Ghost Arm: Aligning Human and Robot Kinematics through AR Overlays in MoCap-Based Teleoperation of Robot Arm

Qiushi Zhou
The Hong Kong University of Science and Technology (Guangzhou)
China
qiushizhou@hkust-gz.edu.cn

Antony Chacon
University of Melbourne
Australia
antony.chacon@unimelb.edu.au

Jiahe Pan
ETH Zürich
Switzerland
jiapan@student.ethz.ch

Wafa Johal
University of Melbourne
Australia
wafa.johal@unimelb.edu.au

Abstract

Teleoperating a robot arm typically requires either positioning its end-effector directly or programming joint movements. The former limits posture and trajectory control, while the latter lacks intuitive, real-time feedback. Unlike controlling their own arms through integrated visual and proprioceptive cues, humans face difficulties with robot arms due to mismatched orientations and joint kinematic configuration. In this work, we investigate how different augmented reality (AR) visualisations can support teleoperation via motion capture (MoCap). Across two user studies (N=24, N=24), we evaluate how overlaying a virtual human arm alongside the robot arm affects user performance and experience. Our results show that the humanoid AR overlay facilitates learning the control mapping, but its benefits diminish once users become familiar with the task. We contribute empirical evidence and design insights for AR- and MoCap-based approaches to more effective robot arm teleoperation.

CCS Concepts

• **Human-centered computing** → *Interactive systems and tools*; **Virtual reality**; **Empirical studies in HCI**.

Keywords

teleoperation, robot arm, augmented reality

ACM Reference Format:

Qiushi Zhou, Antony Chacon, Jiahe Pan, and Wafa Johal. 2026. Ghost Arm: Aligning Human and Robot Kinematics through AR Overlays in MoCap-Based Teleoperation of Robot Arm . In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '26), April 26–May 1, 2025, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

CHI '26, April 26–May 1, 2025, Barcelona, Spain

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Robot arms are popular for a wide range of general-purpose tasks thanks to their efficient form-factors. These robots can be controlled in teleoperation tasks by positioning the end-effector alone or by defining the rotation vector of each joint [26, 28]. However, robot arms also present a unique affordance of behaving like human arms, thanks to their similar structures and movement ranges. This unique feature suggests an interesting opportunity for a novel approach of teleoperation—through motion-capture (MoCap) control that maps the rotations of the joints of the human operator's arms to the rotations of the corresponding joints on a robot arm. This approach to teleoperation is particularly advantageous in real-world tasks requiring both rapid reaching and precise posing, such as remote surgical assistance or complex assembly, especially where the operator must maintain high spatial awareness and posture control to navigate cluttered environments.

Robot arms are typically situated on desktops with the base joint fixed in a vertical orientation that affords the maximum movement range in the space over the supporting platform. A similar movement range of the human arm is the space on one side of the body, which is perpendicular to the orientation of the movement range of the robot arm. This difference in **orientation** presents a challenge for MoCap-based teleoperation control for human operators to intuitively anticipate the movement directions of the robot arm that reacts to their arm movements [28]. Further, the ambiguity in the perceived **kinematic structures** of the robot arms [28] challenges operators to understand how the joint rotations are mapped between the human and the robot arms.

In this work, we explore how augmented reality (AR) can assist MoCap-based teleoperation of a robot arm by rendering a virtual arm as visual reference that mediates the inconsistencies between the human arm and the robot arm. Across two user studies, we aim to address two research questions. **RQ1**: What is the optimal visualisation, regarding the orientation and the kinematic structure, of an AR arm to facilitate MoCap-based teleoperation? **RQ2**: How does an AR overlay of a virtual arm with a human-like kinematic structure facilitate MoCap-based teleoperation of a robot arm?

In Study 1, we investigate how a virtual arm rendered in AR next to the robot arm could help improve user performance and experience of a target reaching teleoperation task. We implemented three conditions of the AR arm that are either in a humanoid or robotic kinematic structure, and either in the same orientation with the robot arm or with the human arm. We concluded that the optimal configuration is a humanoid arm overlaid on the physical robot arm in the same orientation with it to assist the understanding of the control mapping and to ensure easy visual access. However, findings from Study 1 also suggested limitations, that the end-effector reaching task may not be able to fully reflect the effect of the AR visualization, and that the divided attention between the AR arm and the robot arm may prevent the users from referencing it.

In Study 2, we evaluated the overlay of a humanoid AR arm on the robot arm using a posture matching teleoperation task against a baseline condition without the AR arm. We found that the AR arm helped reduce the perceived physical demand, effort, and frustration. While most participants found the AR arm helpful in learning the control, it was considered more suitable for this purpose than an always-on visual guidance for teleoperation. Finally, we offer an in-depth discussion of the key findings from the two studies that directly inform future work exploring the use of AR visualisations to facilitate MoCap-based teleoperation of robot arms.

2 Background

2.1 Robot Teleoperation and Control

Robot teleoperation enables human operators to remotely control a robot to perform tasks that are difficult for the human body or in areas not suitable for human habitation [1], typically through usable manipulation interfaces that provide visual feedback [7]. It can be a difficult task that demands extensive training and expertise, especially for robots with high degree-of-freedom (DOF) (e.g., industrial manipulators and aerial robots) that challenge human operators to anticipate how the robot would move in the environment using the control interfaces provided [14]. While it is possible to teleoperate a robot through abstract programming, such as directly editing the rotation angles of a robot arm in real-time, more intuitive metaphors have been explored, such as "virtual fixture" that provides a simulated abstract sensory information through the command interface that is mapped to the sensory feedback from a remote environment [31], which has shown to reduce operation time and increase teleoperation efficiency and accuracy [21].

Other works have explored varying extent of motion tracking for robot control. For instance, Franzluebbers and Johnsen compared the uses of motion-tracked VR controllers and computer mouses for simulated robot teleoperation and found that motion controls provided higher peak performance, likely due to faster gross movement planning [8]. Kurpath et al. proposed an Inertial Measurement Unit (IMU)- and potentiometer-based controller that can capture the human arm-hand functionalities up to 17 DoF for complex teleoperation tasks, and demonstrated its efficacy through simulation and experimental results [18]. Rakita et al. proposed a motion retargeting method for mimicry-based teleoperation of robot arms while recognising the difficulty of implementing a direct mapping between the user's hand and the robot's end effector due to their different kinematic and speed capabilities [28]. In this work, we investigate whether this control mapping could be eased with the mediation of real-time visual feedback rendered in AR.

2.2 AR for Human-Robot Interaction (HRI)

AR has been used for HRI for different purposes [44], such as to support real-time control and teleoperation [22, 25, 38]. Without AR, teleoperation are typically conducted with visual feedback on 2D screens, suffering from distorted perceptions of cues (e.g., egocentric distances) relevant to task execution [3], while requiring users to check the screen and the physical robot back-and-forth [30]. For remote teleoperation, AR visualisations enhance situational awareness and reduce cognitive load by immersing the operator in a representation of the remote environment and overlaying information related to the task [47]. These types of visualisations afford high-fidelity control, such as those based on hand tracking, by rendering the remote space with reference to the size and range of the hand movement of the human operator [20]. AR has also been explored for real-time 3D rendering of robotic instruments, hand-held instruments, and endoscope to aid robot-assisted surgeries [27].

For co-located HRI tasks, AR can be used to visualise motion intent of robots intuitively by directly showing their potential movements within the physical space [30], allowing users to understand the motion intent more easily [10]. Specifically, motion intent communicated through anthropomorphic virtual kinematics has been found effective, thanks to the ease of understanding human-like movement from robot arms that share anatomical structures with human arms [11, 36]. In this work, we investigate an AR visual-feedback method that overlays an anthropomorphic arm on the robot to convey its motion.

2.3 AR for Anthropomorphic Robot Teleoperation

Following early works on screen-based feedback for robot teleoperation [32, 33], later works rendered remote scenes in AR, enabled by advanced sensing and rendering capabilities [19]. These works were motivated by the intuitive mapping from the anticipated consequences of the movement of the remote robot to what the human operator can directly see and feel. This potential benefit for the ease of control is particularly relevant to the teleoperation of anthropomorphic robots, such as human-like robots and robot arms that share relatively similar anatomical features and movement DOF with human arms [42, 46]. These robots afford the unique possibility to perform human-like movements for tasks that not only benefit from the robotic capabilities but also demand anthropomorphic kinematic structures of movements, inspiring works such as [12] that generates human-like naturalistic upper-limb for humanoid robot arms. While previous work have found that anthropomorphism in robots help human users perceive robot actions [15, 29, 50], to our knowledge, no previous work has explored how to enable better robot teleoperation control that capitalise on the ease of understanding anthropomorphic robotic movements.

AR is a plausible medium to help users learn to teleoperate robots with anthropomorphic features. Whereas previous work explored AR-enabled robot control, such as using virtual shadows on physical floors to position drones [4] and interactive robot programming with virtual movement cues and anchors rendered in the physical environment [24], it remains to be investigated how AR can help teleoperation by visualising interactive anthropomorphic movement cues for operators to understand and learn the control of robots using their own body movement. This vision of AR-mediated teleoperation control is inspired by the widely-used feature of MoCap-based avatar control in XR. Controlling a distant avatar (not from a first-person perspective (1PP)) relies on a successful integration of spatially dissociated proprioceptive information from the user's physical body and visual feedback from the virtual avatar [2]. Numerous previous work have found that it is plausible to control and develop a sense of ownership over a distant avatar as long as synchronous movement with the user is present [17, 23]. Specifically, a series of motor training studies featuring throwing tasks have found that training provided through a third-person perspective mixed reality view (similar to controlling an avatar) yielded better results than from the 1PP, thanks to the visual feedback that places the virtual representation of the user in reference to the task objects [5, 6, 41], suggesting its suitability "for displacement actions and interaction with moving objects [34]." Further, while most previous works on avatar control featured full-body representations of users, some also proved that it is also viable to control a separate part of the body, such as the hand [35, 39]. Inspired by these explorations in avatar control, and by previous work on teleoperating robots through virtual surrogates rendered in AR [45], we explore how a virtual arm in AR can help mediate the teleoperation of a robot arm through MoCap-based control by the human arm.

3 MoCap-Based AR Teleoperation System

By exploiting the similarities in the anatomical features between the human arm and robot arms, we explore how a virtual arm visualised in AR can help users understand and learn the control while mitigating the inconsistencies between the robot and the human arms regarding *Orientation* and *Kinematic Structure* in Study 1, and determine the optimal configuration of the AR arm. In Study 2, we explore how it affects user performance and experience of a posture-matching task. Both studies were approved by the IRB of our university. In this section, we describe our novel implementation of the MoCap-based teleoperation system for a robot arm with AR integration as apparatus for the studies reported in this paper.

3.1 System Structure

We utilised three key hardware components: 1) a 7-DOF robotic arm Franka Research 3 (FR3), 2) HoloLens 2, and 3) OptiTrack motion capture system. The system enables teleoperation of the FR3 by capturing the operator's right arm movements, in particular the upper arm, forearm, and hand orientations. The OptiTrack 8-camera system captures the arm's orientation, while Unity, running on a Windows machine, processes the motion data and translates it into real-time commands for the FR3. HoloLens 2 provides real-time rendering of a virtual arm that corresponds to the movement of the physical robot arm in AR.

Robot control is handled by a Linux subsystem, which processes the motion commands through a custom joint trajectory controller and communicates with the Franka Control Interface. A ROS TCP Endpoint facilitates real-time data transfer between Unity and the robot controller, while joint position tracking monitors the robot's kinematics. Figure 1 illustrates the communication flow between these components.

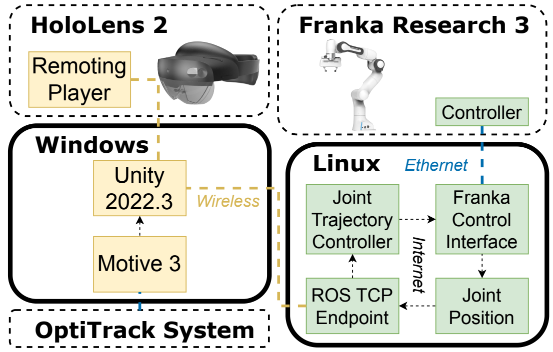

Figure 1: System architecture that describes the communication between a Linux machine running ROS 2 and the Franka Control Interface, a Windows machine running Unity and Motive, an OptiTrack system, a Franka Research 3 robot, and a HoloLens 2 that renders through a remoting player.

3.2 Human to Robot Mapping

We employed the FR3 robot because it resembles the structure of a human arm while maintaining the typical layout of an industrial 7-DOF manipulator. It features groups of joints that correspond to the shoulder (which rotates in three degrees of freedom), elbow (flexion/extension and forearm supination/pronation) and hand (flexion/extension).While the robot has some joints with non-zero offsets, this does not impact the mapping process, as we rely solely on the orientation of the arm segments (i.e., upper arm, forearm, and hand). Adapting this mapping to other manipulators with a similar configuration to the FR3 robot would be straightforward.

The robot's joint angles are denoted as θ_1 through θ_7. The tracked positions of the shoulder, elbow, and wrist in three-dimensional space are represented as vectors: $\mathbf{v}_{\text{shoulder}}$, $\mathbf{v}_{\text{elbow}}$, and $\mathbf{v}_{\text{wrist}}$, respectively. Each vector is expressed in the form $\mathbf{v}_i = (x_i, y_i, z_i)$. Figures 2 and 3 illustrate the relationship between human arm motion and the robot's joint angles in a left-handed coordinate. Given the vector

$$\mathbf{v}_{\text{upper\_arm}} = \mathbf{v}_{\text{elbow}} - \mathbf{v}_{\text{shoulder}} \tag{1}$$

and the robot orientation, we define the second joint θ_2 as

$$\theta_2 = \text{atan2}(z_{\text{upper\_arm}}, x_{\text{upper\_arm}}) \tag{2}$$

The first joint θ_1 maps the elevation of the upper arm during flexion/extension. Thus, we determine it as

$$\theta_1 = \text{atan2}\left(y_{\text{upper\_arm}}, \sqrt{x^2_{\text{upper\_arm}} + z^2_{\text{upper\_arm}}}\right) \tag{3}$$

With

$$\mathbf{v}_{\text{forearm}} = \mathbf{v}_{\text{wrist}} - \mathbf{v}_{\text{elbow}} \tag{4}$$

we determine the fourth joint angle θ_4 as

$$\theta_4 = \arccos\left(-\frac{\mathbf{v}_{\text{upper\_arm}} \cdot \mathbf{v}_{\text{forearm}}}{|\mathbf{v}_{\text{upper\_arm}}| \cdot |\mathbf{v}_{\text{forearm}}|}\right) \tag{5}$$

The third and fifth joint angles, θ_3 and θ_5, map internal/external rotation of the upper arm and supination/pronation of the forearm respectively. Thus, we used the local rotation of these objects along the axes $\mathbf{v}_{\text{upper\_arm}}$ and $\mathbf{v}_{\text{forearm}}$ in Unity. Similarly, we derived

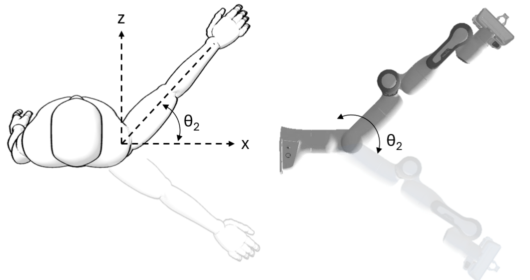

Figure 2: Robot orientation comparing with human arm.

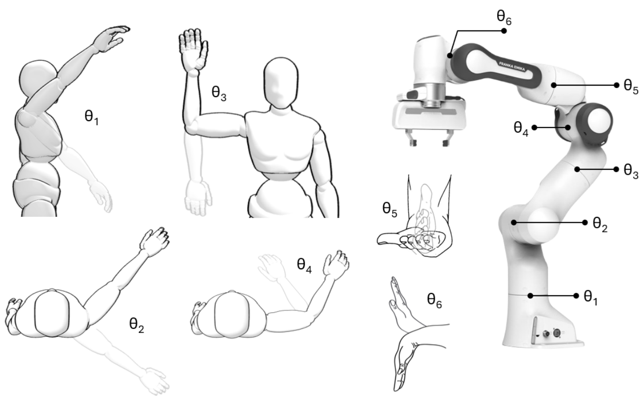

Figure 3: Corresponding joint angles of human and robot.

the final two joint angles using the relative orientations obtained directly from Unity.

3.3 MoCap implementation

We employed an OptiTrack system and its software interface, Motive 3. Eight Prime-13W cameras were mounted on a tracking rig attached to the ceiling of the usability lab, spaced out and focused on the participants' right arm. We employed 3 sets of markers to track the orientation of the upper arm, forearm, and hand as seen in Figure 4. Motive 3 transmits these data to Unity at an average speed of 200 KB/s.

3.4 AR implementation

We used the HoloLens 2 and the Holographic Remoting Player app, connecting it via LAN to a PC. The Mixed Reality Toolkit 3 (MRTK3) allowed us to control the HoloLens 2 in real-time through our Unity implementation. For calibration, we placed a QR code on the back of the robot at a height of 18 cm and 6 cm behind the center of its base. We used the Reality Collective package[1] to recognise the QR code via the HoloLens 2 cameras and anchor it in the real 3D space within Unity. Before starting the tasks, we instructed participants to look directly at the QR code from a distance of 30 cm, enabling the system to recognise the robot's position.

[1]https://github.com/realitycollective

Figure 4: MoCap markers placed on the arm. Markers for the upper arm and forearm were attached with kinesiology tape. We use the Quantum Metaglove's markers for wrist tracking.

4 Study 1: configuring the AR visualisation

When a typical 7-DOF robotic arm (e.g., Franka Research 3, Kinova Gen3, etc.) is situated on a horizontal platform, its range of movement is similar to a human arm on the side of the body (Figure 2). Therefore, in MoCap-based teleoperation of robot arms, the human arm's motor space needs to be converted to cover the movement range of the robot arm. A challenge that arises from this is the mental effort for the human operator to mentally convert the rotation direction of their arm joints to those of the robot. Further, the different kinematic structures between the human and the robot arm may challenge the user to understand how the robot arm would respond to their control. In light of these questions, we formalise our first research question **(RQ1): What is the optimal visualisation, regarding the orientation and the kinematic structure, of an AR arm to facilitate MoCap-based teleoperation?** We employ a within-subject design with four Visualisation conditions (Figure 5) to compare and determine the optimal configuration:

- HumanHorizontal (HH): A virtual humanoid arm in the same orientation as the human arm (Figure 5 (b)).
- HumanVertical (HV): A virtual humanoid arm in the same orientation as the physical robot arm (Figure 5 (c)).
- RobotHorizontal (RH): Virtual replica of robot arm in the same orientation as the human arm (Figure 5 (d)).
- Baseline: No AR visualisation was rendered.

Using these conditions, we investigate how AR visualisations help users understand the control mapping between their arm joints and those of the robot to perform teleoperation tasks by providing a visual reference of the orientation (RH), kinematic structure (HV), or both features (HH) of the human arm.

4.1 Physical Setup

Studies 1 and 2 were conducted in a room with dimensions of 5.5×3.6 m. We set the AR visualizations for HH, HV, and RH with respective offsets of (-0.8, 0.33, 0.4) m, (-0.5, 0.33, 0.4) m, and (-1, 0.33, 0.4) m along the x, y, and z axes, relative to the robot base in a left-handed coordinate. The participant sat 1 m to the left and 1.2 m behind the robot's base. These distances were chosen to ensure comfortable visual access for the virtual arms within the field of view of the HoloLens 2. Participant were not positioned directly behind the robot to avoid occlusion of the virtual targets.

4.2 Task

The task entailed tracing a ring of 11 virtual circular targets by reaching for them with the robot's end-effector in a pre-defined sequence, as illustrated in Figure 5 (e). We used the same sequence across all trials to ensure that performance variations reflected the effectiveness of different visualization conditions rather than participants' ability to adapt to randomized target sequences. The ring of circular targets was parameterised by the radius $R = 22.5$ cm and the diameter $d = 5$ cm. The ring was positioned on a vertical plane directly in front of the robot. The centre of the ring was located at a displacement of (0, 0.56, 0.9) m relative to the robot's

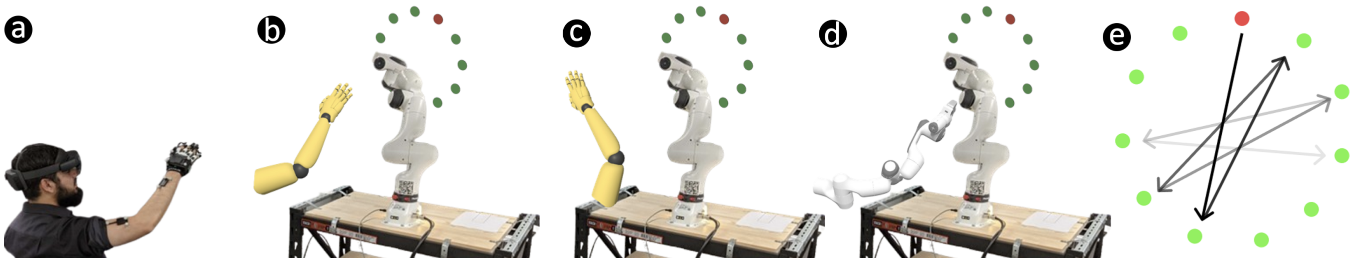

Figure 5: AR Visualisation conditions and apparatus: (a) Participant situated away from the robot; (b) HumanHorizontal arm in AR next to the physical robot; (c) HumanVertical; (d) RobotHorizontal; (e) Target order of the reaching task.

base, which was determined through pilot testing to allow the user's arm to stay within a comfortable range of movement during the task. The target reaching task was chosen as it has been employed to benchmark robot efficiency in manipulation tasks [9, 37] and to evaluate interface design and human movement performance in HCI [43, 48, 49]. During each trial, the active target was always in red, and turns green upon a successful selection, which requires the robot's end-effector position to be within a perpendicular distance of 10 cm from the target plane. Participants were asked to complete the task as fast as possible, while minimising sudden movements to avoid triggering emergency-stop of the robot.

4.3 Participants

We recruited 24 participants (11 female, 13 male) with a mean age of 26.9 years ($Min = 19, Max = 36, SD = 4.5$) using the University's online notice board. Participants rated their prior experience with VR and AR on a 7-point discrete scale from 1 (never used) to 7 (use frequently) with a mean rating of 2.13 ($Min = 1, Max = 6, SD = 1.45$). Each participant took approximately 1.5 hours to finish the experiment and was compensated with a $20 voucher.

4.4 Procedure

Participants were first introduced to the study and asked to filled out consent and demographics forms. We calibrated Microsoft HoloLens' built-in eye-tracker for each participant, and placed the OptiTrack markers on their right arms. We then explained the tasks and the conditions, and used a 2-minute training block for participants to practice teleoperating the robot and reaching towards the ring of targets under the Baseline condition to familiarise with the system and minimize potential learning effects. Then, participants completed four rounds (two trials for each condition) of tasks with breaks between consecutive trials. We limited the procedure to two trials per condition to minimize physical fatigue, as pilot testing indicated that sustaining arm postures led to discomfort and decreased performance over extended sessions. The order of the conditions were fully counterbalanced using a Latin square.

After each condition, participants filled out a questionnaire on their perception of the task and system during that round and therefore had a rest of 2-3 minutes. After completing all conditions, participants were asked to complete a questionnaire ranking their preferences for the different conditions. The study duration was approximately 90 minutes.

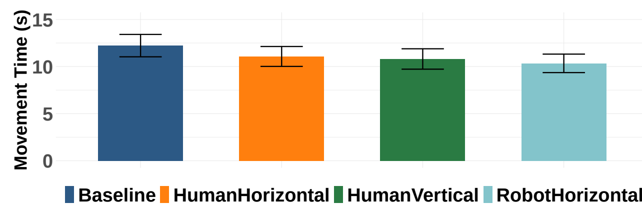

Figure 6: Movement Time by Visualisation in Study 1 (Error bars indicate standard error). A one-way ANOVA revealed no statistically significant differences.

4.5 Measures

4.5.1 Task Performance. Performance was evaluated through *Movement Time*, which was the time interval between the appearance of the target and the successful selection. For each trial, 11 timestamps of successful selection were recorded, from which 10 intervals were computed as the *Movement Time* data points. This resulted in: 10 *time intervals* × 2 *trials* × 4 *Visualisations* = 80 data points per participant.

4.5.2 Questionnaire. We administered a NASA-TLX questionnaire [13] with all six original sub-scales (20-point discrete scales from 1-*very low* to 20-*very high*) in randomised orders at the end of each Visualisation condition. After finishing all tasks, participants ranked the four conditions based on their preferences, and answered interview questions: ***Q1***. *Do you think the AR arms were helpful for your controlling of the robot arm?* ***Q2***. *Do you think it was more helpful with the virtual arm in the same orientation with the physical robot or with your own arm?* ***Q3***. *Do you think it was more helpful seeing a virtual human arm or seeing a virtual robot arm?*

4.6 Quantitative Results

All measures were verified as normally distributed using the Shapiro-Wilk test, with test statistics ranging from $W = 0.87$ to 0.95 and all $p > .05$. We performed one-way repeated-measures ANOVA with Visualisation as independent variable.

4.6.1 Movement Time. We first examined the movement time data for potential learning effects across the two trials (trial 1 & trial 2). For this, we pooled data across all participants and Visualisation conditions, resulting in 96 pairs for a paired-samples t-test, which revealed a significant difference in movement time ($\Delta = 1.63s, t(95) = 3.53, p < .001$). We subsequently performed

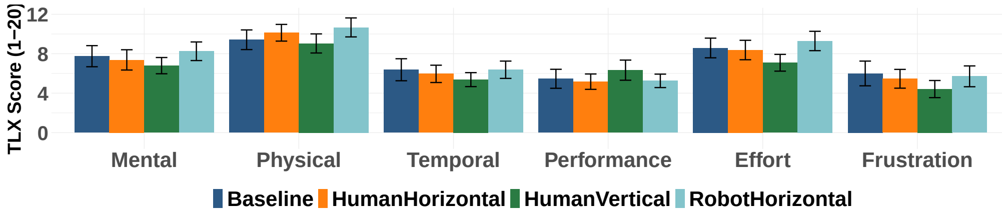

Figure 7: Study 1 questionnaire results: NASA-TLX scores (Error bars indicate standard error). ANOVA indicated no statistically significant differences for any sub-scale.

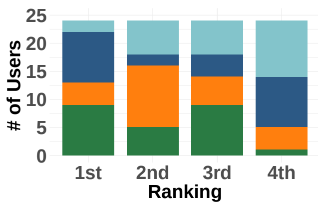

Figure 8: Study 1 questionnaire results: post-study rankings of the visualisations.

paired sample t-tests between trial 1 and trial 2 separately for each of the Visualisation conditions, pooling data across all participants for each condition. HumanVertical yielded a significant difference ($\Delta = 1.75s, t(23) = 2.07, p < .05$) , and a marginal difference was also observed for HumanHorizontal ($\Delta = 1.95s, t(23) = 2.06, p = .05$) . No clear differences were observed for the other conditions. Given the observed learning effect across the two trials, for the subsequent ANOVA we removed all trial 1 movement time data from the analysis. One-way ANOVA revealed no clear difference in movement time ($F_{3,69} = .85, p = .471$) by Visualisation (Figure 6).

4.6.2 NASA-TLX and Preference Ranking. ANOVA indicated no statistically significant differences in the NASA-TLX scores ($F_{3,69} = .74, p = .529$), where the six sub-scales showed internal consistency (Cronbach's $\alpha = .82$) (Figure 7). With the rankings for user preference of different Visualisation conditions (1 being highest), HV was the most preferred ($M = 2.08, SD = .97$), followed by HH ($M = 2.38, SD = .97$), Baseline ($M = 2.54, SD = 1.35$), and finally RH ($M = 3.00, SD = 1.02$) (Figure 8).

4.7 Qualitative Results

The post-study questionnaire yielded 90 responses. Two researchers collaboratively carried out a general inductive analysis, using independent parallel coding to categorise notable factors mentioned by participants that corresponded with their layout creation and adaptation style [40]. We present categories of the answers from participants in Table 1 with the number of participants whose answers fall in each category.

4.7.1 Q1: Was AR helpful for the task and why? Twelve participants reported that the AR visualisation was helpful for completing the task. The common reasons include that the visualisation provided visual reference for them to conveniently see how the robot moves in correspondence to their control, without needing to to look back-and-forth between the robot and their arms: "*It's very helpful especially the vertical human arm ... I don't need to think of the direction (where) my arm goes to control the robot* (R3)." Participants also mentioned that the virtual arm gave them confidence in teleoperation (P22), especially in how far the robot moves away into distance (P15).

Five participants reported that their attention was drawn to the targets and the robot, and did not actively use the visualisation to aid the task. Four participants mentioned that the AR arm would be helpful for learning how the movement directions are mapped between the human arm joints and the robot arm joints at the beginning of the teleoperation, however, they may not be necessary after the user masters the control: "*As I got used to the movements, (the visualisation) could probably be ignored* (P19)." "*(They were) sort of helpful. But as I was quite fast getting used to (the task) so I gradually reduced the frequency I referred to the AR arms* (P21)."

Table 1: Answers to interview questions in Study 1.

Q1: Was AR helpful for the task and why?	**# of P**
Yes	12
Did not pay attention	5
Learning	4
Not helpful with no specific reason	3
Q2: Which orientation was more helpful?	**# of P**
Vertical	13
Horizontal	7
Did not pay attention	4
Q3: Which kinematic structure was more helpful?	**# of P**
Human Arm	15
No Preference	7
Robot Arm	2

4.7.2 Q2: Which orientation was more helpful? Thirteen participants reported that the vertical orientation of the AR arm was more helpful because it was in the same orientation as the physical robot arm, making it easier for them to understand the control. Seven participants preferred the horizontal orientation because it directly represents how their own arm behaves.

4.7.3 Q3: Which kinematic structure was more helpful? Fifteen participants thought the Human arm was more helpful because it was like moving their own arm close to the physical robot arm for easier visual reference. Among these participants, 9 specified that the HumanVertical arm was most helpful to convey the mapping of joint rotations between their own arms and the robot arm. Seven participants did not have preference, while two preferred the robotic kinematic structure.

5 Discussion of Study 1

While the Visualisation conditions did not yield statistically significant differences in movement time or perceived cognitive load across the NASA-TLX sub-scales, qualitative feedback from most participants suggested that they still found the virtual arm helpful. For other participants, the main reason why the virtual arm did not help was that they did not pay attention to it, which also associates with the lack of differentiation in the resulting perceived workload. The repetitive nature of the target reaching task allowed them to adapt to the rhythm and the alternating direction of arm movements without really understanding the control mapping. Further, it was easier to ignore the virtual arm because it was rendered away from the physical robot, while participants needed to pay attention to the targets and the end-effector during the task.

Among the Visualisation conditions, only HumanVertical yielded a significant improvement between two consecutive trials, indicating learning effect. While the vertical orientation and the human-like kinematic structure of the virtual arm were most preferred by participants respectively (Table 1), it is also the only configuration of the AR visualisation that can be overlaid on the physical robot arm to provide easier visual access for participants. This finding echoes previous work that realistic representations of users in XR benefit control. For instance, the authors of [23] compared the performance of a reaching task in VR while varying the level of details of the avatars from a cardboard body, a virtual humanoid avatar, and a real-time point cloud capture. They found that it is possible to develop a sense of ownership and spatial awareness towards a distant avatar that is similar to the 1PP only when it is coupled with a realistic representation, such as using the RGB point cloud capture [23]. Similarly in Study 1, the humanoid arm presents a more realistic kinematic structure for the user to understand, and potentially to develop a better sense of embodiment. For these reasons, to address **RQ1**, we propose HV as the optimal among the tested configuration of the virtual arm for human operators to benefit from AR visualisation. However, we need further understanding of the effect of AR visualisation in tasks where conscious control of the arm posture (beyond the end effector) is needed, and where the user does not need to split attention between teleoperating the robot and observing the AR arm.

6 Study 2: Posture Control

In light of the findings from Study 1, we ask our second research question **(RQ2): How does an AR overlay of a virtual arm with a human-like kinematic structure facilitate MoCap-based teleoperation of a robot arm?** Accordingly, we made the following design choices for Study 2:

- **Overlaying the AR arm**: Participants in Study 1 reported that rendering the AR arm next to the physical robot arm led to an extra attentional cost that prevented them to observe and benefit from the visual reference of the AR arm. To prevent this potential confound of the study, we overlay the AR arm on the physical robot arm to ensure easy visibility.

Figure 9: Study 2 (left to right): Participant; Condition of no visualisation of AR arm on robot; Condition with visualisation of AR arm on robot. Blue and red sphere rendered on elbow and wrist respectively. Lighter blue and red denote targets to match posture.

- **Posture control task**: The task in Study 1 only required participants to reach the targets using the end effector of the physical robot arm, allowing them to adapt to the rhythm and the predictable changes of direction of movement, hence reducing the task difficulty while minimising the effect of the AR visualisations. In Study 2, we adopt a posture control task that demands the participants to move and orient the whole arm correctly to finish a trial. While maximising the potential benefit of the AR visualisations with higher difficulty, the posture control task represents another type of real-world scenarios of teleoperation tasks where the robot arm needs to be controlled to manoeuvre in complex environments to perform operations while avoiding obstacles.

We employ a within-subject design that compares task performance and subjective measures between Visualisation conditions: AR Arm where the virtual arm is presented during the tasks, and No Arm as a baseline without virtual arm.

6.1 Task

The posture matching task entailed sequentially matching a series of virtual target robot postures using the real robot arm, as shown in Figure 9. A posture was defined by the Cartesian positions of the elbow and wrist joints of the robot relative to its base, where their orientations were not considered. We rendered a blue sphere on the robot elbow and a red sphere on the robot wrist for visual reference. A target posture is successfully matched when the elbow and wrist joint positions of the robot are both within 5 cm from their respective target positions given by the target posture (light blue and red). We selected 4 target postures from a pool of 16 candidate postures through pilot testing, as illustrated in Figure 10, which ensured users always stayed within a comfortable range of arm movements. The posture matching task was designed to model real-world scenarios where joint-space control of the robot arm is required to avoid collisions with the environment during teleoperation, which may arise in highly cluttered workspaces.

In each trial, participants were asked to match four target postures in orders that were counterbalanced using a Latin square. The targets disappeared once the participant matched both during the task. Participants were asked to complete the task as fast as possible while minimising sudden movements.

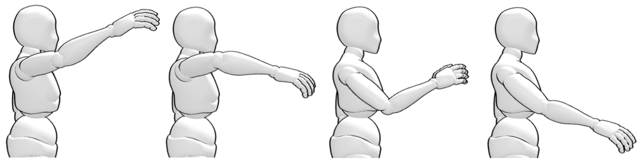

Figure 10: S2 postures (left to right): Elbow up, wrist up; Elbow up, wrist down; Elbow down, wrist up; Elbow down, wrist down.

6.2 Participants

We recruited 24 participants (14 female, 10 male) with a mean age of 25.6 years ($Min = 19, Max = 37, SD = 4.6$) using the University's online notice board. Participants rated their prior experience with VR and AR on a 7-point discrete scale from 1 (never used) to 7 (use frequently) with a mean rating of 2.04 ($Min = 1, Max = 7, SD = 1.27$). Each participant took approximately 50 minutes to finish the experiment and was compensated with a $20 voucher.

6.3 Procedure

After filling out consent and demographics forms, the HoloLens' eye-tracker was calibrated for each participant, and the reflective markers were placed onto their right arms. Participants completed 2 rounds (4 trials each) of the task under the two conditions respectively. In each trial, participants reached towards the target posture, while always starting with the physical robot arm in a "straight" posture with full extension. After each condition, participants filled a questionnaire during a 2-min break. After all conditions, an interview was conducted to collect feedback on their perception of the task and visualisation. The study took approximately 50 minutes.

6.4 Measures

6.4.1 Task Performance. *Movement Time* is defined as the time taken to successfully match each target posture from when they first appear. Each condition block contains 4 trials corresponding to the 4 target postures, and hence 4 *Movement Time* data points were recorded. This resulted in: 4 *time intervals* × 2 *conditions* = 8 data points per participant.

6.4.2 Usability Questionnaire. We administered the NASA-TLX questionnaire [13] in the same manner as in Study 1. In the end, participants answered the following interview questions: ***Q1****. Do you think the AR arms were helpful for your controlling of the robot arm?* ***Q2****. Have you learned how to control the robot arm from seeing the virtual arm?*

6.5 Quantitative Results

We first preprocessed the movement time data by removing outliers on a per-trial basis using the interquartile range (IQR) method [16], where individual data points outside the range [Q1-1.5 IQR, Q3+1.5 IQR] were removed. This resulted in 15 (8%) outlier trials being removed. We also verified that no statistically significant learning effect existed across any of the measures. Finally, after all measures were verified to be normally distributed using the Shapiro-Wilk test, we performed one-way ANOVAs with Visualisation for the performance and TLX measures.

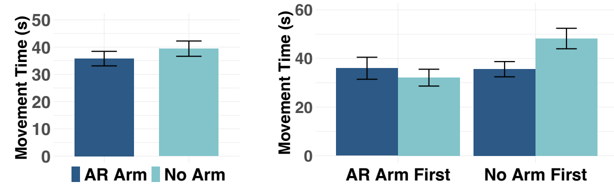

Figure 11: Study 2 results (Error bars indicate standard error): Per-trial movement time by Visualisation (left), and separate plots by condition order (right). A Linear Mixed Model revealed no statistically significant differences.

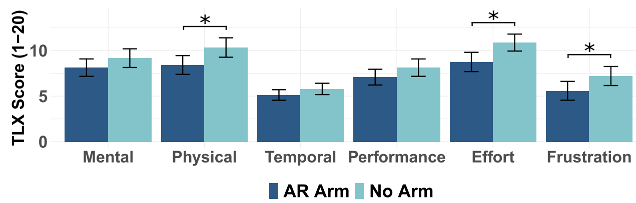

Figure 12: Study 2 results on the NASA-TLX questionnaire (Error bars indicate standard error). Statistically significant differences are indicated by asterisks (*).

6.5.1 Movement Time. We analysed movement time using a Linear Mixed Model (LMM), with the *Visualization* condition as the fixed effect and a random intercept for each participant. *P*-values were derived using Satterthwaite's approximation for degrees of freedom. We found no statistically significant effect of *Visualisation* ($F_{1,21.00} = 2.649, p = .119$) on movement time. We present the plots comparing the two *Visualisation* conditions altogether, and also from each group of participants who performed the task in different orders to aid interpretation of results (Figure 11).

6.5.2 NASA-TLX. The ratings across six sub-scales showed internal consistency (Cronbach's $\alpha = .86$). The perceived Physical Demand yielded a significant effect of *Visualisation* ($F_{1,23} = 5.20, p < .05$), and similarly for Effort ($F_{1,23} = 5.06, p < .05$) and Frustration ($F_{1,23} = 5.14, p < .05$).

6.6 Qualitative Results

We analysed the questionnaire answers from participants in the same manner as Study 1, and present results in Table 2.

6.6.1 Q1: Was AR helpful for the task and why? Thirteen participants reported that the AR visualisation was helpful for them to perform the task, for instance: "*It helped me with the position of the arm to have a reference. Without that, I couldn't imagine what was the best position to meet the final destination* (P2)."; " *(It) helped me learn what angle I should rotate to get to the point because the direction the robot (moves) doesn't match the direction of my arm* (P10)." Six participants mentioned that the AR visualisation was helpful specifically because it helped them learn the control at the beginning, indicating that the task becomes easy after a few rounds, for which they did not need the visualisation any more. Two participants reported that they did not find AR helpful because they felt the visualisation was either lagging or not perfectly overlaying the physical robot.

Table 2: Answers to interview questions in Study 2.

Q1: Was AR helpful for the task and why?	# of P
Yes	13
Yes for initial learning	6
No	2
No for technical issue	2
Did not pay attention	1
Q2: Was AR helpful for learning control and why?	**# of P**
Yes	15
Yes but only at the beginning	6
No (no specific reason)	1
No for technical issue	1
Did not pay attention	1

6.6.2 Q2: Was AR helpful for learning the control and why? Fifteen participants reported that the AR visualisation was helpful for them to learn the control mapping, for instance: "*Yeah, it was very intuitive. I didn't have to learn anything beforehand, just having that reference was enough for me to do the task* (P2)."; " *It's like when the arm is there, (I feel like I have) more connection with all this* (P17)."; " *Because the arm just looks like mine, so I can see if I was wrong* (P24)." Six participants mentioned that AR helped them learn the control but only at the beginning: "*I think it's really important that you have that at the first time to get the dot. But (then)) you will know how it works* (P5)."; "*I thought initially it ... helped me adapting to the rotation of the robot. But then after I did the second (round), I think it sort of distracted me* (P11)."

7 Discussion of Study 2

While we did not find significant difference in movement time between the two conditions, we can observe interesting trends from the data. Figure 11 (right) presents the performance results in each user group who completed the task in one of two orders of the two Visualisation conditions. While we can observe that AR Arm yielded similar performance regardless of the order, Figure 11 (right) indicates a trend where No Arm yielded observably longer Movement Time when it was the first condition than when it was administered after the AR Arm condition. This suggests that while individual differences and learning effects are important factors in the data, experiencing the AR Arm condition first may have helped participants learn the control better, which is reflected in their performance in subsequent trials without the AR visual feedback.

This observation of the data is consistent with the subjective feedback from the questionnaires. While a total of 21 participants agreed that the AR Arm was helpful for learning the control, six of them commented that it was only helpful at the beginning, while reporting that the AR Arm was either redundant or distracting after they had mastered the control with its earlier help. These results suggests that the AR visualisation may serve as a useful learning aid for notice users to understand how the teleoperation control is done through the mappings between joint rotation of their arms and of the robot, which is visualised through the mediation of the AR Arm between the two physical arms.

The visual reference of the human-like kinematic structure of the AR Arm made the task perceivably easier for participants, as supported by its significantly lower scores in Physical Demand, Effort, and in Frustration. While the physical effort demanded by the task was the same across the two conditions, participants perceived less Physical Demand and Effort with AR Arm likely because they did not need to move their arms blindly toward different directions to test the control at the beginning. The visual reference from the virtual arm enables them to quickly grasp how the rotations of their arm joints are mapped to the robot by observing the movement of the humanoid AR Arm, which visualises the same structure as their own arms and is rendered as overlaying on the physical robot arm. Similarly, participants experienced significantly less Frustration because the AR Arm provides direct and intuitive visual feedback for their arm movements, saving mental effort of actively interpreting how different parts of the robot responded to the rotation of different joints of their own arms.

Overall, to address **RQ2**, our findings suggests that the overlay of the human-like AR arm on the robot arm facilitates teleoperation and results in better user experience and less perceived effort. However, the performance benefit, reflected in time taken to complete the teleoperation task, may diminish over time once users become familiar with the control.

8 Summary of Results from Study 1 and 2

From Study 1, we learned that for MoCap-based teleoperation tasks that require human operators to position the end-effector of the robot arm at target positions, AR visualisation that aids the task should not be away from the robot and the target to be effective. Given this reason, along with user preferences for seeing a human-like virtual arm as visual reference to help them understand the mapping between the movements of their own arms and that of the robot, a **virtual humanoid arm overlaying the physical robot is the most likely effective configuration for AR visualisation to help with MoCap-based teleoperation**. The evaluation of this visualisation in Study 2 indicated that the **AR humanoid arm overlay on the robot arm improves user experience of MoCap-based teleoperation** by providing an intuitive visual feedback that mediates the otherwise confusing mapping between the movement of two arms with different visual structures and in different orientations.

Observations from the performance data and subjective feedback from participants consistently suggest that the AR overlay of humanoid arm on the robot arm exhibits potential to assist users learn its control at the beginning of teleoperation tasks, but may become redundant after users become familiar with the control. Based on these results, we propose that the AR overlay may help reduce the cognitive load of users for developing the understanding of the mapping between the kinematics of the robot arm and of their own arms. This is possible thanks to the anthropomorphic structure visualised through the AR overlay on the robot arm that makes the control mapping easier to interpret and understand, as suggested by previous work [15, 29, 50]. Other potential benefits of seeing anthropomorphic visual feedback through AR is that users may develop a better sense of embodiment (ownership and agency) over the robot through the mediation of the AR overlay, which may contribute to the easier learning of control [17, 23]. However, as the results suggest, these benefit may only take effect for learning the control, but not reflected in the performance results any longer after the users master the control with the help of the AR visualisation. In sum, we propose that **anthropomorphic AR overlay aids the comprehension of control-mapping in MoCap-based robot teleoperation, but not long-term performance.**

9 Limitations and Future Work

The infrastructure of the studies faced technical limitations. For instance, the movement speed of the robot arm was limited by the refresh rate of the communication channel that sent MoCap data to the robot, and by the hardware limitation of the robot itself. The movement range of the robot arm was also limited to prevent it from going into extreme postures (e.g., singularity positions), which limits the set of postures that we could use in the study. The perceived technical issues by the two participants (Table 2) regarding spatial accuracy of the AR overlay and the latency of robot movement could potentially be alleviated by a more capable technical setup. If more advanced technologies are accessible, future work could explore how latency between the MoCap system and the robot-control module affects the efficacy of AR-mediated teleoperation performance. In this work, we employed the state-of-the-art 7DOF robot arm and AR headset. In particular, 7-DoF robotic arms still remain critical for industrial and medical applications due to their dynamic stability and kinematic redundancy, which allows for easier obstacle avoidance, for instance, through “null-space" elbow movements. However, future work could also validate our findings using other devices and robotic platforms, such as humanoids like the Unitree G1, to investigate the relationship between robot morphology and teleoperation intuitiveness.

Though subjective feedback from Study 1 indicated that displaying the AR visual feedback away from the physical robot may cause distraction and hinder its effectiveness in helping learning the control, future work could investigate more variations in the positioning of the AR visualisations relative to the physical robot. While we cover an end-effector manipulation task and a posture control task, future works could explore the effect of AR visualisations in specific use contexts of teleoperation that can directly benefit the application domain, such as home care [28] and surgery [27].

Regarding performance results, we observed that individual variance of participants' capabilities to learn and perform teleoperation may have played a factor in their movement time. While the results suggested that the AR visualization may be helpful for learning at early stages of the task, future work could specifically investigate the effectiveness of the AR arm to help learn MoCap-based teleoperation control by comparing transfers from different AR visualisations. Moreover, while limited the number of trials per condition to preserve participant comfort and data quality during a physically demanding task, we acknowledge that the relatively low number of repetitions may limit the ability to observe long-term learning effects.

Finally, based on our findings, we propose that future work could explore **adaptive AR visualisation approaches** that use anthropomorphic overlays to aid the comprehension of the alignment between human- and robot- kinematics **at the beginning** of teleoperation tasks, and make them fade away as the users exhibit mastery over the control.

10 Conclusion

Robot arms present a unique affordance of behaving like human arms, thanks to their similar structures of movement ranges. This unique feature suggests an interesting opportunity for a novel approach of teleoperating robot arms through MoCap control that maps the rotations of the joints of the human operator's arms to the rotations of the corresponding joints on a robot arm. In this work, we explore how AR can assist this task by rendering a virtual arm as visual reference that mediates the inconsistencies between the human arm and the robot arm. In Study 1, we concluded that the optimal configuration of the AR visualisation is a humanoid arm overlaid on the physical robot arm in the same orientation with it to assist the understanding of the control mapping and to ensure easy visual access and attention. In Study 2, we evaluated the AR arm in this configuration and found that it helped reduce the perceived physical demand, effort, and frustration. Most participants found the AR arm helpful in learning the control, while considering it more suitable as a learning tool rather than an always-on visual guidance for teleoperation tasks. Finally, we summarise key findings that directly inform future work that explore MoCap-based teleoperation of robot arms, along with the documentation of a system integrating ROS, MoCap, and AR components.

References

[1] Bárbara Álvarez, Andrés Iborra, Alejandro Alonso, and Juan Antonio De la Puente. 2001. Reference Architecture for Robot Teleoperation: Development Details and Practical Use. *Cont. Eng. Practice* 9, 4 (2001), 395–402.

[2] Olaf Blanke, Theodor Landis, Laurent Spinelli, and Margitta Seeck. 2004. Out-of-body experience and autoscopy of neurological origin. *Brain* 127, 2 (02 2004), 243–258. https://doi.org/10.1093/brain/awh040

[3] Filippo Brizzi, Lorenzo Peppoloni, Alessandro Graziano, Erika Di Stefano, Carlo Alberto Avizzano, and Emanuele Ruffaldi. 2018. Effects of Augmented Reality on the Performance of Teleoperated Industrial Assembly Tasks in a Robotic Embodiment. *IEEE Transactions on Human-Machine Systems* 48, 2 (2018), 197–206. https://doi.org/10.1109/THMS.2017.2782490

[4] Linfeng Chen, Kazuki Takashima, Kazuyuki Fujita, and Yoshifumi Kitamura. 2021. PinpointFly: An Egocentric Position-control Drone Interface using Mobile AR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 150, 13 pages. https://doi.org/10.1145/3411764.3445110

[5] Alexandra Covaci, Anne-Hélène Olivier, and Franck Multon. 2014. Third person view and guidance for more natural motor behaviour in immersive basketball playing. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology (VRST '14)*. Association for Computing Machinery, New York, NY, USA, 55–64. https://doi.org/10.1145/2671015.2671023

[6] Alexandra Covaci, Anne-Hélène Olivier, and Franck Multon. 2015. Visual Perspective and Feedback Guidance for VR Free-Throw Training. *IEEE Computer Graphics and Applications* 35, 5 (Sept. 2015), 55–65. https://doi.org/10.1109/MCG.2015.95 Conference Name: IEEE Computer Graphics and Applications.

[7] Jianhong Cui, Sabri Tosunoglu, Rodney Roberts, Carl Moore, and Daniel W Repperger. 2003. A Review of Teleoperation System Control. In *Proc. Florida Conf. Recent Adv. Robot.* Citeseer, 1–12.

[8] Anton Franzluebbers and Kyle Johnson. 2019. Remote Robotic Arm Teleoperation through Virtual Reality. In *Symposium on Spatial User Interaction* (New Orleans, LA, USA) *(SUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 27, 2 pages. https://doi.org/10.1145/3357251.3359444

[9] Niklas Funk, Charles Schaff, Rishabh Madan, Takuma Yoneda, Julen Urain De Jesus, Joe Watson, Ethan K Gordon, Felix Widmaier, Stefan Bauer, Siddhartha S Srinivasa, et al. 2021. Benchmarking structured policies and policy optimization for real-world dexterous object manipulation. *IEEE Robot. Autom. Lett.* 7, 1 (2021), 478–485.

[10] Uwe Gruenefeld, Lars Prädel, Jannike Illing, Tim Stratmann, Sandra Drolshagen, and Max Pfingsthorn. 2020. Mind the ARm: realtime visualization of robot motion intent in head-mounted augmented reality. In *Proceedings of Mensch Und Computer 2020* (Magdeburg, Germany) *(MuC '20)*. Association for Computing Machinery, New York, NY, USA, 259–266. https://doi.org/10.1145/3404983.3405509

[11] Gianpaolo Gulletta, Eliana Costa e Silva, Wolfram Erlhagen, Ruud Meulenbroek, Maria Fernanda Pires Costa, and Estela Bicho. 2021. A Human-like Upper-limb Motion Planner: Generating naturalistic movements for humanoid robots. *International Journal of Advanced Robotic Systems* 18, 2 (2021), 1729881421998585. https://doi.org/10.1177/1729881421998585 arXiv:https://doi.org/10.1177/1729881421998585

[12] Gianpaolo Gulletta, Eliana Costa e Silva, Wolfram Erlhagen, Ruud Meulenbroek, Maria Fernanda Pires Costa, and Estela Bicho. 2021. A Human-like Upper-limb Motion Planner: Generating naturalistic movements for humanoid robots. *International Journal of Advanced Robotic Systems* 18, 2 (2021), 1729881421998585.

[13] Sandra G Hart. 2006. Development of NASA TLX: Result of Empirical and Theoretical Research. *California: San Jose State University* (2006).

[14] Hooman Hedayati, Michael Walker, and Daniel Szafir. 2018. Improving Collocated Robot Teleoperation with Augmented Reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. Association for Computing Machinery, New York, NY, USA, 78–86. https://doi.org/10.1145/3171221.3171251

[15] Frank Hegel, Soren Krach, Tilo Kircher, Britta Wrede, and Gerhard Sagerer. 2008. Understanding social robots: A user study on anthropomorphism. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*. 574–579. https://doi.org/10.1109/ROMAN.2008.4600728

[16] Victoria Hodge and Jim Austin. 2004. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review* 22 (2004), 85–126.

[17] Matthias Hoppe, Andrea Baumann, Patrick Chofor Tamunjoh, Tonja-Katrin Machulla, Paweł W. Woźniak, Albrecht Schmidt, and Robin Welsch. 2022. There Is No First- or Third-Person View in Virtual Reality: Understanding the Perspective Continuum. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3491102.3517447

[18] Mohammed Irfan Kurpath, Pradeep Kumar Adwai, Janardhan Bodireddy, Chandrasekaran K, and Navin Sam K. 2024. An IMUs and potentiometer-based controller for robotic arm-hand teleoperation. *Sensors and Actuators A: Physical* 367 (2024), 115019. https://doi.org/10.1016/j.sna.2024.115019

[19] Donghyeon Lee and Young Soo Park. 2018. Implementation of Augmented Teleoperation System Based on Robot Operating System (ROS). In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 5497–5502. https://doi.org/10.1109/IROS.2018.8594482

[20] Yuan Lin, Shuang Song, and Max Q.-H. Meng. 2016. The implementation of augmented reality in a robotic teleoperation system. In *2016 IEEE International Conference on Real-time Computing and Robotics (RCAR)*. 134–139. https://doi.org/10.1109/RCAR.2016.7784014

[21] Zhengxiong Liu, Zhenyu Lu, Yang Yang, and Panfeng Huang. 2019. Teleoperation for space manipulator based on complex virtual fixtures. *Robotics and Autonomous Systems* 121 (2019), 103268.

[22] Zhanat Makhataeva and Huseyin Atakan Varol. 2020. Augmented Reality for Robotics: A Review. *Robotics* 9, 2 (2020). https://doi.org/10.3390/robotics9020021

[23] Daniel Medeiros, Rafael K. dos Anjos, Daniel Mendes, João Madeiras Pereira, Alberto Raposo, and Joaquim Jorge. 2018. Keep my head on my shoulders! why third-person is bad for navigation in VR. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3281505.3281511

[24] M. Ostanin and A. Klimchik. 2018. Interactive Robot Programing Using Mixed Reality. *IFAC-PapersOnLine* 51, 22 (2018), 50–55. https://doi.org/10.1016/j.ifacol.2018.11.517 12th IFAC Symposium on Robot Control SYROCO 2018.

[25] Ornnalin Phaijit, Mohammad Obaid, Claude Sammut, and Wafa Johal. 2022. A Taxonomy of Functional Augmented Reality for Human-Robot Interaction. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 294–303. https://doi.org/10.1109/HRI53351.2022.9889622

[26] Carsten Preusche. 2008. Telerobotics. In *Springer handbook of robotics*. Springer, 741–757.

[27] Long Qian, Anton Deguet, and Peter Kazanzides. 2018. ARssist: augmented reality on a head-mounted display for the first assistant in robotic surgery. *Healthcare technology letters* 5, 5 (2018), 194–200.

[28] Daniel Rakita, Bilge Mutlu, and Michael Gleicher. 2017. A Motion Retargeting Method for Effective Mimicry-based Teleoperation of Robot Arms. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 361–370. https://doi.org/10.1145/2909824.3020254

[29] E. Roesler, D. Manzey, and L. Onnasch. 2021. A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science Robotics* 6, 58 (2021), eabj5425. https://doi.org/10.1126/scirobotics.abj5425 arXiv:https://www.science.org/doi/pdf/10.1126/scirobotics.abj5425

[30] Eric Rosen, David Whitney, Elizabeth Phillips, Gary Chien, James Tompkin, George Konidaris, and Stefanie Tellex. 2019. Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays. *The International Journal of Robotics Research* 38, 12-13 (2019), 1513–1526.

[31] Louis B Rosenberg. 1992. The use of virtual fixtures as perceptual overlays to enhance operator performance in remote environments. *Air force material command* (1992), 1–42.

[32] Louis B Rosenberg. 1993. The use of virtual fixtures to enhance operator performance in time delayed teleoperation. (1993).

[33] Louis B Rosenberg. 1993. Virtual fixtures: Perceptual tools for telerobotic manipulation. In *Proceedings of IEEE virtual reality annual international symposium*. Ieee, 76–82.

[34] Patrick Salamin, Daniel Thalmann, and Frédéric Vexo. 2006. The Benefits of Third-Person Perspective in Virtual and Augmented Reality? Association for Computing Machinery, 27–30. https://doi.org/10.1145/1180495.1180502

[35] Jonas Schjerlund, Kasper Hornbæk, and Joanna Bergström. 2021. Ninja Hands: Using Many Hands to Improve Target Selection in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 130, 14 pages. https://doi.org/10.1145/3411764.3445759

[36] Noriyoshi Shimizu, Maki Sugimoto, Dairoku Sekiguchi, Shoichi Hasegawa, and Masahiko Inami. 2008. Mixed reality robotic user interface: virtual kinematics to enhance robot motion. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology* (Yokohama, Japan) *(ACE '08)*. Association for Computing Machinery, New York, NY, USA, 166–169. https://doi.org/10.1145/1501750.1501789

[37] Alejandro Suarez, Victor M Vega, Manuel Fernandez, Guillermo Heredia, and Anibal Ollero. 2020. Benchmarks for aerial manipulation. *IEEE Robot. Autom. Lett.* 5, 2 (2020), 2650–2657.

[38] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 553, 33 pages. https://doi.org/10.1145/3491102.3517719

[39] Santawat Thanyadit and Ting-Chuen Pong. 2017. Desktop VR using a Mirror Metaphor for Natural User Interface. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17)*. Association for Computing Machinery, New York, NY, USA, 92–99. https://doi.org/10.1145/3132272.3134131

[40] David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27, 2 (2006), 237–246. https://doi.org/10.1177/1098214005283748

[41] Yuki Ueyama and Masanori Harada. 2022. Effects of first- and third-person perspectives created using a head-mounted display on dart-throwing accuracy. *Virtual Reality* 26, 2 (June 2022), 687–695. https://doi.org/10.1007/s10055-021-00562-x

[42] Alberto Villani, T Lisini Baldi, Nicole D'Aurizio, Giulio Campagna, and Domenico Prattichizzo. 2024. Does Robot Anthropomorphism Improve Performance and User Experience in Teleoperation?. In *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)*.

[43] Uta Wagner, Mathias N Lystbæk, Pavel Manakhov, Jens Emil Sloth Grønbæk, Ken Pfeuffer, and Hans Gellersen. 2023. A fitts' law study of gaze-hand alignment for selection in 3d user interfaces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[44] Michael Walker, Thao Phung, Tathagata Chakraborti, Tom Williams, and Daniel Szafir. 2023. Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy. *ACM Transactions on Human-Robot Interaction* 12, 4 (2023), 1–39.

[45] Michael E. Walker, Hooman Hedayati, and Daniel Szafir. 2019. Robot Teleoperation with Augmented Reality Virtual Surrogates. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 202–210. https://doi.org/10.1109/HRI.2019.8673306

[46] Xiaolong Yang, Furong Chen, Feilong Wang, Long Zheng, Shukun Wang, Wen Qi, and Hang Su. 2023. Sensor Fusion-Based Teleoperation Control of Anthropomorphic Robotic Arm. *Biomimetics* 8, 2 (2023), 169.

[47] AWW Yew, SK Ong, and AYC Nee. 2017. Immersive augmented reality environment for the teleoperation of maintenance robots. *Procedia Cirp* 61 (2017), 305–310.

[48] Difeng Yu, Qiushi Zhou, Benjamin Tag, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2020. Engaging Participants during Selection Studies in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 500–509. https://doi.org/10.1109/VR46266.2020.00071

[49] Qiushi Zhou, Brandon Victor Syiem, Beier Li, Jorge Goncalves, and Eduardo Velloso. 2024. Reflected Reality: Augmented Reality through the Mirror. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 202 (Jan. 2024), 28 pages. https://doi.org/10.1145/3631431

[50] Jakub Złotowski, Diane Proudfoot, Kumar Yogeeswaran, and Christoph Bartneck. 2015. Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *International Journal of Social Robotics* 7, 3 (June 2015), 347–360. https://doi.org/10.1007/s12369-014-0267-6