Assisting MoCap-Based Teleoperation of Robot Arm using Augmented Reality Visualisations

Qiushi Zhou
Aarhus University
0000-0001-6880-6594

Antony Chacon
The University of Melbourne
0000-0001-5888-1646

Jiahe Pan
ETH Zurich
0000-0003-2272-0925

Wafa Johal
The University of Melbourne
0000-0001-9118-0454

***Abstract*—Teleoperating a robot arm involves the human operator positioning the robot's end-effector or programming each joint. Whereas humans can control their own arms easily by integrating visual and proprioceptive feedback, it is challenging to control an external robot arm in the same way, due to its inconsistent orientation and appearance. We explore teleoperating a robot arm through motion-capture (MoCap) of the human operator's arm with the assistance of augmented reality (AR) visualisations. We investigate how AR helps teleoperation by visualising a virtual reference of the human arm alongside the robot arm to help users understand the movement mapping. We found that the AR overlay of a humanoid arm on the robot in the same orientation helped users learn the control. We discuss findings and future work on MoCap-based robot teleoperation.**

***Index Terms*—teleoperation; robot arm; augmented reality**

I. Introduction

Robot arm is a popular type of robot for a wide range of general-purpose tasks thanks to their efficient form-factors. They are typically situated on desktops in a vertical orientation. A similar movement range of the human arm is the space on one side of the body (Figure 2), perpendicular to the orientation of the movement range of the robot arm. This difference in **orientation** presents a challenge for Motion Capture (MoCap) based teleoperation for human operators to anticipate the movement of the robot. Further, the difference in **appearances** of the human and robot arms challenges operators to understand how the joint rotations are mapped.

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. 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 task. We implemented three conditions of the AR arm that are either in a human-like or robotic appearance, and either in the same orientation with the robot or with the human arm. We concluded that the optimal configuration is a human-like arm overlaid on the physical robot in the same orientation with it to assist the understanding of the control mapping and to ensure easy visual access. In Study 2, we evaluated this AR arm using a posture matching task. We found that the AR arm helped reduce the perceived physical demand, effort, and frustration. Most participants found the AR arm more helpful for learning the control at the beginning than as an always-on visual guidance for teleoperation.

This research is partially supported by the Australian Research Council Discovery Early Career Research Award (Grant No. DE210100858).

II. Background

A. Robot Teleoperation via Human Motion Mapping

Robot teleoperation enables human operators to remotely control a robot to perform tasks that are difficult for the human body through manipulation interfaces [1]. It requires extensive training, especially for robots with high DOF that challenge operators to anticipate the robot movement [2]. Anthropomorphic robots afford the unique possibility to perform human-like movements for tasks that demand anthropomorphic appearances of movements, inspiring works such as [3] that generates human-like movement for human-like robot arms. While previous work found that anthropomorphism help users perceive robot actions [4]–[6], no previous work has explored robot teleoperation control that capitalise on the ease of understanding anthropomorphic robotic movements.

B. AR Visualisations and Robot Control

AR has been used for HRI for different purposes [7], such as to support real-time control and teleoperation (for reviews, see [8], [9]). 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 [10]. 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 [11], allowing users to understand the motion intent more easily [12]. Whereas previous work explored AR-enabled robot control, such as using virtual shadows on physical floors to position drones [13] and interactive robot programming with virtual movement cues and anchors rendered in the physical environment [14], it remains to be investigated how AR can help teleoperation by visualising interactive anthropomorphic movement cues for operators to learn the control using their own body movement. We explore how a virtual arm next to or overlaid on a physical robot arm can help with MoCap-based teleoperation.

III. MoCap-Based AR Teleoperation System

We explore how a virtual arm visualised in AR can help users learn the control while mitigating inconsistencies between the robot and the human arms regarding their *Orientations* and *Appearances* in Study 1, and determine the optimal

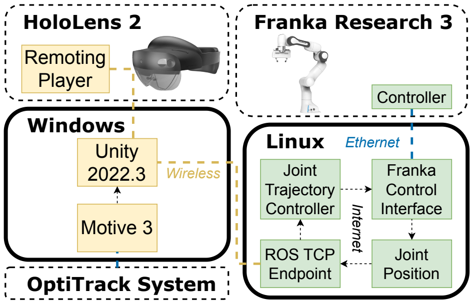

Fig. 1: System architecture: a Linux machine running ROS 2 and the Franka Control Interface, a Windows machine running Unity and Motive, a Franka Research 3 robot, and a HoloLens 2 that renders through a remoting player.

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.

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 to enable teleoperation of the FR3 by capturing the operator's right arm movements. The OptiTrack 8-camera system captures the arm's orientation, while HoloLens 2 renders 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.

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). Adapting this mapping to other manipulators with a similar configuration 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 illustrates 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

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

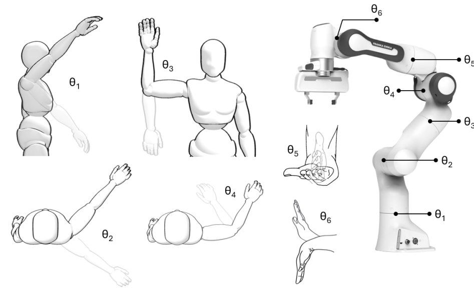

Fig. 2: Corresponding joint angles of human and robot arms.

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

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. Motive 3 transmits these data to Unity at an average speed of 200 KB/s. We used the HoloLens 2 and the Holographic Remoting Player app connected 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 (https://github.com/realitycollective) to recognise the QR code via the HoloLens 2 cameras and to log its position.

IV. 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). A challenge of teleoperating these robots 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 appearances between the human and the robot arm challenge the user to understand how the robot arm would respond to their control. We employ a within-subject design with four Visualisation conditions (Figure 3) to compare and determine the optimal configuration: HumanHorizontal (HH): A virtual human-like arm in the same orientation as the human arm. HumanVertical (HV): A virtual human-like arm in the same orientation as the physical robot arm. RobotHorizontal (RH): Virtual

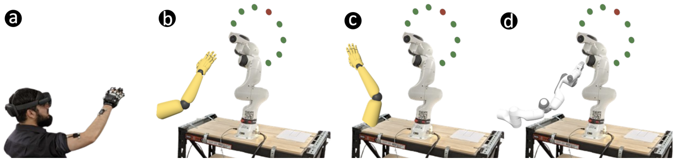

Fig. 3: AR VISUALISATION conditions and apparatus: (a) Participant situated away from the robot; (b) HUMANHORIZONTAL arm in AR next to the physical robot; (c) HUMANVERTICAL; (d) ROBOTHORIZONTAL.

replica of robot arm in the same orientation as the human arm. BASELINE: No AR visualisation rendered (Figure 3).

The studies 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 relative to the robot. The participant sat 1 m to the left and 1.2 m behind the robot's base.

The task was reaching for a ring of 11 virtual circular targets with the robot's end-effector in a pre-defined sequence, as illustrated in Figure 3 (e) [15]–[18]. The ring of 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 base, which was determined through pilot testing to allow the user's arm to move within a comfortable range. The active target (red) 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.

We recruited 24 participants (11 F, 13 M) with a mean age of 26.9 years ($Min = 19, Max = 36, SD = 4.5$) using the University's online billboard. Participants rated their prior experience with VR and AR on a 7-point scale from 1 (never used) to 7 (use frequently) with a mean rating of 2.13 ($Min = 1, Max = 6, SD = 1.45$). We calibrated HoloLens' built-in eye-tracker for each participant, and placed OptiTrack markers on their right arms. We used a 2-minute training block for participants to practice teleoperating the robot under the BASELINE. Then, participants completed four rounds (two trials per condition) of tasks with breaks between trials. The condition orders were counterbalanced using a Latin square. After each condition, participants filled out a questionnaire on their perception of the task and system. After all tasks, participants ranked their preferences for the conditions.

A. Results

Performance was evaluated through *Movement Time*, the time interval between the appearance of the target and the successful selection. We administered a NASA-TLX questionnaire [19] at the end of each 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?*

While we did not find statistically significant results in Movement Time or in NASA-TLX, user rankings of different VISUALISATION conditions showed that 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 ($M = 3.00, SD = 1.02$).

Twelve participants reported that the AR visualisation was helpful because it 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 (P22), especially in how far the robot moves into distance (P15). Five reported that their attention was drawn to the targets and the robot instead of using the visualisation.

Thirteen participants reported that the vertical orientation of the AR arm was more helpful because it was consistent with the physical robot, making it easier to understand the control. Seven participants preferred the horizontal orientation as it represents how their own arm behaves. Fifteen participants thought the HUMAN arm was more helpful because it resembled their own arm close to the robot for visual reference. Nine of them deemed HUMANVERTICAL most helpful.

B. Discussion

While VISUALISATION did not yield statistically significant differences in movement time, qualitative feedback suggested that participants still found the virtual arm helpful. While the vertical orientation and the human-like appearance of the virtual arm were most preferred, it is also the only configuration of AR that can be overlaid on the physical robot for easier visual access. We choose HV as the optimal configuration.

V. STUDY 2: POSTURE CONTROL

In Study 2, we investigate how an AR overlay of a human-like arm on the physical robot arm can assist MoCap-based teleoperation. We employ a posture matching task to avoid participants adapting to the repetitive movements as in target reaching, while requiring them to understand the mapping between their own arms and the robot. We employ a within-subject design that compares task performance and subjective measures between two VISUALISATION conditions: AR ARM where the virtual arm is presented during the tasks, and NO ARM as a baseline without virtual arm.

The task was sequentially matching postures using the real robot arm (Figure 5). We rendered a blue sphere on the robot elbow and a red sphere on the wrist for visual reference. A target posture is matched when the elbow and wrist of the robot are both within 5 cm from their respective target positions (light blue and red). We selected 4 target postures (Figure 4) within a comfortable range of movement. The task was designed to model real-world scenarios where joint-space

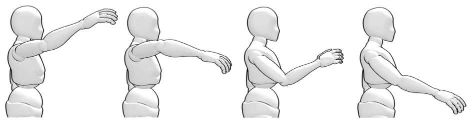

Fig. 4: S2 postures: a) Elbow up, wrist up; b) Elbow up, wrist down; c) Elbow down, wrist up; d) Elbow down, wrist down.

Fig. 5: Study 2: participant, condition NO ARM and AR ARM. Blue and red sphere rendered on elbow and wrist respectively. Lighter blue and red denote targets to match posture.

control of the robot is required to avoid collisions during teleoperation, which may arise in highly cluttered workspaces.

We recruited 24 participants (14F, 10M) with a mean age of 25.6 years ($SD = 4.6$), mean rating of prior experience with VR and AR (1-7) of 2.04 ($SD = 1.27$). Participants completed 2 rounds of the task under each condition. Each trial always started with the robot arm in a "straight" posture with full extension. The order of postures were counterbalanced using a Latin square. Participants filled out a questionnaire after each condition, and received an interview after all conditions for feedback on their perception of the task and visualisation.

A. Results

Movement Time is defined as the time taken to successfully match each target posture from when they first appear. We administered NASA-TLX [19] 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?*

We found no statistically significant effects of *Visualisation* or *viz order* on *Movement Time* (Figure 6). The perceived PHYSICAL DEMAND was significantly affected by *Visualisation* ($F_{1,23} = 5.20, p < .05$), similarly for EFFORT ($F_{1,23} = 5.06, p < .05$) and FRUSTRATION ($F_{1,23} = 5.14, p < .05$).

Thirteen participants found that the AR visualisation helpful "*It helped me with the position of the arm to have a reference. Without that, I couldn't imagine the best position* (P2)." Six participants mentioned that the visualisation helped them learn the control at the beginning while the task becomes easy after a few rounds, and they did not need the visualisation any more.

Fifteen participants found the AR visualisation helpful to learn the control mapping: "*It was very intuitive. I didn't have to learn anything beforehand, just having that reference was enough for the task* (P2)."; " *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. But then you will know how it works* (P5)."

B. Discussion of Study 2

In Figure 6 (right), we can observe that while the earlier condition always yielded longer movement time, the difference

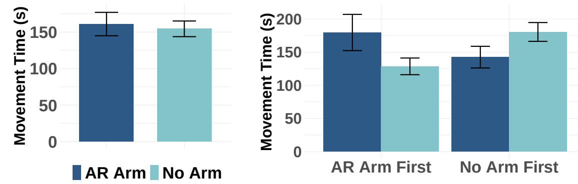

Fig. 6: Study 2 results: movement time by VISUALISATION (left), and separate plots by condition order (right).

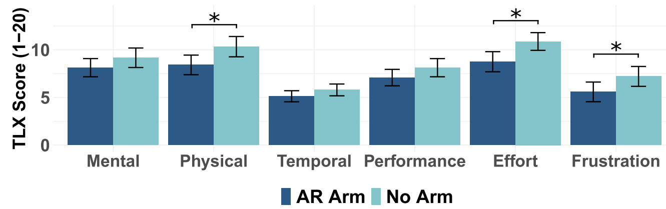

Fig. 7: Study 2 results on the NASA-TLX questionnaire.

seems more pronounced when AR ARM was administered first. This suggests that while individual differences and learning effects are present, experiencing the AR ARM first may have helped participants learn the control better. This observation is consistent with the subjective feedback. While 21 participants agreed that AR ARM was helpful for learning the control, six of them commented that it was only helpful at the beginning, and became redundant or distracting afterwards. These results suggests that the AR visualisation may serve as a useful learning aid for novice users to understand how the teleoperation control is done through the mappings between their arms and the robot, visually mediated by the AR ARM.

The visual reference of the human-like appearance of the AR ARM made the task perceivably easier for participants, as supported by its significantly lower scores in PHYSICAL DEMAND and EFFORT. This is likely because they did not need to move their arms blindly to test the control at the beginning. They were able to quickly grasp how the rotations of their arm joints are mapped to the robot by observing the movement of the human-like AR ARM, which visualises the same structure as their own arms and is rendered on the physical robot. Similarly, they experienced less FRUSTRATION because the AR ARM provides straightforward visual feedback for their movements, saving the mental conversion effort.

VI. CONCLUSION

Anthropomorphic robot arms suggests a novel approach of teleoperation through MoCap control that maps the rotations of the joints of the human operator's arms to a robot arm. We explore how AR can assist this by rendering a virtual arm as visual reference that mediates the inconsistencies between the human and the robot arm. In Study 1, we concluded that the optimal configuration of AR is a human-like arm overlaid on the physical robot in the same orientation. In Study 2, we evaluated this configuration and found that it helped reduce the perceived physical demand, effort, and frustration. Most participants found the AR arm suitable as a learning tool rather than an always-on visual guidance for teleoperation tasks.

REFERENCES

[1] R. Pepper and J. Hightower, "Research issues in teleoperator systems," in *Proceedings of the Human Factors Society Annual Meeting*, vol. 28, pp. 803–807, SAGE Publications Sage CA: Los Angeles, CA, 1984.

[2] H. Hedayati, M. Walker, and D. Szafir, "Improving collocated robot teleoperation with augmented reality," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18, (New York, NY, USA), p. 78–86, Association for Computing Machinery, 2018.

[3] G. Gulletta, E. C. e. Silva, W. Erlhagen, R. Meulenbroek, M. F. P. Costa, and E. Bicho, "A human-like upper-limb motion planner: Generating naturalistic movements for humanoid robots," *International Journal of Advanced Robotic Systems*, vol. 18, no. 2, p. 1729881421998585, 2021.

[4] F. Hegel, S. Krach, T. Kircher, B. Wrede, and G. Sagerer, "Understanding social robots: A user study on anthropomorphism," in *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 574–579, 2008.

[5] J. Złotowski, D. Proudfoot, K. Yogeeswaran, and C. Bartneck, "Anthropomorphism: Opportunities and challenges in human–robot interaction," *International Journal of Social Robotics*, vol. 7, pp. 347–360, June 2015.

[6] E. Roesler, D. Manzey, and L. Onnasch, "A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction," *Science Robotics*, vol. 6, no. 58, p. eabj5425, 2021.

[7] M. Walker, T. Phung, T. Chakraborti, T. Williams, and D. Szafir, "Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 4, pp. 1–39, 2023.

[8] Z. Makhataeva and H. A. Varol, "Augmented reality for robotics: A review," *Robotics*, vol. 9, no. 2, 2020.

[9] 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 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, (New York, NY, USA), Association for Computing Machinery, 2022.

[10] A. Yew, S. Ong, and A. Nee, "Immersive augmented reality environment for the teleoperation of maintenance robots," *Procedia Cirp*, vol. 61, pp. 305–310, 2017.

[11] 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*, vol. 38, no. 12-13, pp. 1513–1526, 2019.

[12] U. Gruenefeld, L. Prädel, J. Illing, T. Stratmann, S. Drolshagen, and M. Pfingsthorn, "Mind the arm: realtime visualization of robot motion intent in head-mounted augmented reality," in *Proceedings of Mensch Und Computer 2020*, MuC '20, (New York, NY, USA), p. 259–266, Association for Computing Machinery, 2020.

[13] L. Chen, K. Takashima, K. Fujita, and Y. Kitamura, "Pinpointfly: An egocentric position-control drone interface using mobile ar," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, (New York, NY, USA), Association for Computing Machinery, 2021.

[14] M. Ostanin and A. Klimchik, "Interactive robot programing using mixed reality," *IFAC-PapersOnLine*, vol. 51, no. 22, pp. 50–55, 2018. 12th IFAC Symposium on Robot Control SYROCO 2018.

[15] A. Suarez, V. M. Vega, M. Fernandez, G. Heredia, and A. Ollero, "Benchmarks for aerial manipulation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2650–2657, 2020.

[16] N. Funk, C. Schaff, R. Madan, T. Yoneda, J. U. De Jesus, J. Watson, E. K. Gordon, F. Widmaier, S. Bauer, S. S. Srinivasa, *et al.*, "Benchmarking structured policies and policy optimization for real-world dexterous object manipulation," *IEEE Robot. Autom. Lett.*, vol. 7, no. 1, pp. 478–485, 2021.

[17] D. Yu, Q. Zhou, B. Tag, T. Dingler, E. Velloso, and J. Goncalves, "Engaging participants during selection studies in virtual reality," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 500–509, 2020.

[18] Q. Zhou, B. V. Syiem, B. Li, J. Goncalves, and E. Velloso, "Reflected reality: Augmented reality through the mirror," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 7, Jan. 2024.

[19] S. G. Hart, "Development of NASA TLX: Result of empirical and theoretical research," *California: San Jose State University*, 2006.