Extended Nonverbal Expressions in Hybrid Robots with Physical and AR-Based Presentations

Ikkaku Kawaguchi
University of Tsukuba
Tsukuba, Japan
kawaguchi@cs.tsukuba.ac.jp

Keiichi Ihara
University of Colorado Boulder
University of Tsukuba
Boulder, USA

Yusuke Ashizawa
University of Tsukuba
Tsukuba, Japan

Shintaro Mori
University of Tsukuba
Tsukuba, Japan

Miki Hasegawa
University of Tsukuba
Tsukuba, Japan

Taito Ishii
University of Tsukuba
Tsukuba, Japan

Asahi Yamada
University of Tsukuba
Tsukuba, Japan

***Abstract*—Robotic embodiment and virtual embodiment have been proposed as methods for conveying nonverbal expressions. Despite their potential, methods using physical robots often encounter high implementation costs, while AR devices face limitations due to their restricted field of view (FoV), which results in the loss of peripheral information. This study introduces a hybrid robot system integrating physical robotic parts with optical see-through AR devices for AR-based presentations. The research defines extended nonverbal expressions to overcome robots' physical limitations, aiming to establish design guidelines for hybrid robots in remote communication. The study proposes a design space organizing hybrid robots' configurations and expressive capabilities based on communication purposes. A prototype system was developed to demonstrate the feasibility of incorporating extended nonverbal expressions into hybrid robot designs, showcasing potential real-world applications. This approach offers a promising solution to enhance nonverbal communication in remote settings, bridging the gap between physical and virtual interactions.**

***Index Terms*—component, formatting, style, styling, insert**

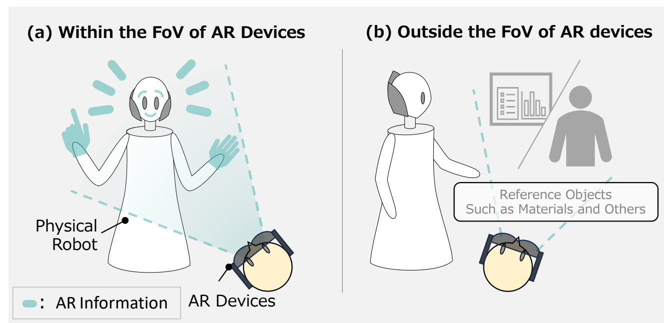

Fig. 1. Concept of the Hybrid Robot System. (a) Within the field of view (FOV) of AR devices, the physical robot's representation is augmented through AR display. (b) Outside the FOV of AR devices, the physical robot remains constantly visible to compensate for information loss in the peripheral vision.

I. Introduction

In face-to-face interactions, nonverbal information plays a crucial role in enabling smooth and meaningful communication. For example, gaze direction strongly influences the flow of conversation and turn-taking [1]–[3]. Likewise, gestures such as pointing are essential for establishing shared understanding when communication involves joint attention to physical objects [4]. However, conveying nonverbal information through 2D video is challenging. For instance, 2D video displays can lead to spatial misinterpretations, hindering the accurate transmission of nonverbal cues [5], [6]. Furthermore, screens reduce the richness of interpersonal awareness, making it more difficult to perceive subtle nonverbal signals [7].

To address these challenges, the use of physical robots, referred to as **robotic embodiment**, has been proposed as a potential solution [8]. For example, methods have been developed to convey the gaze and body orientation of remote users through the rotation or movement of displays [9]–[11]. Additionally, robots equipped with human-like physical features such as heads and arms can use these body parts to convey more sophisticated nonverbal cues [12]–[14]. However, implementing the physical components necessary for transmitting the desired nonverbal information involves significant costs in terms of system design, implementation, and maintenance.

As an alternative, methods using Augmented Reality (AR) devices like the Hololens to present nonverbal information through CG avatars, known as **virtual embodiment**, have also been proposed [15]–[17]. Unlike robots, CG-based nonverbal information presentation is not subject to physical constraints, allowing for relatively cost-effective implementation of advanced expressions. Furthermore, AR enables expressive extensions that transcend physical limitations, such as stretching an arm [18] or indicating gaze directions with arrows or bubbles [19]. However, current AR headsets are limited by their restricted field of view (FOV) [20], which can result in missing information in the peripheral vision.

This study addresses the trade-offs between robotic embodiment and virtual embodiment by integrating optical see-through AR devices, which preserve peripheral vision, with

This work was supported by JSPS KAKENHI Grant Number 24K20808.

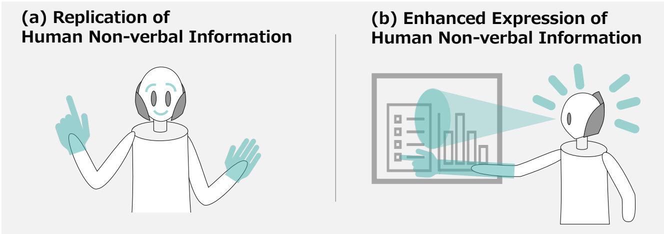

Fig. 2. Directions for Augmented Nonverbal Communication in Hybrid Robots. (a) Augment the physical robot through AR to replicate nonverbal cues. (b) Utilize AR to create enhanced expressions of nonverbal information.

physical robots. We define this integration for nonverbal communication as **hybrid embodiment** and propose a hybrid robot leveraging this concept (Fig. 1). Within the AR device's FOV, AR extends the physical robot's expressive capabilities (Fig. 1-a). Beyond the AR device's view, the physical robot's persistent presence compensates for peripheral vision limitations (Fig. 1-b).

To enhance robot expressions, we propose two approaches. The first approach supplements nonverbal cues that are difficult or expensive to implement physically, such as facial expressions or hand gestures, using AR (Fig. 2-a). The second approach surpasses physical constraints by enhancing nonverbal communication, such as extending arms during pointing gestures, visualizing gaze with rays, or displaying emotions through special effects (Fig. 2-b).

In this paper, we first review related research on robotic embodiment, virtual embodiment, and hybrid embodiment. Next, we explore the design space of hybrid robots, organizing potential configurations and expressions based on usage objectives. Finally, we describe a prototype system implementing part of the proposed design space and conclude with a discussion of future directions.

II. Related Work

A. Robotic Embodiment

As a method for non-verbal communication using robots, telepresence robots such as Kubi [21] enable remote participants to direct their gaze toward specific directions on-site by rotating a display equipped with a camera. This allows remote participants to convey references to local participants effectively. Such display rotation techniques can be combined with video feeds and are widely adopted in numerous systems [9]–[11], [22]–[25]. Research has shown that transmitting such references contributes to enhancing the sense of presence of remote participants and facilitates smoother conversations [10], [11]. However, challenges such as misinterpretation of gaze direction have been reported when rotating displays showing facial images [5], [6]. An alternative approach involves systems that utilize human-like body parts, such as heads or arms, to enhance communication [12], [13]. However, employing physical body parts entails high costs for system design, implementation, and maintenance. In this study, we explore the potential of hybrid robots, which could mitigate such costs while retaining effective communication capabilities.

B. Virtual Embodiment

In the context of non-verbal communication using Augmented Reality (AR), several approaches have been proposed. These include methods that present simplified representations of body parts such as the head or hands [26] and methods employing full-body CG avatars that replicate human appearances [17]. Additionally, rather than using pre-rendered CG avatars, techniques utilizing point clouds captured by depth cameras to convey the real-time appearance of users have also been explored [15], [16]. Despite these advances, the head-mounted displays (HMDs) commonly used in such systems have limitations, particularly in their field of view (FOV), which makes it challenging to fully replicate human vision [20]. To address these FOV constraints, methods such as representing avatars as smaller, scaled-down figures to fit within the user's view have been proposed [27], [28]. Moreover, compared to physical objects in real space, AR representations are also constrained by brightness and resolution. This study aims to mitigate these limitations of AR systems by integrating physical robots with AR, leveraging the strengths of both to enhance the embodiment experience. Some of the aforementioned systems [16], [17] combine physical robots with AR. However, in these systems, the embodiment is fully realized within AR. Therefore, we distinguish these systems from hybrid embodiment that combines both physical bodies and AR to enable the embodiment.

C. Hybrid Embodiment

Research on hybrid embodiment has explored various methods to enhance robotic expressiveness by combining physical and AR-based elements. For example, Groechel et al. proposed a technique to enhance the expressiveness of social robots without physical arms by adding AR-rendered arms [29]. Han et al. compared the effectiveness of deictic gestures performed by physical arms and AR-rendered arms in robots with mobile platforms [30]. Kawaguchi et al. developed a robot capable of switching between physical and AR representations for its head and arms [31].

While these studies demonstrated the potential of hybrid robots, they were limited in the scope of AR-presented information and did not explore extended forms of expression, as illustrated in Fig. 2-b. To address these limitations, this study investigates the design space of hybrid robots. We systematically organize potential configurations and expressive capabilities tailored to various use cases and present a prototype system implementing part of these configurations to showcase their feasibility.

III. Proposed System

A. Design Space

The design space for the proposed hybrid robot in this study is shown in Fig. 3. The vertical axis of the figure represents key aspects of nonverbal information presentation by the robot: **reference**, which is crucial for interactions with people and objects in the physical space; **appearance and expressions**, which play a significant role in conveying social presence and emotions; and **other considerations**. The horizontal axis represents the types of AR-based information presentation: **information augmentation**, **information refinement**, and **enhanced expression**.

Information augmentation aims to add information that does not physically exist by leveraging AR. Examples include adding a head or arms to displays that lack physical body parts to facilitate reference (Fig. 3-i) or enhancing simple robot designs by adding specific facial features (e.g., mouth or eyebrows) or fingers for expression (Fig. 3-ii). Information refinement focuses on improving the detail and realism of the presented information, bringing it closer to human-like representations. This could involve overlaying holograms using point clouds or pre-rendered realistic avatars (Fig. 3-iii). Enhanced expression takes advantage of AR's ability to transcend physical constraints, enabling unique visual representations. Examples include projecting rays along gaze or pointing directions, or simulating extended arms to indicate references (Fig. 3-iv), using effects like color-changing lights to convey emotions (Fig. 3-v), or displaying textual information or related images alongside conversational content (Fig. 3-vi). To comprehensively realize the capabilities suggested in this design space, we implemented a prototype system.

B. Prototype System

In this study, we implemented a prototype system by extending a previously developed system [31]. The configuration of the prototype system is shown in Fig. 4.

The system implemented in this study consists of a robot, HoloLens 2, an AR control PC, and a remote control PC. The external appearance of the robot is shown in Fig. 4-B. The robot comprises a head with two degrees of freedom (pan-tilt), arms with three degrees of freedom at the shoulder and one at the elbow per arm, and a base with a single degree of freedom (pan). Both the head and arms are detachable. Each joint is actuated by servo motors, which are controlled via serial communication from a Surface Go 3 (10.5-inch, Intel Core i3, 8GB RAM) mounted on the robot's body. The angles of the joints are updated based on data received from the remote control PC via UDP communication. The control program, written in Python, handles both the communication with the remote control PC and the serial communication with the motors and runs on the Surface Go 3. When operating the robot, the Surface Go 3 also runs a video calling application (e.g., Microsoft Teams [1]) for audio communication.

HoloLens 2 is used to overlay AR information onto the robot. The AR content is controlled by an AR control program running on the AR control PC and rendered on HoloLens 2 using Microsoft's Holographic Remoting Player.

On the remote control PC, a posture acquisition program and a facial expression acquisition program are executed to operate the robot remotely. Control information obtained by these programs is transmitted to the robot and the AR control PC via UDP communication. The posture acquisition program uses MediaPipe [2] to capture upper-body posture data and employs Final IK [3] to calculate joint angles for the robot's arms. The head posture and facial expression acquisition program uses Dlib [4] to extract head orientation and facial landmark information.

On the AR control PC, an AR control program interprets the control information received via UDP from the remote control PC and performs the rendering and control of AR elements, including the robot's body parts and augmented expressions. Currently, the system supports adding arms and a head to the robot, adding facial expressions and fingers, visualizing rays for pointing directions, and displaying emotion-related effects (Fig. 5). For AR-based augmentation of the robot's body parts, CG models of the robot's body parts are animated

[1] https://www.microsoft.com/en-us/microsoft-teams/group-chat-software
[2] https://github.com/google/mediapipe
[3] https://assetstore.unity.com/packages/tools/animation/final-ik-14290
[4] https://github.com/davisking/dlib

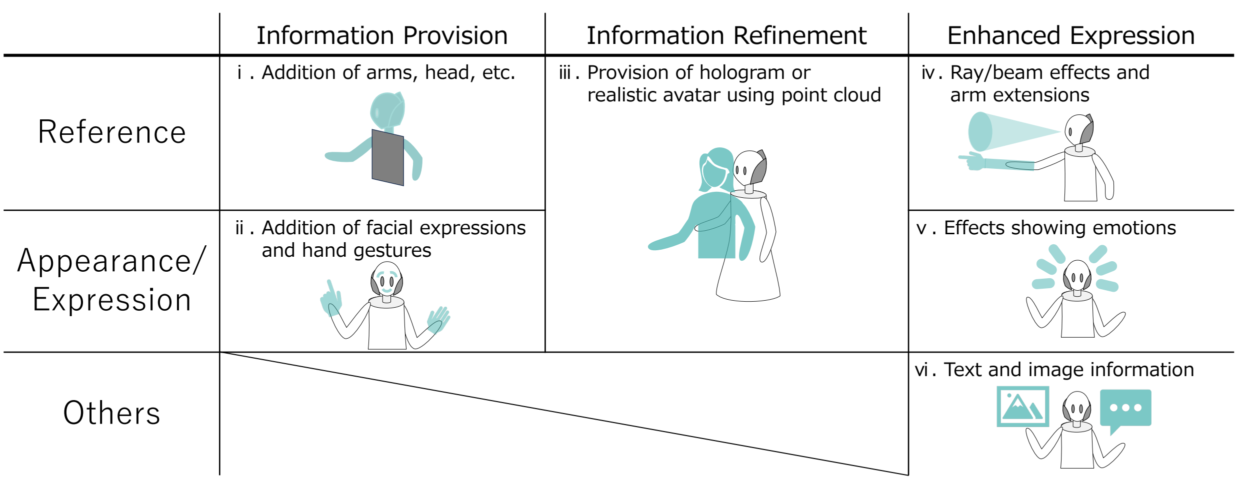

Fig. 3. Design space of the extended nonverbal expressions.

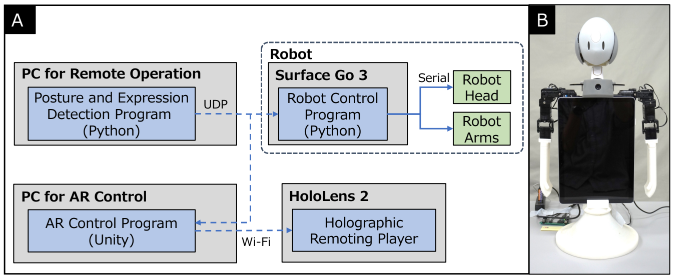

Fig. 4. System configuration. (A) Block diagram of our system, (B) Appearance of our robot.

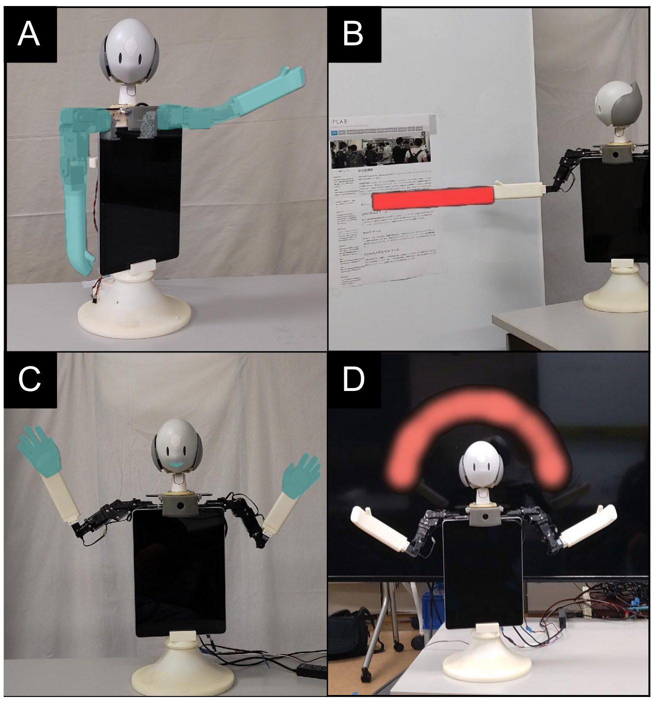

Fig. 5. Features of the prototype system: (A) Augmentation of body parts, (B) Ray visualization for pointing directions, (C) Facial expression and finger augmentation, (D) Visualization of emotions like 'happy' using effects.

based on the posture information received from the remote control PC (Fig. 5-A). For visualizing rays pointing in specific directions, Unity's LineRenderer component is used to display rays emanating from the robot's fingertips (Fig. 5-B). To augment facial expressions, the curvature of a Unity cube object representing a mouth is adjusted based on the received control data (Fig. 5-C). Emotion-related effects are implemented using Unity's Particle System component to display light effects that represent emotions like 'happy' (Fig. 5-D). While the system plans to incorporate automatic emotion recognition from facial expressions in the future, currently, the emotions displayed are manually controlled via keyboard input.

IV. Conclusion and Future Work

In this paper, we proposed extended nonverbal expressions for a hybrid robot that combines physical and AR-based presentations. We described the design space for these expressions as well as the current prototype system. As future work, we plan to implement features that have not yet been realized within the design space, such as holographic presentations using point clouds and the ability to display text or image information based on conversational context. Subsequently, we aim to evaluate the effectiveness of the proposed extended nonverbal expressions by conducting comparative experiments. These experiments will compare the hybrid approach to conditions where all information is presented via AR or where only the physical robot is used.

References

[1] A. Kendon, "Conducting interaction: Patterns of behavior in focused encounters". CUP Archive, 1990.
[2] G. S. Nielsen, "Studies in Self Confrontation: Viewing a Sound Motion Picture of Self and Another Person in a Stressful Dyadic Interaction". Munksgaard., 1964.
[3] H. Admoni, and B. Scassellati, "Social eye gaze in human-robot interaction: a review". Journal of Human-Robot Interaction, vol. 6, no. 1, pp. 25–63, 2017.
[4] S. R. Fussell, L. D. Setlock, J. Yang, J. Ou, E. Mauer, and A. D. I. Kramer, "Gestures over video streams to support remote collaboration on physical tasks". Human-Computer Interaction, vol. 19, no. 3, pp. 273–309, 2004.
[5] S. M. Anstis, J. W. Mayhew, and T. Morley, "The perception of where a face or television'portrait'is looking". The American journal of psychology, vol. 82, no. 4, pp. 474–489, 1969.
[6] I. Kawaguchi, H. Kuzuoka, and Y. Suzuki, "Study on gaze direction perception of face image displayed on rotatable flat display". in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1729–1737, 2015.
[7] C. Heath, and P. Luff, "Disembodied conduct: Communication through video in a multi-media office environment". in Proceedings of the SIGCHI conference on human factors in computing systems, pp. 99–103, 1991.
[8] E. Deng, B. Mutlu, M. J. Mataric, and others, "Embodiment in socially interactive robots". Foundations and Trends® in Robotics, vol. 7, no. 4, pp. 251–356, 2019.
[9] N. Yankelovich, N. Simpson, J. Kaplan, and J. Provino, "Porta-person: Telepresence for the connected conference room". in Proceedings of the CHI'07 extended abstracts on Human factors in computing systems, pp. 2789–2794, 2007.
[10] D. Sirkin, and W. Ju, "Consistency in physical and on-screen action improves perceptions of telepresence robots". in Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, pp. 57–64, 2012.
[11] J. T. Biehl, D. Avrahami, and A. Dunnigan, "Not really there: Understanding embodied communication affordances in team perception and participation". in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 1567–1575, 2015.
[12] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki, "Reconfiguring spatial formation arrangement by robot body orientation". in Proceedings of the 2010 5th ACM/IEEE international conference on human-robot interaction (HRI), pp. 285–292, 2010.
[13] S. O. Adalgeirsson, and C. Breazeal, "MeBot: A robotic platform for socially embodied telepresence". in Proceedings of the 2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 15–22, 2010.
[14] Y. Onishi, K. Tanaka, and H. Nakanishi, "PopArm: A robot arm for embodying video-mediated pointing behaviors". in Proceedings of the 2014 International Conference on Collaboration Technologies and Systems (CTS), pp. 137–141, 2014.
[15] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, and others, "Holoportation: Virtual 3d teleportation in real-time". in Proceedings of the 29th annual symposium on user interface software and technology, pp. 741–754, 2016.

[16] K. Ihara, M. Faridan, A. Ichikawa, I. Kawaguchi, and R. Suzuki, "HoloBots: Augmenting Holographic Telepresence with Mobile Robots for Tangible Remote Collaboration in Mixed Reality". in Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp. 1–12, 2023.

[17] B. Jones, Y. Zhang, P. N. Y. Wong, and S. Rintel, "Belonging there: VROOM-ing into the uncanny valley of XR telepresence". Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. CSCW1, pp. 1–31, 2021.

[18] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, "The go-go interaction technique: non-linear mapping for direct manipulation in VR". in Proceedings of the 9th annual ACM symposium on User interface software and technology, pp. 79–80, 1996.

[19] J. Ichino, M. Ide, T. Yoshiki, H. Yokoyama, H. Asano, H. Miyachi, and D. Okabe, "How gaze visualization facilitates initiation of informal communication in 3D virtual spaces". ACM Transactions on Computer-Human Interaction, vol. 31, no. 1, pp. 1–32, 2023.

[20] L.-H. Lee, T. Braud, S. Hosio, and P. Hui, "Towards augmented reality driven human-city interaction: Current research on mobile headsets and future challenges". ACM Computing Surveys (CSUR), vol. 54, no. 8, pp. 1–38, 2021.

[21] Xandex, Inc., "KUBI Telepresence Robot". 2024. Available: https://www.kubiconnect.com/, Accessed on January 23, 2025.

[22] E. Paulos, and J. Canny, "PRoP: Personal roving presence". in Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 296–303, 1998.

[23] N. T. Fitter, L. Rush, E. Cha, T. Groechel, M. J. Matarić, and L. Takayama, "Closeness is key over long distances: Effects of interpersonal closeness on telepresence experience". in Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction, pp. 499–507, 2020.

[24] A. Boudouraki, J. E. Fischer, S. Reeves, and S. Rintel, "" I can't get round" Recruiting Assistance in Mobile Robotic Telepresence". Proceedings of the ACM on Human-computer Interaction, vol. 4, no. CSCW3, pp. 1–21, 2021.

[25] M. Sakashita, E. A. Ricci, J. Arora, and F. Guimbretière, "RemoteCoDe: Robotic Embodiment for Enhancing Peripheral Awareness in Remote Collaboration Tasks". in Proceedings of the ACM on Human-Computer Interaction, vol. 6, no. CSCW1, pp. 1–22, 2022.

[26] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst, "A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing". in Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1–13, 2020.

[27] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, "Mini-me: An adaptive avatar for mixed reality remote collaboration". in Proceedings of the 2018 CHI conference on human factors in computing systems, pp. 1–13, 2018.

[28] T. Piumsomboon, G. A. Lee, A. Irlitti, B. Ens, B. H. Thomas, and M. Billinghurst, "On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction". in Proceedings of the 2019 CHI conference on human factors in computing systems, pp. 1–17, 2019.

[29] T. Groechel, Z. Shi, R. Pakkar, and M. J. Matarić, "Using socially expressive mixed reality arms for enhancing low-expressivity robots". in Proceedings of the 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), pp. 1–8, 2019.

[30] Z. Han, Y. Zhu, A. Phan, F. S. Garza, A. Castro, and T. Williams, "Crossing Reality: Comparing Physical and Virtual Robot Deixis". in Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2023.

[31] I. Kawaguchi, A. Ichikawa, K. Ihara, R. Ishibashi, and A. Tanokashira, "Hybrid Robot with Physical and AR Body Presentation". in Proceedings of the ICAT-EGVE 2022 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments - Posters and Demos, 2022.