics and Computer-Integrated Manufacturing*, 61:101830.

[35] Chan, K. Y., Yiu, C. K., Dillon, T. S., Nordholm, S., and Ling, S. H. (2012). Enhancement of speech recognitions for control automation using an intelligent particle swarm optimization. *IEEE Transactions on Industrial Informatics*, 8(4):869–879.

[36] Chan, W. P., Hanks, G., Sakr, M., Zhang, H., Zuo, T., Van der Loos, H. M., and Croft, E. (2022). Design and evaluation of an augmented reality head-mounted display interface for human robot teams collaborating in physically shared manufacturing tasks. *ACM Transactions on Human-Robot Interaction (THRI)*, 11(3):1–19.

[37] Chan, W. P., Hanks, G., Sakr, M., Zuo, T., Van der Loos, H. M., and Croft, E. (2020). An augmented reality human-robot physical collaboration interface design for shared, large-scale, labour-intensive manufacturing tasks. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11308–11313.

[38] Chandan, K., Kudalkar, V., Li, X., and Zhang, S. (2021). Arroch: Augmented reality for robots collaborating with a human. In *Proceedings of the 38th IEEE International Conference on Robotics and Automation (ICRA)*, pages 3787–3793, Xi'an, China. IEEE.

[39] Chen, H., Leu, M. C., and Yin, Z. (2022). Real-time multi-modal human–robot collaboration using gestures and speech. *Journal of Manufacturing Science and Engineering*, 144(10):101007.

[40] Chitta, S., Sucan, I., and Cousins, S. (2012). Moveit! *IEEE Robotics & Automation Magazine*, 19(1):18–19.

[41] Coovert, M. D., Lee, T., Shindev, I., and Sun, Y. (2014). Spatial augmented reality as a method for a mobile robot to communicate intended movement. *Computers in Human Behavior*, 34:241–248.

[42] Darvish, K., Wanderlingh, F., Bruno, B., Simetti, E., Mastrogiovanni, F., and Casalino, G. (2018). Flexible human–robot cooperation models for assisted shop-floor tasks. *Mechatronics*, 51:97–114.

[43] De Simone, V., Di Pasquale, V., Giubileo, V., and Miranda, S. (2022). Human-robot collaboration: An analysis of worker's performance. *Procedia Computer Science*, 200:1540–1549.

[44] Devin, S. and Alami, R. (2016). An implemented theory of mind to improve human-robot shared plans execution. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 319–326, Christchurch, New Zealand.

[45] Dornelles, J. d. A., Ayala, N. F., and Frank, A. G. (2023). Collaborative or substitutive robots? effects on workers' skills in manufacturing activities. *International Journal of Production Research*, 61(22):7922–7955.

[46] Dragan, A. D., Bauman, S., Forlizzi, J., and Srinivasa, S. S. (2015). Effects of robot motion on human-robot collaboration. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 51–58, Portland OR, USA.

[47] Dragan, A. D., Lee, K. C., and Srinivasa, S. S. (2013). Legibility and predictability of robot motion. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 301–308, Tokyo, Japan.

[48] Du, G., Chen, M., Liu, C., Zhang, B., and Zhang, P. (2018). Online robot teaching with natural human–robot interaction. *IEEE Transactions on Industrial Electronics*, 65(12):9571–9581.

[49] Duan, J., Fang, Y., Zhang, Q., and Qin, J. (2023). Hrc of intelligent assembly system based on multimodal gesture control. *The International Journal of Advanced Manufacturing Technology*, pages 1–13.

[50] El Zaatari, S., Marei, M., Li, W., and Usman, Z. (2019). Cobot programming for collaborative industrial tasks: An overview. *Robotics and Autonomous Systems*, 116:162–180.

[51] Faibish, T., Kshirsagar, A., Hoffman, G., and Edan, Y. (2022). Human preferences for robot eye gaze in human-to-robot handovers. *International Journal of Social Robotics*, 14(4):995–1012.

[52] Ferrari, D., Pupa, A., Signoretti, A., and Secchi, C. (2023). Safe multimodal communication in human-robot collaboration. *arXiv preprint arXiv:2308.03690*.

[53] Fitzgerald, C. (2013). Developing Baxter. In *Proceedings of the 5th IEEE Conference on Technologies for Practical Robot Applications (TePRA)*, pages 1–6, Woburn MA, USA.

[54] Frank, J. A., Moorhead, M., and Kapila, V. (2016). Realizing mixed-reality environments with tablets for intuitive human-robot collaboration for object manipulation tasks. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 302–307, New York, USA.

[55] Ganesan, R. K., Rathore, Y. K., Ross, H. M., and Amor, H. B. (2018). Better teaming through visual cues: how projecting imagery in a workspace can improve human-robot collaboration. *IEEE Robotics & Automation Magazine*, 25(2):59–71.

[56] George, A. S. and George, A. H. (2023). The cobot chronicles: Evaluating the emergence, evolution, and impact of collaborative robots in next-generation manufacturing. *Partners Universal International Research Journal*, 2(2):89–116.

[57] Gervasi, R., Aliev, K., Mastrogiacomo, L., and Franceschini, F. (2022). User experience and physiological response in human-robot collaboration: A preliminary investigation. *Journal of Intelligent & Robotic Systems*, 106(2):36.

[58] Gervasi, R., Mastrogiacomo, L., and Franceschini, F. (2020). A conceptual framework to evaluate human-robot collaboration. *The International Journal of Advanced Manufacturing Technology*, 108:841–865.

[59] Gong, Z. and Zhang, Y. (2018). Behavior explanation as intention signaling in human-robot teaming. In *Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1005–1011, Nanjing, China.

[60] Grieves, M. (2014). Digital twin: manufacturing excellence through virtual factory replication. *White paper*.

[61] Gruenefeld, U., Prädel, L., Illing, J., Stratmann, T., Drolshagen, S., and Pfingsthorn, M. (2020). Mind the arm: realtime visualization of robot motion intent in head-mounted augmented reality. In *Proceedings of the Conference on Mensch und Computer (MuC)*, pages 259–266, Magdeburg, Germany.

[62] Gu, M., Cosgun, A., Chan, W. P., Drummond, T., and Croft, E. (2021). Seeing thru walls: Visualizing mobile robots in augmented reality. In *Proceedings of the 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 406–411.

[63] Guerra-Zubiaga, D., Kuts, V., Mahmood, K., Bondar, A., Nasajpour-Esfahani, N., and Otto, T. (2021). An approach to develop a digital twin for industry 4.0 systems: manufacturing automation case studies. *International Journal of Computer Integrated Manufacturing*, 34(9):933–949.

[64] Guo, J., Zhao, N., Sun, L., and Zhang, S. (2019). Modular based flexible digital twin for factory design. *Journal of Ambient Intelligence and Humanized Computing*, 10(3).

[65] Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908, Los Angeles, CA, USA.

[66] Hetherington, N. J., Croft, E. A., and Van der Loos, H. M. (2021). Hey robot, which way are you going? nonverbal motion legibility cues for human-robot spatial interaction. *IEEE Robotics and Automation Letters*, 6(3).

[67] Hietanen, A., Changizi, A., Lanz, M., Kämäräinen, J., Ganguly, P., Pieters, R., and Latokartano, J. (2019). Proof of concept of a projection-based safety system for human-robot collaborative engine assembly. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7, New Delhi, India.

[68] Hietanen, A., Pieters, R., Lanz, M., Latokartano, J., and Kämäräinen, J.-K. (2020). Ar-based interaction for human-robot collaborative manufacturing. *Robotics and Computer-Integrated Manufacturing*, 63:101891.

[69] Hoang, K. C., Chan, W. P., Lay, S., Cosgun, A., and Croft, E. (2022). Virtual barriers in augmented reality for safe and effective human-robot cooperation in manufacturing. In *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1174–1180, Naples, Italy.

[70] Hoffman, G. (2019). Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3):209–218.

[71] Horváth, G., Kardos, C., Kovács, A., Pataki, B., and Tóth, S. (2021). Worker assistance suite for efficient human–robot collaboration. In *Advanced Human-Robot Collaboration in Manufacturing*, pages 375–394. Springer.

[72] Huang, C.-M., Andrist, S., Sauppé, A., and Mutlu, B. (2015). Using gaze patterns to predict task intent in collaboration. *Frontiers in psychology*, 6:1049.

[73] Inkulu, A. K., Bahubalendruni, M. R., and Dara, A. (2022). Challenges and opportunities in human robot collaboration context of industry 4.0-a state of the art review. *Industrial Robot: the international journal of robotics research and application*, 49(2):226–239.

[74] Jones, D., Snider, C., Nassehi, A., Yon, J., and Hicks, B. (2020). Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29.

[75] Kadir, B. A. and Broberg, O. (2021). Human-centered design of work systems in the transition to industry 4.0. *Applied ergonomics*, 92:103334.

[76] Kannan, S. S., Lee, A., and Min, B.-C. (2021). External human-machine interface on delivery robots: Expression of navigation intent of the robot. In *Proceedings of the 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 1305–1312, Vancouver, Canada. IEEE.

[77] Karami, H., Carfì, A., and Mastrogiovanni, F. (2021a). Branched and/or graphs: Toward flexible and adaptable human-robot collaboration. In *Proceedings of the 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 527–533.

[78] Karami, H., Thomas, A., and Mastrogiovanni, F. (2021b). Task allocation for multi-robot task and motion planning: a case for object picking in cluttered workspaces. In *Proceedings of the 20th International Conference of the Italian Association for Artificial Intelligence (AIxIA)*, Milan, Italy.

[79] Kenett, R. S. and Bortman, J. (2022). The digital twin in industry 4.0: A wide-angle perspective. *Quality and Reliability Engineering International*, 38(3):1357–1366.

[80] Klein, G., Feltovich, P. J., Bradshaw, J. M., and Woods, D. D. (2005). Common ground and coordination in joint activity. *Organizational Simulation*, 53:139–184.

[81] Kousi, N., Gkournelos, C., Aivaliotis, S., Lotsaris, K., Bavelos, A. C., Baris, P., Michalos, G., and Makris, S. (2021). Digital twin for designing and reconfiguring human–robot collaborative assembly lines. *Applied Sciences*, 11(10).

[82] Kritzinger, W., Karner, M., Traar, G., Henjes, J., and Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11).

[83] Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.

[84] Kshirsagar, A., Lim, M., Christian, S., and Hoffman, G. (2020). Robot gaze behaviors in human-to-robot handovers. *IEEE Robotics and Automation Letters*, 5(4):6552–6558.

[85] Kumar, S., Savur, C., and Sahin, F. (2020). Survey of human–robot collaboration in industrial settings: Awareness, intelligence, and compliance. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1):280–297.

[86] Lambrecht, J. and Nimpsch, S. (2019). Human prediction for the natural instruction of handovers in human robot collaboration. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–6, New Delhi, India.

[87] Lastrico, L., Ferreira Duarte, N., Carfì, A., Rea, F., Mastrogiovanni, F., Sciutti, A., and Santos-Victor, J. (2022). If you are careful, so am i! how robot communicative motions can influence human approach in a joint task. In *Proceedings of the 14th International Conference on Social Robotics (ICSR)*, Florence, Italy.

[88] Lee, K. M., Krishna, A., Zaidi, Z., Paleja, R., Chen, L., Hedlund-Botti, E., Schrum, M., and Gombolay, M. (2023). The effect of robot skill level and communication in rapid, proximate human-robot collaboration. In *Proceedings of the 18th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 261–270, Stockholm, Sweden.

[89] Lemasurier, G., Bejerano, G., Albanese, V., Parrillo, J., Yanco, H. A., Amerson, N., Hetrick, R., and Phillips, E. (2021). Methods for expressing robot intent for human–robot collaboration in shared workspaces. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(4):1–27.

[90] Li, H., Ma, W., Wang, H., Liu, G., Wen, X., Zhang, Y., Yang, M., Luo, G., Xie, G., and Sun, C. (2022). A framework and method for human-robot cooperative safe control based on digital twin. *Advanced Engineering Informatics*, 53.

[91] Liu, C., Hamrick, J. B., Fisac, J. F., Dragan, A. D., Hedrick, J. K., Sastry, S. S., and Griffiths, T. L. (2018a). Goal inference improves objective and perceived performance in human-robot collaboration. *arXiv preprint arXiv:1802.01780.*

[92] Liu, H., Fang, T., Zhou, T., Wang, Y., and Wang, L. (2018b). Deep learning-based multimodal control interface for human-robot collaboration. *Procedia CIRP*, 72:3–8.

[93] Liu, H. and Wang, L. (2021). Latest developments of gesture recognition for human–robot collaboration. In *Advanced Human-Robot Collaboration in Manufacturing*, pages 43–68. Springer.

[94] Liu, Y., Habibnezhad, M., and Jebelli, H. (2021). Brainwave-driven human-robot collaboration in construction. *Automation in Construction*, 124:103556.

[95] Lohse, M., Rothuis, R., Gallego-Pérez, J., Karreman, D. E., and Evers, V. (2014). Robot gestures make difficult tasks easier: the impact of gestures on perceived workload and task performance. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1459–1466.

[96] Lotsaris, K., Fousekis, N., Koukas, S., Aivaliotis, S., Kousi, N., Michalos, G., and Makris, S. (2021). Augmented reality (ar) based framework for supporting human workers in flexible manufacturing. *Procedia CIRP*, 96:301–306.

[97] Luebbers, M. B., Brooks, C., Kim, M. J., Szafir, D., and Hayes, B. (2019). Augmented reality interface for constrained learning from demonstration. In *Proceedings of the 2nd International Workshop on Virtual, Augmented and Mixed Reality for HRI (VAM-HRI)*, Daegu, Korea.

[98] Luebbers, M. B., Brooks, C., Mueller, C. L., Szafir, D., and Hayes, B. (2021). Arc-lfd: Using augmented reality for interactive long-term robot skill maintenance via constrained learning from demonstration. In *Proceedings of the 38th IEEE International Conference on Robotics and Automation (ICRA)*, pages 3794–3800, Xi'an, China.

[99] Ma, L. M., IJtsma, M., Feigh, K. M., and Pritchett, A. R. (2022). Metrics for human-robot team design: A teamwork perspective on evaluation of human-robot teams. *ACM Transactions on Human-Robot Interaction (THRI)*, 11(3):1–36.

[100] Macciò, S., Carfì, A., and Mastrogiovanni, F. (2022). Mixed reality as communication medium for human-robot collaboration. In *Proceedings of the 39th IEEE International Conference on Robotics and Automation (ICRA)*, pages 2796–2802, Philadelphia PA, USA.

[101] Macciò, S., Shaaban, M., Carfì, A., Zaccaria, R., and Mastrogiovanni, F. (2023). RICO-MR: An open-source architecture for robot intent communication through mixed reality. In *Proceedings of the 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Busan, Korea.

[102] Maio, R., Santos, A., Marques, B., Almeida, D., Dias, P., and Sousa-Santos, B. (2022). Pervasive augmented reality (ar) for assistive production: Comparing the use of a head-mounted display (hmd) versus a hand-held device (hhd). In *Proceedings of the 21st International Conference on Mobile and Ubiquitous Multimedia (MUM)*, pages 279–281, Lisbon, Portugal.

[103] Makris, S. and Aivaliotis, P. (2022). Ai-based vision system for collision detection in hrc applications. *Procedia CIRP*, 106:156–161.

[104] Malik, A. A. and Bilberg, A. (2018). Digital twins of human robot collaboration in a production setting. *Procedia manufacturing*, 17.

[105] Malik, A. A. and Brem, A. (2021). Digital twins for collaborative robots: A case study in human-robot interaction. *Robotics and Computer-Integrated Manufacturing*, 68.

[106] Marvel, J. A., Bagchi, S., Zimmerman, M., and Antonishek, B. (2020). Towards effective interface designs for collaborative hri in manufacturing: Metrics and measures. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(4):1–55.

[107] Matheson, E., Minto, R., Zampieri, E. G., Faccio, M., and Rosati, G. (2019). Human–robot collaboration in manufacturing applications: a review. *Robotics*, 8(4):100.

[108] Matsumoto, S., Washburn, A., and Riek, L. D. (2022). A framework to explore proximate human-robot coordination. *ACM Transactions on Human-Robot Interaction (THRI)*, 11(3):1–34.

[109] Meissner, A., Trübswetter, A., Conti-Kufner, A. S., and Schmidtler, J. (2020). Friend or foe? understanding assembly workers' acceptance of human-robot collaboration. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(1):1–30.

[110] Michalos, G., Karagiannis, P., Makris, S., Tokçalar, Ö., and Chryssolouris, G. (2016). Augmented reality (ar) applications for supporting human-robot interactive cooperation. *Procedia CIRP*, 41:370–375.

[111] Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(12):1321–1329.

[112] Mohammadi Amin, F., Rezayati, M., van de Venn, H. W., and Karimpour, H. (2020). A mixed-perception approach for safe human–robot collaboration in industrial automation. *Sensors*, 20(21):6347.

[113] Mourtzis, D., Angelopoulos, J., and Panopoulos, N. (2022). Closed-loop robotic arm manipulation based on mixed reality. *Applied Sciences*, 12(6).

[114] Moutinho, D., Rocha, L. F., Costa, C. M., Teixeira, L. F., and Veiga, G. (2023). Deep learning-based human action recognition to leverage context awareness in collaborative assembly. *Robotics and Computer-Integrated Manufacturing*, 80:102449.

[115] Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. (2009). Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 69–76, La Jolla CA, USA.

[116] Neto, P., Simão, M., Mendes, N., and Safeea, M. (2019). Gesture-based human-robot interaction for human assistance in manufacturing. *The International Journal of Advanced Manufacturing Technology*, 101(1):119–135.

[117] Newbury, R., Cosgun, A., Crowley-Davis, T., Chan, W. P., Drummond, T., and Croft, E. A. (2022). Visualizing robot intent for object handovers with augmented reality. In *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1264–1270, Naples, Italy.

[118] Ng, A. et al. (2022). Challenges for manufacturing smes in the introduction of collaborative robots. In *Proceedings of the 10th Swedish Production Symposium (SPS2022)*, Skove, Sweden.

[119] Nikolaidis, S., Kwon, M., Forlizzi, J., and Srinivasa, S. (2018). Planning with verbal communication for human-robot collaboration. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(3):1–21.

[120] Ostanin, M., Mikhel, S., Evlampiev, A., Skvortsova, V., and Klimchik, A. (2020). Human-robot interaction for robotic manipulator programming in mixed reality. In *Proceedings of the 37th International Conference on Robotics and Automation (ICRA)*, pages 2805–2811, Paris, France.

[121] Pages, J., Marchionni, L., and Ferro, F. (2016). Tiago: the modular robot that adapts to different research needs. In *International workshop on robot modularity, IROS*, volume 290.

[122] Palmarini, R., del Amo, I. F., Bertolino, G., Dini, G., Erkoyuncu, J. A., Roy, R., and Farnsworth, M. (2018). Designing an ar interface to improve trust in human-robots collaboration. *Procedia CIRP*, 70:350–355.

[123] Paluch, S., Tuzovic, S., Holz, H. F., Kies, A., and Jörling, M. (2022). "my colleague is a robot"–exploring frontline employees' willingness to work with collaborative service robots. *Journal of Service Management*, 33(2):363–388.

[124] Papageorgiou, D., Stavridis, S., Papakonstantinou, C., and Doulgeri, Z. (2021). Task geometry aware assistance for kinesthetic teaching of redundant robots. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7285–7291.

[125] Papcun, P., Cabadaj, J., Kajati, E., Romero, D., Landryova, L., Vascak, J., and Zolotova, I. (2019). Augmented reality for humans-robots interaction in dynamic slotting "chaotic storage" smart warehouses. In *IFIP International Conference on Advances in Production Management Systems (AMPS)*, pages 633–641, Austin, TX, USA.

[126] Pinto, A. R., Kildal, J., and Lazkano, E. (2020). Multimodal mixed reality impact on a hand guiding task with a holographic cobot. *Multimodal Technologies and Interaction*, 4(4):78.

[127] Puljiz, D., Stöhr, E., Riesterer, K. S., Hein, B., and Kröger, T. (2019). General hand guidance framework using microsoft hololens. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5185–5190, Macau, China.

[128] Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., Ng, A. Y., et al. (2009). ROS: an open-source robot operating system. In *ICRA workshop on open source software*, Kobe, Japan.

[129] Quintero, C. P., Li, S., Pan, M. K., Chan, W. P., Van der Loos, H. M., and Croft, E. (2018). Robot programming through augmented trajectories in augmented reality. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1838–1844, Madrid, Spain.

[130] Ragaglia, M., Zanchettin, A. M., and Rocco, P. (2018). Trajectory generation algorithm for safe human-robot collaboration based on multiple depth sensor measurements. *Mechatronics*, 55.

[131] Rahmayanti, R., Alvarez, J. C., Alvarez, D., and López, A. M. (2022). Model-based design of a collaborative human-robot workspace. In *Proceedings of the 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1291–1296, Naples, Italy.

[132] Ravichandar, H., Polydoros, A. S., Chernova, S., and Billard, A. (2020). Recent advances in robot learning from demonstration. *Annual review of control, robotics, and autonomous systems*, 3:297–330.

[133] Reardon, C., Lee, K., Rogers, J. G., and Fink, J. (2019). Augmented reality for human-robot teaming in field environments. In *Proceedings of the 11th International Conference on Virtual, Augmented and Mixed Reality (VAMR)*, pages 79–92, Orlando, FL, USA.

[134] Ren, T., Dong, Y., Wu, D., and Chen, K. (2018). Collision detection and identification for robot manipulators based on extended state observer. *Control Engineering Practice*, 79.

[135] Rivera-Pinto, A., Kildal, J., and Lazkano, E. (2023). Toward programming a collaborative robot by interacting with its digital twin in a mixed reality environment. *International Journal of Human–Computer Interaction*, pages 1–13.

[136] Rosen, E., Whitney, D., Fishman, M., Ullman, D., and Tellex, S. (2020). Mixed reality as a bidirectional communication interface for human-robot interaction. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11431–11438, Las Vegas, USA.

[137] Rosen, E., Whitney, D., Phillips, E., Chien, G., Tompkin, J., Konidaris, G., and Tellex, S. (2019). Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays. *The International Journal of Robotics Research*, 38(12-13):1513–1526.

[138] Roveda, L., Maskani, J., Franceschi, P., Abdi, A., Braghin, F., Molinari Tosatti, L., and Pedrocchi, N. (2020). Model-based reinforcement learning variable impedance control for human-robot collaboration. *Journal of Intelligent & Robotic Systems*, 100(2).

[139] Ruffaldi, E., Brizzi, F., Tecchia, F., and Bacinelli, S. (2016). Third point of view augmented reality for robot intentions visualization. In *Proceedings of the 3rd International Conference on Augmented Reality, Virtual Reality and Computer Graphics (AVR)*, pages 471–478, Otranto, Italy.

[140] Sahin, M. and Savur, C. (2022). Evaluation of human perceived safety during hrc task using multiple data collection methods. In *Proceedings of the 17th Annual System of Systems Engineering Conference (SOSE)*, pages 465–470. IEEE.

[141] Salehzadeh, R., Gong, J., and Jalili, N. (2022). Purposeful communication in human-robot collaboration: A review of modern approaches in manufacturing. *IEEE Access*.

[142] Sandini, G., Sciutti, A., Rea, F., Goswami, A., and Vadakkepat, P. (2017). Movement-based communication for humanoid-human interaction. *Humanoid Robotics: A Reference*, pages 1–29.

[143] Scheffe, H. (1999). *The analysis of variance*, volume 72. John Wiley & Sons.

[144] Schrepp, M., Thomaschewski, J., and Hinderks, A. (2017). Construction of a benchmark for the user experience questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(4):40–44.

[145] Shaaban, M., Carfì, A., and Mastrogiovanni, F. (2023). Digital twins for human-robot collaboration: A future perspective. *arXiv preprint arXiv:2311.02421*.

[146] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

[147] Sheikholeslami, S., Lee, G., Hart, J. W., Srinivasa, S., and Croft, E. A. (2018). A study of reaching motions for collaborative human-robot interaction. In *Proceedings of the International Symposium on Experimental Robotics*, pages 584–594, Buenos Aires, Argentina.

[148] Shrestha, M. C., Kobayashi, A., Onishi, T., Uno, E., Yanagawa, H., Yokoyama, Y., Kamezaki, M., Schmitz, A., and Sugano, S. (2016). Intent communication in navigation through the use of light and screen indicators. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Christchurch, New Zealand.

[149] Shukla, D., Erkent, Ö., and Piater, J. (2018). Learning semantics of gestural instructions for human-robot collaboration. *Frontiers in neurorobotics*, 12:7.

[150] Simões, A. C., Pinto, A., Santos, J., Pinheiro, S., and Romero, D. (2022). Designing human-robot collaboration (hrc) workspaces in industrial settings: A systematic literature review. *Journal of Manufacturing Systems*, 62:28–43.

[151] Sonawani, S. and Amor, H. (2022). When and where are you going? a mixed-reality framework for human robot collaboration. In *Proceedings of the 5th International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, Sapporo, Japan.

[152] Strabala, K., Lee, M. K., Dragan, A., Forlizzi, J., and Srinivasa, S. S. (2012). Learning the communication of intent prior to physical collaboration. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 968–973, Paris, France.

[153] Sun, X., Zhang, R., Liu, S., Lv, Q., Bao, J., and Li, J. (2022). A digital twin-driven human–robot collaborative assembly-commissioning method for complex products. *The International Journal of Advanced Manufacturing Technology*, 118(9).

[154] Szafir, D. (2019). Mediating human-robot interactions with virtual, augmented, and mixed reality. In *Proceedings of the 11th International Conference on Virtual, Augmented and Mixed Reality (VAMR)*, pages 124–149, Orlando, FL, USA.

[155] Szafir, D., Mutlu, B., and Fong, T. (2015). Communicating directionality in flying robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 19–26, Portland OR, USA.

[156] Tang, G., Webb, P., and Thrower, J. (2019). The development and evaluation of robot light skin: A novel robot signalling system to improve communication in industrial human–robot collaboration. *Robotics and Computer-Integrated Manufacturing*, 56:85–94.

[157] Terzioğlu, Y., Mutlu, B., and Şahin, E. (2020). Designing social cues for collaborative robots: the role of gaze and breathing in human-robot collaboration. In *Proceedings of the 15th ACM/IEEE international conference on human-robot interaction (HRI)*, pages 343–357.

[158] Tsamis, G., Chantziaras, G., Giakoumis, D., Kostavelis, I., Kargakos, A., Tsakiris, A., and Tzovaras, D. (2021). Intuitive and safe interaction in multi-user human robot collaboration environments through augmented reality displays. In *Proceedings of the 30th IEEE international conference on robot & human interactive communication (RO-MAN)*, pages 520–526.

[159] Tsiakas, K., Papakostas, M., Theofanidis, M., Bell, M., Mihalcea, R., Wang, S., Burzo, M., and Makedon, F. (2017). An interactive multisensing framework for personalized human robot collaboration and assistive training using reinforcement learning. In *Proceedings of the 10th international conference on pervasive technologies related to assistive environments*, pages 423–427.

[160] Tuli, T. B., Kohl, L., Chala, S. A., Manns, M., and Ansari, F. (2021). Knowledge-based digital twin for predicting interactions in human-robot collaboration. In *Proceedings of the 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Västerås, Sweden.

[161] Unhelkar, V. V., Li, S., and Shah, J. A. (2020). Decision-making for bidirectional communication in sequential human-robot collaborative tasks. In *Proceedings of the 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 329–341.

[162] van Delden, S., Umrysh, M., Rosario, C., and Hess, G. (2012). Pick-and-place application development using voice and visual commands. *Industrial Robot: An International Journal*, 39(6):592–600.

[163] Villani, V., Pini, F., Leali, F., and Secchi, C. (2018). Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. *Mechatronics*, 55:248–266.

[164] Vogel, C., Schulenburg, E., and Elkmann, N. (2020). Projective-ar assistance system for shared human-robot workplaces in industrial applications. In *Proceedings of the 25th IEEE International conference on emerging technologies and factory automation (ETFA)*, volume 1, pages 1259–1262.

[165] Vogel, C., Walter, C., and Elkmann, N. (2017). Safeguarding and supporting future human-robot cooperative manufacturing processes by a projection-and camera-based technology. *Procedia Manufacturing*, 11:39–46.

[166] Walker, M., Hedayati, H., Lee, J., and Szafir, D. (2018). Communicating robot motion intent with augmented reality. In *Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 316–324, Chicago, IL, USA.

[167] Wang, C. and Belardinelli, A. (2022). Investigating explainable human-robot interaction with augmented reality. In *Proceedings of the 5th International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*.

[168] Wang, X. V., Wang, L., Lei, M., and Zhao, Y. (2020). Closed-loop augmented reality towards accurate human-robot collaboration. *CIRP Annals*, 69(1):425–428.

[169] Weidemann, A. and Rußwinkel, N. (2021). The role of frustration in human–robot interaction–what is needed for a successful collaboration? *Frontiers in psychology*, 12:707.

[170] Wengefeld, T., Höchemer, D., Lewandowski, B., Köhler, M., Beer, M., and Gross, H.-M. (2020). A laser projection system for robot intention communication and human robot interaction. In *Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 259–265, Naples, Italy.

[171] Wilcoxon, F. (1947). Probability tables for individual comparisons by ranking methods. *Biometrics*, 3(3):119–122.

[172] Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in Statistics*, pages 196–202. Springer.

[173] Wilhelm, J., Petzoldt, C., Beinke, T., and Freitag, M. (2021). Review of digital twin-based interaction in smart manufacturing: Enabling cyber-physical systems for human-machine interaction. *International Journal of Computer Integrated Manufacturing*, 34(10).

[174] Williams, T., Bussing, M., Cabrol, S., Boyle, E., and Tran, N. (2019). Mixed reality deictic gesture for multi-modal robot communication. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 191–201, Daegu, Korea.

[175] Ye, S., Neville, G., Schrum, M., Gombolay, M., Chernova, S., and Howard, A. (2019). Human trust after robot mistakes: Study of the effects of different forms of robot communication. In *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1–7, New Delhi, India.

[176] Zhang, J., Liu, H., Chang, Q., Wang, L., and Gao, R. X. (2020). Recurrent neural network for motion trajectory prediction in human-robot collaborative assembly. *CIRP annals*, 69(1):9–12.

[177] Zhu, H., Gabler, V., and Wollherr, D. (2017). Legible action selection in human-robot collaboration. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 354–359. IEEE.

[178] Zou, C., Chandan, K., Ding, Y., and Zhang, S. (2023). Ardie: Ar, dialogue, and eye gaze policies for human-robot collaboration. *arXiv preprint arXiv:2305.04685*.