Kinesthetic Teaching in Robotics: a Mixed Reality Approach

Simone Macciò
Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa
Via Opera Pia 13, 16145 Genoa, Italy
simone.maccio@edu.unige.it

Mohamad Shaaban
Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa
Via Opera Pia 13, 16145 Genoa, Italy
mohamad.shaaban@edu.unige.it

Alessandro Carfì
Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa
Via Opera Pia 13, 16145 Genoa, Italy
alessandro.carfi@unige.it

Fulvio Mastrogiovanni
Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa
Via Opera Pia 13, 16145 Genoa, Italy
fulvio.mastrogivoanni@edu.unige.it

Abstract: As collaborative robots become more common in manufacturing scenarios and adopted in hybrid human-robot teams, we should develop new interaction and communication strategies to ensure smooth collaboration between agents. In this paper, we propose a novel communicative interface that uses Mixed Reality as a medium to perform Kinesthetic Teaching (KT) on any robotic platform. We evaluate our proposed approach in a user study involving multiple subjects and two different robots, comparing traditional physical KT with holographic-based KT through user experience questionnaires and task-related metrics.

Keywords: Human-Robot Interaction, Mixed Reality, Kinesthetic Teaching

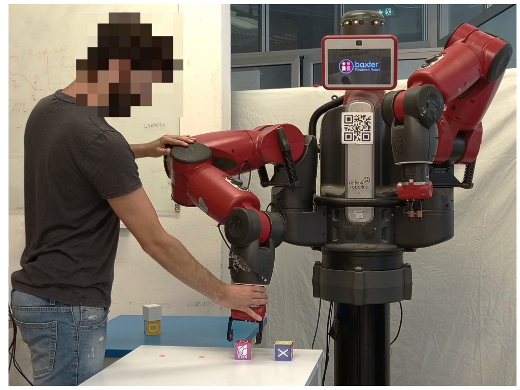
1 Introduction

In smart factories, robots are designed to coexist and collaborate with humans rather than replace them. This manufacturing shift has led to the rise of collaborative robots, which are adaptive and versatile platforms [1]. One critical aspect of Human-Robot Collaboration (HRC) is developing structured communication systems that allow agents to exchange information [2] intuitively. Research has shown that effective bi-directional communication is vital for successful collaboration, enabling agents to understand each other's actions, synchronize, and provide feedback. Conversely, poor communication can lead to misunderstandings and distrust in robot teammates [3].

Creating a robust communication interface is complex, requiring the selection of appropriate communicative channels. A promising approach combines Mixed Reality (MR) with wearable Head-Mounted Displays (HMD), allowing for engaging holographic interfaces where users see 3D digital content overlaid onto their environment [4]. This virtual layer can serve as a communicative channel for intuitive human-robot interaction. However, few studies have explored using MR to preview

8th Conference on Robot Learning (CoRL 2024), Munich, Germany.

(a) An experimenter conducts a holographic KT session with the Tiago++ robot, teaching actions by manipulating the grey holographic sphere, superimposed on the robot's wrist.



(b) An experimenter interacting with Baxter during a physical KT session. The operator drives the robot's arm through gestural interaction, teaching the needed sequence of pick-and-place actions.

Figure 1: Demonstrations of KT sessions with the Tiago++ (left) and Baxter (right) robots, highlighting both holographic and physical interaction modalities.

a robot's intentions and actions [5, 6, 7], which could provide valuable visual feedback to human teammates during collaboration.

In this paper, we explore human-to-robot communication using Mixed Reality (MR) to enable operators to *teach* robots through holographic communication. We focus on the Learning from Demonstration (LfD) approach [8], positing that LfD sessions serve as communication acts to transfer skills from human operators to robot teammates via explicit actions or gestures. Our work targets Kinesthetic Teaching (KT), a well-established technique where human operators manually guide the robot's arm or end-effector to teach new actions through direct demonstration. We aim to integrate this teaching methodology into the holographic communicative framework proposed in [7]. Throughout the paper, we formalize KT within this framework and translate it into a modular software component, enabling KT in human-robot interactions via holographic communication. Our approach leverages MR for intuitive interaction and aligns with the LfD paradigm, offering a holographic tool for demonstrating skills to robot teammates in Human-Robot Collaboration (HRC). Furthermore, the MR space's flexibility allows for KT on any robotic platform compatible with the Universal Robot Description Format (URDF).

In addition to presenting a holographic-based tool for KT, we assess its effectiveness in task demonstration and user experience (UX). We propose that the holographic KT approach can be a viable alternative to traditional, hand-guided KT when the latter is unavailable or not applicable to a specific robot platform. To test this hypothesis, we conducted a preliminary user study with 12 participants and two robots, comparing the two KT methods using task-based metrics and UX questionnaires.

2 Background

Over the years, various communication strategies have been explored in Human-Robot Collaboration (HRC), incorporating both explicit media (e.g., voice [9], upper limb gestures [10, 11], light and visual cues [12, 13]) and implicit ones (e.g., gaze [14], body posture and motion [15]). However, many of these approaches have inherent limitations, hindering the development of a bi-directional communication interface and restricting their applicability to a narrow range of collaborative scenarios. For example, while human-like communication through gestures and gaze can be expressive and intuitive, most collaborative platforms lack the physical features to replicate these cues.

The advent of Augmented Reality (AR) on mobile devices such as smartphones and tablets has introduced a new virtual layer for researchers to facilitate intuitive communication between human and robot teammates [16, 17, 18]. This has become increasingly relevant with the rise of Mixed Reality (MR) Head-Mounted Displays (HMDs), which enhance immersion and enable the development of interfaces for programming robot behaviors [19, 20, 21] and provide intuitive feedback during interactions [22, 23]. Researchers have also focused on conveying a robot's intentions through MR, exploring intuitive strategies to enable robots to anticipate actions via holographic cues during interactive tasks [5, 6, 7].

While substantial research has examined how robots can communicate effectively with humans through MR, few studies have explored leveraging this medium for intuitive human-to-robot communication, particularly in Learning from Demonstration (LfD). Popular LfD methods often rely on computer vision to passively observe and transfer desired motions from human actions [24, 25] or use hand-tracking devices for teleoperation-based LfD [26]. Although these methods offer a straightforward interface for skill transfer, they typically require structured environments and complex calibration routines, which may limit real-world applications. In contrast, MR can address these challenges, as MR-HMDs are designed for unstructured environments and can provide similar demonstration capabilities with minimal setup.

In the realm of Kinesthetic Teaching (KT), early attempts at integrating KT and MR relied on physical robot guidance for demonstrations while using holographic media primarily for visualizing learned actions and imposing motion constraints [27, 28]. An example is [29], where authors utilize MR-HMD hand-tracking capabilities to manually drive an industrial robotic manipulator's joints during teaching. Similarly, [30] introduced a system for teaching a tabletop holographic robot a simple pick-and-place task via holographic hand guidance. A more recent study [31] proposed an MR interface for intuitively teaching trajectories to a holographic collaborative manipulator. However, these works lack a structured representation of communication acts for skill transfer and empirical assessments of the demonstration capabilities and user experiences of these solutions.

In contrast, this article aims to frame KT consistently within the holographic communication space outlined in [7], presenting a standalone approach for MR-based KT applicable to any robot describable in the URDF format. Additionally, we conduct an experimental evaluation of the communicative capabilities of our MR-based KT tool, assessing the robot skills acquired during interactive tasks. The proposed framework adheres to the open-source paradigm, making it publicly available for researchers and companies to use as an alternative to traditional KT with any URDF-compatible robot, requiring minimal hardware setup. [1].

3 Formalization

Recalling the definition provided in [7], we describe communication as the act of conveying or transmitting *pieces of information* (I) through one or more communicative channels. It is noteworthy to mention that, in general, conveying a single piece of information may involve simultaneously multiple channels to strengthen the clarity of the communicative act itself. For example, human-human communication often combines verbal and gestural media to be meaningful and unambiguous. Following this principle, and denoting $M = \{m_1, \ldots, m_{|M|}\}$ the set of all possible communicative media available (e.g., voice, gestures, gaze and so on), we provided the general formulation of a communicative act involving N communication media, namely

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

where $\boldsymbol{t}$ represents the time interval associated with the overall communication, whereas the intervals $\boldsymbol{t}_i$ span the duration of the individual components of the communication act.

Here, we leverage such formalization to frame KT inside the holographic communication space developed for [7]. The first step requires identifying the relevant information exchanged during KT

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

sessions. In particular, we argue that the act of KT implies teaching robots about their future *states*, denoted as $\boldsymbol{\tau}$. Without loss of generality, such a notion of robot state includes the robot's pose $\boldsymbol{x}(t)$ (that is, its position and orientation in the environment) and its joint configuration $\boldsymbol{q}(t)$,where $\boldsymbol{q}(t)$ also includes the state of the gripper. Consequently, we can formalize the robot's state as

$$\boldsymbol{\tau}(t) = \{\boldsymbol{x}(t), \boldsymbol{q}(t)\} \ . \tag{2}$$

This, in turn, provides us with a suitable representation of the set of information I which can be conveyed through KT, namely $I = \{\boldsymbol{\tau}(t)\}$. Having defined the set I, we observe that KT is achieved by hand-guiding the robot's wrist or end-effector. According to our proposed formalism, this act involves a gesture-mediated communication C_{gest} that enables users to teach robots about their future states in a simple way and can be described as follows:

$$C_{\text{gest}}(I, \boldsymbol{t}_{\text{gest}}) = \mathbf{T}(\boldsymbol{t}_{\text{gest}}) \,, \tag{3}$$

where $\mathbf{T}(\boldsymbol{t}_{\text{gest}})$ describes the robot trajectory that is conveyed via gestural guidance during the interval $\boldsymbol{t}_{\text{gest}}$ spanning the KT session

With this formalization in mind, we claim that KT can be translated and framed into the holographic communication space envisioned in [7] by letting users convey robots' trajectories via gestural guidance on a virtual counterpart of the robot. As already mentioned, the unconstrained nature of the MR space allows for such a form of KT while solely relying on the built-in hand-tracking capabilities of the MR-HMD device. Additionally, such decoupling between physical and holographic layers could be particularly effective in production environments, as the operators could leverage the virtual robot to program or teach upcoming tasks, without halting the execution of real robotic chains.

To further strengthen the communicative framework and ensure a more natural interaction, we postulate that adding the vocal medium 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 modelling, the holographic-based KT process is translated into a communication act combining gestural and vocal interaction and, as such, can be formalized as follows:

$$C^{KT}(I, \boldsymbol{t}) = C_{\text{gest}}(I, \boldsymbol{t}_{\text{gest}}) \ \cup \ C_{\text{voc}}(I, \boldsymbol{t}_{\text{voc}}) \,. \tag{4}$$

This formalization, combined with equation (3), describes the building blocks of the communication act taking place during the proposed holographic-based KT process. In the following paragraph, these building blocks are translated into modular software components and integrated into a preexisting MR-based architecture.

4 Software Architecture

The software components developed in the context of this work constitute a modular extension of the open-source architecture, named *Robot Intent Communication through Mixed Reality* (RICO-MR), which is introduced and detailed in [32]. The features described in this paragraph are publicly available under MIT licence in a separate branch of the main RICO-MR repository. A link to the repository is included at the end of Section 2.

The proposed architecture exploits functionalities developed for RICO-MR to achieve the holographic KT envisioned in Section 3. However, currently, the architecture allows holographic KT with fixed manipulators only. As such, we introduce a simplification in the formalization provided in (2), and we hereafter refer to the notion of robot state to indicate its joint configuration $\boldsymbol{q}(t)$ only.

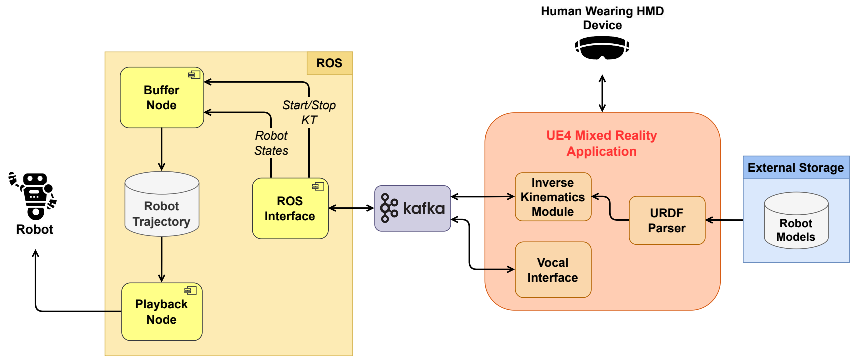


Figure 2: Overview of the proposed architecture implementing holographic KT.

4.1 Mixed Reality Application

A *MR Application*, built with Unreal Engine 4.27 (UE4) and deployed on the embedded HMD device worn by the user, drives the whole holographic interface. A hand-attached menu enables the user to select robot models from a list of predefined ones, making it possible to load and spawn holographic robots in the environment. Aside from the pre-loaded models that ship with the current architecture version, the list of supported robots can be extended by uploading relevant resources (i.e., URDF files) to a remote repository, which can be customized in the application's settings. 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 a QR code as a spatial anchor, taking advantage of Unreal's marker detection capabilities. Along with the robot model, a grey holographic sphere, visible in Fig. 1a, is spawned and superimposed 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 human 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). Specifically, the IK computation occurs with a rate of 30 Hz. 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 3, a voice interface is also active inside the MR 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.

4.2 Recording and Playback

While the MR application provides the holographic interface to perform KT, recording and subsequent playback of the robot's actions are respectively managed through Apache Kafka and the Robot Operating System (ROS) [33] framework. On the one hand, we take advantage of Kafka, an open-source, high-performant data streaming platform, for input / output data exchange with the MR application. Kakfa provides numerous advantages for real-time data streaming applications, including cloud integration and scalability, and it has been adopted for developing RICO-MR [32]. In this context, we use Kafka to stream the robot's states at a rate of 20 Hz, 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 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, a *ROS-Kafka Interface* has been developed 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.

5 Experimental Validation

5.1 Hypotheses and Experimental Scenario

The experimental campaign carried out in this study aims to determine if our proposed holographic KT approach can act as a suitable alternative to standard, physical KT, both in terms of demonstration capabilities and perceived user experience. To achieve our goal, we devised a human-robot interactive scenario to compare traditional physical kinematic teaching (KT), where the operator manually controls the robot's kinematic chain, with our proposed holographic approach. To ensure more generalized results, we conducted experiments using two different robots. In particular, we opted for *Baxter* [34] from Rethink Robotics and *Tiago*++ [35] from Pal Robotics, both being well-known platforms adopted in relevant research studies [5, 7, 11, 23, 36] and natively endowed with the necessary software and hardware components to achieve physical KT. Similarly, the HMD platform employed for rendering the holographic medium is a Microsoft HoloLens 2, a popular MR headset offering many features, including 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, we have come up with the following hypotheses, which have been evaluated through preliminary user study:

- *H1* There is no observable difference between actions taught through physical or holographic KT, namely the two approaches provide equivalent communicative power, leading to similar playback outcomes;
- *H2* No difference can be observed in terms of temporal overhead when demonstrating actions through either physical or holographic KT;
- *H3* No difference can be observed between the two approaches in terms of perceived UX during the demonstration process.

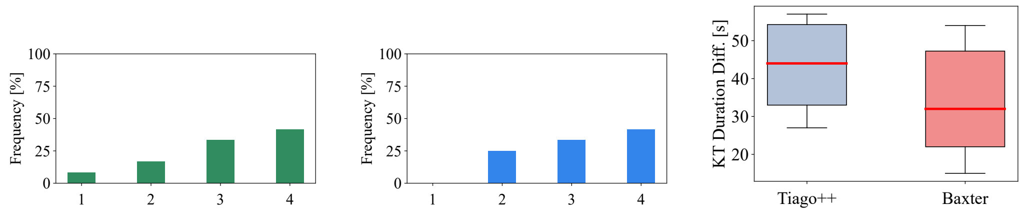
Regarding the interactive task employed to evaluate the two KT alternatives, a simple *stacking task* has been devised. Specifically, the human should use KT to teach a sequence of pick-and-place actions aimed at stacking four cubes on top of each other according to a predefined order. Fig. 1b depicts the experimental scenario, showing a user in the middle of a physical KT session with the Baxter robot.

5.2 User Study

We carried out a within-subject experimental campaign with $K = 12$ volunteers (9 males and 3 females), all aged between 21-32 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 physical, hand-guided KT.
- *C2* Wearing the HMD and performing holographic KT.

To avoid introducing unwanted biases, the starting experimental condition for each subject was randomized. Participants were initially instructed on the stacking task and assigned an arbitrary order for the cubes to be collected. Then, they performed their first trial, in condition *C1* or *C2*. However, before beginning the experiment with HMD on (i.e., condition *C2*), subjects were also briefly instructed on how to interact with the HoloLens holographic menus and interface. Then, once accustomed, they proceeded to carry out their trial. Subsequently, each subject repeated the experiment



(a) Cubes stacked in *C1*. (b) Cubes stacked in *C2*. (c) Temporal overhead.

Figure 3: (Left, Middle) Histograms depicting the number of cubes successfully stacked during the playback phase in conditions *C1* and *C2*. (Right) Differential distributions of the temporal overhead introduced under condition *C2*.

in the opposite condition. To achieve a consistent KT experience, the holographic interface in condition *C2* also included four virtual cubes placed coherently with their real-world counterparts, as shown in Fig. 1a. 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 MR application running on the HoloLens 2, in condition *C1* it was simulated thanks to a *Wizard of Oz* approach.

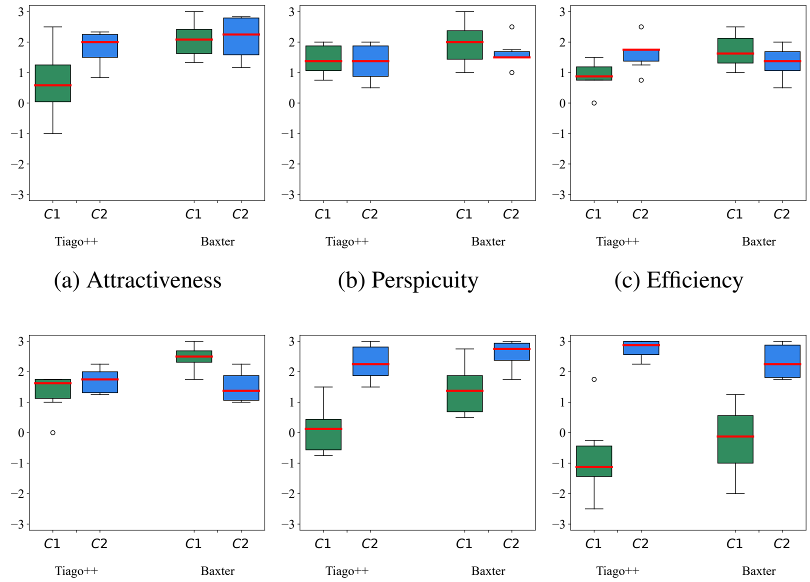
After successfully completing each KT session, the playback phase was manually triggered, causing the robot to reproduce the taught action. This phase allowed us to rank the KT session quantitatively by combining two distinct variables, useful in evaluating *H1* and *H2*. On the one hand, we counted the number of cubes successfully stacked by the robot during playback. As such, we were able to evaluate the communicative capabilities of each KT alternative, assessing how well the combination of vocal and gestural interface translated into the corresponding robot action. On the other hand, we recorded the duration of each demonstration session and employed such quantity to compare the two KT techniques in terms of time necessary to teach the full stacking task.

Finally, after completing their trials, each participant was required to fill out the User Experience Questionnaire (UEQ) [37], a well-known survey useful for ranking and comparing interactive products. In particular, such a questionnaire allows grading the UX of a given product through six evaluation scales, namely *attractiveness*, *perspicuity*, *efficiency*, *dependability*, *stimulation* and *novelty*. In accordance with hypothesis *H3*, 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.

6 Results

We report and discuss the results of our preliminary user study. Notably, we found that participants achieved comparable outcomes in teaching the stacking task across both experimental conditions, regardless of the robot used. Thus, Fig. 3 presents only the aggregated results, comparing conditions *C1* and *C2* without distinguishing between interactions with Tiago++ or Baxter. The histograms indicate that about 40% of subjects successfully executed a flawless KT, enabling the robot to stack all four cubes during playback.

From the plots in Fig. 3, we observe that physical and holographic KT produced similar results. Since the distributions could not be assumed normal, we employed a non-parametric one-tailed Wilcoxon signed-rank test [38]. The test yielded a statistic of $W = 20$ with a p-value greater than 0.3. We compared this result against the critical value $W_c = 17$, as found in the literature [39], at a significance level of $\alpha = 0.05$. Since $W > W_c$, we could not reject the null hypothesis, suggesting our initial hypothesis *H1* was correct: both communicative interfaces (physical and holographic) yield consistent performance in executing KT.



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

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

Figure 4: 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.

Regarding the time needed for KT, participants in condition *C2* were slower due to their limited experience with MR devices. We conducted a differential analysis, calculating the time difference between conditions *C2* and *C1* for each participant. The results, shown in Fig. 3c, indicate that, on average, holographic KT took 44 seconds longer for Tiago++ and 32 seconds longer for Baxter compared to physical sessions. This results in mean temporal overheads of 37% and 33%, respectively. A one-tailed t-test on the original distributions confirmed these findings, yielding p-values less than 0.05, allowing us to reject the null hypothesis for *H2*. While these results imply that the holographic demonstration process is slower, we believe the participants' limited experience with MR devices contributed significantly to the increased teaching time. Future studies could involve a more experienced population to verify or revisit these findings.

Nonetheless, Fig. 3c shows no significant difference between temporal overheads when using one robot or the other. This result is also confirmed by a one-tailed t-test on the two differential distributions, which yielded a p-value> 0.2. In other words, the overhead introduced by the MR medium was consistent among the two robots.

Fig. 4 displays results from the UEQ questionnaires, categorized by evaluation scale and robot type, with scores ranging from $[-3, 3]$, where positive values reflect favourable user perceptions of each interface. Figs. 4c and 4b demonstrate that both KT approaches yielded similar results in terms of *efficiency* and *perspicuity* (intuitive and pragmatic interface perception), regardless of the robot employed. Statistical analysis via the Kruskal-Wallis test [40] returned p-values greater than 0.05 for both scales, indicating no significant differences. This result aligns with our hypothesis *H3*, suggesting similar perceived user experiences across both KT strategies. Notably, holographic KT excelled in *attractiveness*, *stimulation*, and *novelty*, indicating that participants found the holographic environment more engaging and innovative. The only scale where holographic KT performed slightly worse was *dependability*, which assesses perceived safety and predictability. In this case, physical KT was viewed as more predictable, especially with Baxter, although the MR-based approach still received positive ratings with both robots.

7 Conclusions

In this paper, we proposed a novel communicative interface based on MR to achieve KT with any URDF-compatible robotic manipulator platform. We built on top of a state-of-the-art communicative framework [7] to account for holographic-based KT as a form of human-to-robot communication. Then, we presented a software architecture translating the formalization into a practical MR application running on embedded HMD devices. We compared holographic KT with standard, physical KT in a preliminary user study involving multiple subjects and two different robots. The results suggest that holographic KT behaves comparably to physical KT, achieving similar task-based performances and user experience. This finding suggests that the proposed methodology could be adopted as a suitable alternative to physical KT in experimental and manufacturing scenarios, decoupling the demonstration process and enabling operators to program robot tasks in the MR space, without halting the production flow of the machine. In future works, we will evaluate whether these findings can be generalized by conducting user studies on a wider population, considering different robots, and more structured human-robot interaction scenarios where the individual is required to teach more complex tasks through holographic KT.

Acknowledgments

This work has been supported by the European Union Erasmus+ Programme via the Joint Master Degree program European Master on Advanced Robotics Plus (EMARO+), and via the Italian government support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2 Investment 1.5, funded from the European Union NextGenerationEU and awarded by the Italian Ministry of University and Research.

References

[1] L. Wang, S. Liu, H. Liu, and X. V. Wang. Overview of human-robot collaboration in manufacturing. In *Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing (AMP 2020)*, pages 15–58, Belgrade, Serbia, June 2020.

[2] R. Suzuki, A. Karim, T. Xia, H. Hedayati, and N. Marquardt. Augmented reality and robotics: A survey and taxonomy for ar-enhanced human-robot interaction and robotic interfaces. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 1–33, New Orleans, USA, Apr. 2022.

[3] S. Ye, G. Neville, M. Schrum, M. Gombolay, S. Chernova, and A. Howard. 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, Oct. 2019.

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

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

[6] R. Newbury, A. Cosgun, T. Crowley-Davis, W. P. Chan, T. Drummond, and E. A. Croft. 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, Aug. 2022.

[7] S. Macciò, A. Carfì, and F. Mastrogiovanni. 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, May 2022.

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

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

[10] A. Carfì, C. Motolese, B. Bruno, and F. Mastrogiovanni. 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, August 2018.

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

[12] E. Cha and M. Matarić. Using nonverbal signals to request help during human-robot collaboration. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5070–5076, Daejeon, Korea, Otc. 2016.

[13] S. Song and S. Yamada. Bioluminescence-inspired human-robot interaction: designing expressive lights that affect human's willingness to interact with a robot. In *Proceedings of 13th ACM/IEEE International Conference on Human Robot Interaction (HRI)*, pages 224–232, Chicago, USA, Mar. 2018.

[14] A. Kalegina, G. Schroeder, A. Allchin, K. Berlin, and M. Cakmak. Characterizing the design space of rendered robot faces. In *Proceedings of 13th ACM/IEEE International Conference on Human Robot Interaction (HRI)*, pages 96–104, Chicago, USA, Mar. 2018.

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

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

[17] S. M. Chacko and V. Kapila. 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, November 2019.

[18] K. Chandan, V. Kudalkar, X. Li, and S. Zhang. 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, May 2021. IEEE.

[19] C. P. Quintero, S. Li, M. K. Pan, W. P. Chan, H. M. Van der Loos, and E. Croft. 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, October 2018.

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

[21] W. P. Chan, G. Hanks, M. Sakr, H. Zhang, T. Zuo, H. M. Van der Loos, and E. Croft. 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, 2022.

[22] T. Williams, M. Bussing, S. Cabrol, E. Boyle, and N. Tran. 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, Mar. 2019.

[23] E. Rosen, D. Whitney, M. Fishman, D. Ullman, and S. Tellex. 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, Oct. 2020.

[24] Z. Qiu, T. Eiband, S. Li, and D. Lee. Hand pose-based task learning from visual observations with semantic skill extraction. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 596–603. IEEE, 2020.

[25] C. Eze and C. Crick. Learning by watching: A review of video-based learning approaches for robot manipulation. *arXiv preprint arXiv:2402.07127*, 2024.

[26] W. Si, N. Wang, and C. Yang. A review on manipulation skill acquisition through teleoperation-based learning from demonstration. *Cognitive Computation and Systems*, 3(1):1–16, 2021.

[27] M. B. Luebbers, C. Brooks, M. J. Kim, D. Szafir, and B. Hayes. 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, Mar. 2019.

[28] M. B. Luebbers, C. Brooks, C. L. Mueller, D. Szafir, and B. Hayes. 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, June 2021.

[29] D. Puljiz, E. Stöhr, K. S. Riesterer, B. Hein, and T. Kröger. 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, Nov. 2019.

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

[31] A. Rivera-Pinto, J. Kildal, and E. Lazkano. 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, 2023.

[32] S. Macciò, M. Shaaban, A. Carfì, R. Zaccaria, and F. Mastrogiovanni. 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, Aug. 2023. doi:10.48550/arXiv.2309.04765.

[33] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. ROS: an open-source robot operating system. *ICRA workshop on open source software*, 3(3.2):5, May 2009.

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

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

[36] E. Ruffaldi, F. Brizzi, F. Tecchia, and S. Bacinelli. 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, June 2016.

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

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

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

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