That's Iconic! Designing Augmented Reality Iconic Gestures To Enhance Multi-modal Communication For Morphologically Limited Robots

Yifei Zhu, Alexander Torres, Zane Aloia, and Tom Williams

***Abstract*— Robots that use gestures in conjunction with speech can achieve more effective and natural communication with human teammates, however, not all robots have capable and dexterous arms. Augmented Reality technology has effectively enabled deictic gestures for morphologically limited robots in prior work, however, the design space of AR-facilitated iconic gestures remains under-explored. Moreover, existing work largely focuses on closed-world context, where all referents are known a priori. In this work, we present a human-subject study situated in an open-world context, and compare the task performance and subjective perception associated with three different iconic gesture designs (anthropomorphic, non-anthropomorphic, deictic-iconic) against previously studied abstract gesture design. Our quantitative and qualitative results demonstrate that deictic iconic gestures (in which a robot hand is shown pointing to a visualization of a target referent) outperforms all other gestures on all metrics – but that non-anthropomorphic iconic gestures (where a visualization of a target referent appears on its own) is overall most preferred by users. These results represent a significant step forward to enabling effective human-robot interactions in realistic large-scale open-world environments.**

INTRODUCTION

Robots can assist human teammates in many settings by providing information, guidance, or support [1].For robots to be competent, effective helpers and facilitate a positive user experience, they need to communicate effectively with their human teammates in natural and intuitive ways [2]. While robots often use speech to communicate with human teammates, research has also demonstrated the benefits of accompanying that speech with non-verbal cues like gestures [3], [4]. In human-human interaction, gestures that involve moving one's hands and arms are often used in conjunction with speech [5]. These co-speech gestures can improve the clarity of speech, establish rapport with listeners, and provide context through emphasis [6]. Deictic and iconic gestures are two types of frequently used co-speech gestures, where deictic gestures are used to direct joint attention to a certain area or referent [7], [8], and iconic gestures are used to suggest visual characteristics of a referent, like shape and size [9], [10]. Prior work demonstrates that deictic and iconic gestures can enhance the listener's understanding of their environment [11], [12] and that people actively adapt their gestures during complex tasks [13].

Because these types of nonverbal cues are so effective in human-human communication, roboticists have also started to use gestures to accompany robot speech, in order to maximize the clarity and naturalness of human-robot interactions [14], [15]. However, many currently available robots are **morphologically limited**, either lacking the arms necessary to perform gestures, or lacking sufficiently dexterous arms and hands to do so accurately and expressively. Even when robots have physical arms, those arms are subject to significant physical constraints. For instance, if a robot's arm is blocked from view, it may not be able to gesture effectively despite its morphology. A key technology that circumvents these limitations is Augmented Reality (AR).

Recent research has demonstrated that AR-enabled robot gestures are just as effective as physical gestures [16], at least in closed-world environments in which robots must identify objects already *visible* to human teammates. In more realistic open-world environments, however, robots often need to refer to referents that are *not visible* to human teammates at the time of reference. This challenge has been recently recognized by research outside the scope of AR-enabled robot gesture, which has started moving beyond closed-world contexts. For example, Huang et al. investigated how robots might gesture differently to visible vs. non-visible referents, and argued that when referents are non-visible, robots should opt for *abstract deictic gestures* instead of precise deictic gestures. Yet because abstract deictic gestures alone do not achieve high accuracy, Huang et al. acknowledged that they would need to be paired with other modalities of communication such as speech to achieve sufficient accuracy [17].

We note, however, that when referring to non-visible objects in open-world contexts, humans regularly make use of iconic rather than deictic gestures. As such, in this work, we investigate whether robots might communicate more effectively in open-world environments if they were capable of iconic gestures. Moreover, because iconic gestures are notoriously challenging to generate without human-like manual dexterity, we further investigate whether AR might once again serve as a key technology for enabling effective iconic gestures.

To answer these questions, we begin by exploring the design space of AR gestures that might be used to refer to non-visible objects. Then, we perform a human-subject study to investigate the objective and subjective differences when robots use various AR gestures to refer to non-visible objects. Through this work, we thus answer the key **research question**: *When robots must refer to non-visible objects, what form of gesture will lead to the greatest objective and subjective performance?*

The first, second and fourth authors are with the MIRRORLab at the Colorado School of Mines, Golden, CO, USA {zhu1,altorres, twilliams}@mines.edu . The third author is with Fairview High School, Boulder, CO, USA zanealoia@gmail.com. This work was funded in part by ONR Award N00014-21-1-2418.

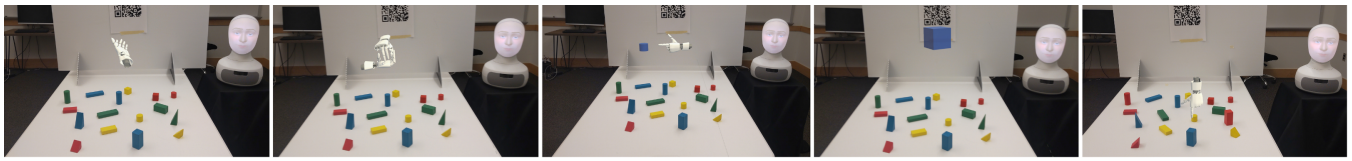


Fig. 1: Screen capture of five different virtual gestures. From left to right (images 1-5), in increasing levels of precision, screen captures show abstract deictic, anthropomorphic iconic, deictic iconic, non-anthropomorphic iconic, and precise deictic gestures.

Related work

Gestures in Human-Robot Interaction (HRI)

Social robots often use non-verbal cues to achieve natural, friendly, and effective communication. Non-verbal cues such as gestures can make robot communication more precise, user-friendly, and intuitive [18], [19], [20]. Sauppe et al. showed that robots that use deictic gestures are highly effective when identifying a referent among groups of objects [21], Ali and Williams found participants correctly identified objects faster when robots use co-speech gestures and gaze [22], and Lohse et al. observed that robot gestures reduced workload and improved task performance during challenging tasks [23]. In addition to improving clarity, gestures can also bring interpersonal benefit. For instance, De Wit et al. found that robot-generated iconic gestures made children more engaged during learning [24], Li et al. showed that robots become more agreeable to participants when they mimic human gesture during interaction [25], Riek et al. found robot gestures to help build rapport with humans [26], and Deshmukh et al. showed that greater speed and amplitude of robot gestures could improve robot likability and animacy [27]. However, enabling gestures for morphologically limited robots can be challenging: mounting physical arms on armless robots for the sole purpose of gesturing may be prohibitively expensive and create safety hazards, and robot appendages used for grasping are rarely dexterous enough to perform accurate and expressive gestures. Since virtual gestures do not involve physical robot appendages and the safety risks that may come with them [28], [29], AR presents unique opportunities to safely and effectively integrate gestures into human-robot interactions.

AR-based robot gestures

A growing body of work has integrated AR into HRI. While often bound to the physical form of robots, interactions can extend to virtual environments using AR as a medium. This can enable more intuitive robot programming [30], [31], enhance real-time monitoring of robot internal states [32], [33], enable robots to overcome morphological limitations to improve both communication [34], [35], [36] and expressivity [37], [38], and even improve the science of HRI by enabling better experimental control [39]. Beyond extending in-person interaction to the virtual space, some work also leverages the visual dimension of AR interfaces to improve human-robot communication by enabling robot gestures. Tran et al. showed that robots can use mixed reality deictic gestures without causing additional cognitive burden [40], Han et al. found that robots can use virtual arms to point to physical objects without causing harm to task performance or subjective experience [16], and Brown et al. found that robots could combine different types of mixed reality deictic gestures to simultaneously enable quicker reaction times and more positive social perceptions [41]. This body of work, while demonstrating the effectiveness of AR-based gestures and providing guidelines for using them, focuses solely on robots that use deictic gestures to refer to *visible objects*. In real-world scenarios, robots may need to use gestures to identify both *visible* and *non-visible objects*. Competent robot helpers may tailor their communication based on whether or not their referent is visible. Specifically, robots may opt to use different gestures based on referent visibility to improve the efficiency of the communication and subjective experience during the interaction. It is unclear, however, what kind of alternative gesture may be suitable for robots to refer to non-visible objects.

Gesturing in open-world context

Existing research in robot gestures focuses on what type of gesture robots should use in closed-world interactions. Compared to closed-world interaction where all objects are known and visible [42], in more realistic open-world interactions interlocutors may need to refer to objects unknown to observers a priori [43], [44]. During real-world human-robot collaborations, the human teammate might not always have the opportunity to become aware of all objects in the environment, and even if they were aware of all objects earlier during the interaction, they might forget information as the interaction goes on; Human teammates and robots might also physically move around in the environment during collaboration, changing the visibility of potential referents as objects become blocked from view.

To further understand gesture use in open-world contexts, Higger et al. present an analysis of human gestures during open-world references, and found that participants often used iconic gestures to aid open-world references [45]. Other prior work also highlights the ways that visualizing relevant objects can improve transparency and lighten the workload of human teammates [46]. While robot iconic gesture is often used to increase expressivity [47] or to communicate past robot behavior [48], iconic gestures can also be particularly helpful for identifying non-visible objects, since deictic pointing gestures are no longer suitable or optimal. Iconic gestures' ability to represent the appearance of the referent can be further enhanced by the AR medium. However, AR-based gestures also bring additional questions about the specific design of the gesture. In the following section, we

detail the design process and design considerations of AR-based iconic gestures for open-world contexts.

AR-BASED ROBOT ICONIC GESTURE DESIGN

In this work, we explore the design space of open-world mixed-reality robot gestures, including deictic gestures, iconic gestures, and the combination thereof. Given that iconic gestures mimic the appearance of the referent, we first propose **non-anthropomorphic iconic gesture**, which is a direct rendering of the referent (see Fig 1 image 4). We then incorporate anthropomorphism into the gesture design in two different ways: **anthropomorphic iconic gesture** (Fig 1 image 2) and **deictic iconic gesture** (Fig 1 image 3). In the deictic iconic gesture design, a virtual hand is used to draw the listener's attention to a visual depiction of the non-visible referent, in order to increase clarity and better convey the robot's intention to identify the referent.

As we consider the performance of these novel gesture designs, we note that aside from improving communication and task performance, gestures are also inherently social [21], [24], therefore, it's important to understand the implications of different robot gesture designs regarding social perception of the robot. In particular, our designs involve anthropomorphic cues. Anthropomorphic designs allow a robot to more readily fit into familiar ontological categories [49], leading to better acceptance [50] and a higher level of trust [51]. While advantageous in these respects, anthropomorphic designs also bring drawbacks. For instance, anthropomorphic designs may incorporate limitations that mechanomorphic or zoomorphic robots can easily circumvent, and in task environments, anthropomorphic designs may do more harm than good by causing distractions [52]. Therefore, in this work, we form the following hypotheses around task performance and subjective experience by considering the effect of precise reference and anthropomorphic designs:

Hypotheses: When used to refer to non-visible objects, *more precise gestures* will achieve greater task performance **(H1)**, and *more anthropomorphic gestures* will achieve greater subjective perceptions **(H2)**.

EXPECTED DESIGN IMPLICATIONS

Given our overarching hypotheses, we would expect key differences to be observed between our four novel AR gesture designs. Specifically, we would expect that more precise gestures will lead to better objective benefits, and thus would expect deictic iconic gestures to enable fastest task completion time greatest accuracy, and lower mental demand when referring to non-visible referents, followed by non-anthropomorphic iconic gestures, then anthropomorphic iconic gestures, with abstract deictic gesture achieving poorest performance on these metrics.

On the other hand, we would expect that more anthropomorphic gestures will lead to better subjective perceptions, and thus would expect anthropomorphic iconic gestures to best promote perceptions of robot competence, intelligence, agency, understandability, and helpfulness, followed by deictic iconic gestures, then abstract deictic gestures, with non-anthropomorphic iconic gestures achieving poorest performance on these metrics.



Fig. 2: Photo capture of the setup of the human-subject study showing a participant being instructed by a Furhat robot while wearing HoloLens

METHODS

Experimental Design

To assess our overarching hypotheses and these specific expectations, we designed and conducted an ethics board approved within-subject study in which participants experienced four conditions, each of which used precise deictic gestures to refer to visible referents, and a different type of abstract gesture (abstract deixis, non-anthropomorphic iconic gesture, anthropomorphic iconic gesture, and deictic iconic gesture) to refer to non-visible entities. This within-subject design allows participants to compare gestures to each other. In each within-subjects block, participants were asked to complete a series of tasks in which they were instructed to identify a wooden block (referred to by a robot using an auditory description and one of the five virtual gestures) as quickly and accurately as possible. Some described blocks were visible, and some were hidden behind a partition. The robot instructions were formatted as "The [color] [size] [shape]", where the color can be red, blue, green or yellow; the size can be short or long, and the shape can be rectangle, cylinder, triangle or cube. The term "long cube" emphasizes the visual similarity to a cube but highlights the elongated nature, differentiating it from the long/short rectangles. To prevent over-familiarizing participants with these shapes, each condition was associated with a different task, and the task area was reset in between experimental blocks. Each participant thus performed four different tasks and experienced four different open-world gestures, counterbalanced through a Graeco-Latin square. Each task involved the same types of blocks and contained the same number of instructions (14 randomly ordered instructions, 7 of which referred to visible objects, and 7 of which referred to non-visible objects).

Materials

Fig 2 shows the task environment and the mixed reality robotic system. We implemented all virtual gesture designs using the Unity 2020.3.4f1 editor, and deployed them to the HoloLens 2 HMD. The robot that spoke the instructions and was responsible for generating the AR gestures was a Furhat robot, with a simple gaze controller running.

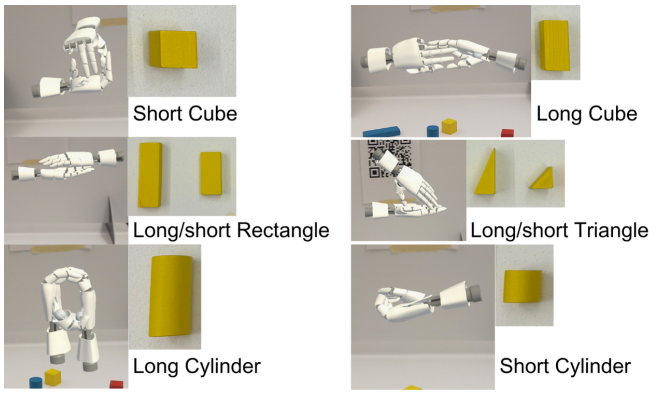


Fig. 3: Figure showing possible anthropomorphic iconic designs for eight types of shapes used

Measures

During study sessions, we collected both objective and subjective measures. To measure objective task performance, we recorded the accuracy and task completion time of each task. To measure subjective perception, we use the NASA TLX single-item scale to measure mental workload [53], [54], the Godspeed questionnaire to measure perceived intelligence [55], the RoSAS scale to measure perceived competence [56], a 6-item questionnaire supported by Korman et al. [57]'s prior work [58] to measure perceived agency, and 7-point Likert items to measure the understandability and the helpfulness of the robot helper ("Overall, I found the robot to be easy to understand" and "Overall, I found the robot to be helpful", 1 = Strongly disagree, 7 = Strongly agree).

Experiment procedure

Participants first entered the general experiment area and engaged in consent procedures. After providing informed consent, participants completed a short demographic survey and then put on the HoloLens 2 device. Participants then entered the task area to complete the four experimental blocks determined by their experimental condition. After each experimental block, participants completed a questionnaire including the measures described in the previous section. After all four experimental blocks, the experimenter conducted a short semi-structured interview, in which the experimenter asked the participant to provide verbal feedback about their preference for different gesture designs. Finally, participants were debriefed and compensated with a $10 gift card. The entire study lasted 42 minutes on average.

Participant demographics

20 participants were recruited from a university. 11 identified as male, 8 identified as female, and 1 identified as non-binary. Participant ages ranged from 18 to 63 ($Mean$=29.3). 14 (70%) reported experience with robots, 2 (11%) were neutral, and 4 (20%) disagreed. 15 (75%) reported experience with augmented reality and 5 (25%) disagreed.

Data Analysis

We used the Bayesian analysis framework [59] to analyze our data. While the Frequentist approach merely allows one to determine whether a null hypothesis can be rejected, the Bayesian approach quantifies evidence *for* and *against* competing hypotheses. This approach allows us to state conclusively when there is evidence *against* an effect. Within this analysis framework, we used version 0.19.1.0 of the JASP statistical software [60] to perform Bayesian Repeated-Measures Analyses of Variance [61]. When a main effect or interaction effect could not be ruled out (i.e., the Inclusion Bayes Factor BF_{incl} in favor of including the main or interaction term was above 0.333), Post hoc Bayesian t-tests were used to examine pairwise comparisons between conditions. All statistical analysis can be found in this repo. For participants' verbal responses to questions about their preference for the robot's usage of gestures, we examined their transcripts and coded participants' responses.

RESULTS

Experimental results are shown in Table I. As shown in that table, the Deictic Iconic gesture outperformed all other gestures on every metric. In the following sections we will step through the results of our Bayesian data analysis, which confirms this while providing further nuance. In these sections we only report evidence found in favor of results; evidence was found against all other effects (See our OSF repository for complete statistical results). Whenever an anecdotal effect is reported, this means that there is overall more evidence for than against the effect, but that more data would be needed to confirm the effect.

Task Accuracy

A Bayesian RM-ANOVA revealed strong evidence for an effect of gesture design on task accuracy (BF_{Incl} = 18.222, Fig. 4). Post hoc analysis revealed strong evidence that deictic iconic gestures enabled higher task accuracy than did abstract deictic gestures (BF_{Incl} = 23.286), moderate evidence that deictic iconic gestures enabled higher task accuracy than did anthropomorphic iconic gestures (BF_{Incl} = 3.236), and anecdotal evidence that deictic iconic gestures enabled higher task accuracy than did non-anthropomorphic iconic gestures (BF_{Incl} = 1.266). Post hoc analysis also revealed anecdotal evidence that non-anthropomorphic iconic gestures may have offered better accuracy than abstract deictic gestures (BF_{Incl} = 2.479).

Task Completion Time

A Bayesian RM-ANOVA revealed strong evidence for an effect of gesture design on task completion time (BF_{Incl} = 15.981, Fig. 5). While deictic iconic gestures overall facilitated the fastest task completion time, no statistical evidence was found for this effect. Rather, post-hoc analysis primarily evidenced the task completion time benefits that deictic iconic, abstract deictic, and non-anthropomorphic iconic gestures all provided as compared to a anthropomorphic iconic gestures. That is, post hoc analysis revealed

Measure	Abstract Deictic	Non-anthropomorphic Iconic	Anthropomorphic Iconic	Deictic Iconic
Accuracy	0.871 (0.153)	0.957 (0.067)	0.917 (0.126)	**0.993 (0.032)**
Task Completion Time	8.338 (1.326)	8.236 (1.567)	9.145 (1.903)	**7.984 (1.437)**
TLX Mental Demand	3.450 (2.982)	2.950 (2.259)	3.900 (3.093)	**2.300 (2.029)**
Understandability	5.550 (1.099)	5.900 (0.968)	5.350 (1.226)	**6.250 (0.716)**
Perceived Intelligence	3.730 (0.925)	3.820 (0.960)	3.740 (0.885)	**4.050 (0.853)**
Perceived Competence	3.475 (0.862)	3.587 (0.832)	3.500 (0.987)	**3.763 (0.705)**
Perceived Agency	4.058 (0.984)	4.042 (1.039)	4.000 (0.927)	**4.233 (1.115)**
Helpfulness	5.050 (1.638)	5.500 (1.469)	5.050 (1.468)	**5.700 (1.342)**

TABLE I: Table showing Mean value followed by SD value of measures of each gesture design

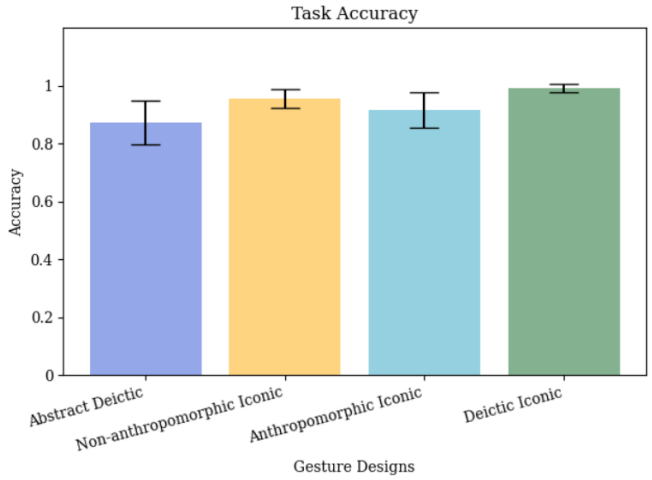


Fig. 4: Mean value of task accuracy of each gesture design, error bar shows standard error

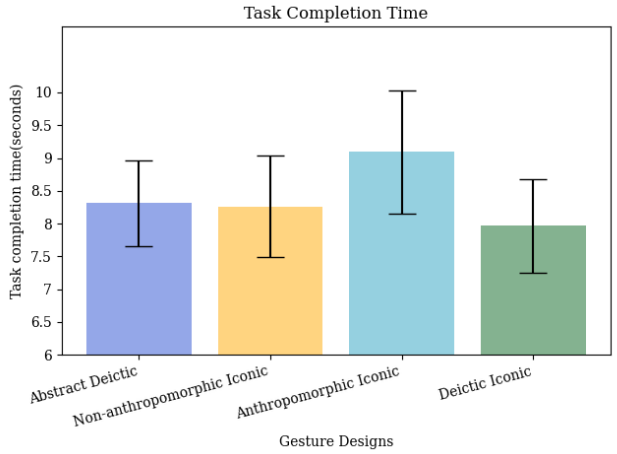


Fig. 5: Mean value of Task Completion Time of each gesture design, error bar shows standard error

very strong evidence that deictic iconic gestures enabled faster task completion times than did anthropomorphic iconic gestures (BF_{Incl} = 52.746), moderate evidence that non-anthropomorphic iconic gestures enabled faster task completion times than did anthropomorphic iconic gestures (BF_{Incl} = 8.157), and moderate evidence that abstract deictic gestures enabled faster task completion times than did anthropomorphic iconic gestures (BF_{Incl} = 3.819).

TLX Mental Demand

A Bayesian RM-ANOVA revealed extreme evidence for an effect of gesture design on mental demand (BF_{Incl} = 115.802, Fig. 6). Post hoc analysis revealed very strong evidence that deictic iconic gestures required less mental demand than did anthropomorphic iconic gestures gestures (BF_{Incl} = 33.935), strong evidence that deictic iconic gestures required less mental demand than did abstract deictic gestures (BF_{Incl} = 11.356), and anecdotal evidence that deictic iconic gestures required less mental demand than did non-anthropomorphic iconic gestures (BF_{Incl} = 2.520). Post hoc analysis also revealed moderate evidence that non-anthropomorphic iconic gestures were less mentally demanding than anthropomorphic iconic gestures (BF_{Incl} = 3.182).

Understandability

Bayesian RM-ANOVA revealed moderate evidence for an effect of gesture design on understandability (BF_{Incl} =

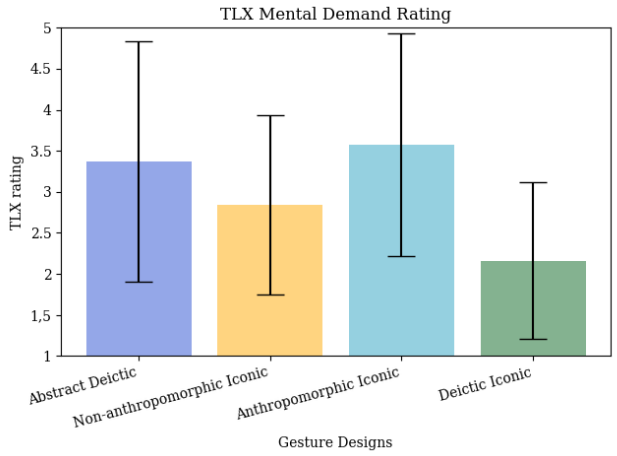


Fig. 6: Mean value of TLX rating of each gesture design, error bar shows standard error

3.417, Fig. 6). Post hoc analysis revealed moderate evidence that deictic iconic gestures led participants to view the robot as more understandable than did anthropomorphic iconic gestures gestures (BF_{Incl} = 3.320), moderate evidence that deictic iconic gestures led participants to view the robot as more understandable than did abstract deictic gestures (BF_{Incl} = 5.669), and anecdotal evidence that deictic iconic gestures led participants to view the robot as more understandable than did non-anthropomorphic iconic gestures

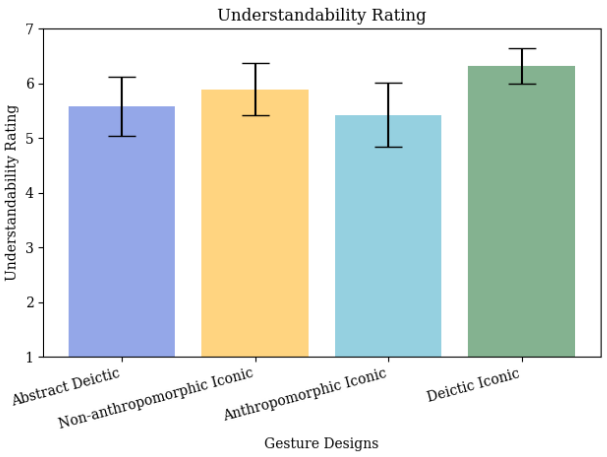


Fig. 7: Mean value of Understandability rating of each gesture design, error bar shows standard error

(BF_{Incl} = 2.234).

Perceived Intelligence, Perceived Competence, Perceived Agency, Helpfulness

A Bayesian RM-ANOVA revealed moderated evidence against the effect of gesture design in regard to the perceived agency (BF_{Incl} = 0.242), and anecdotal evidence against an effect of gesture designs regarding perceived intelligence (BF_{Incl} = 0.695), perceived competence (BF_{Incl} = 0.415) and helpfulness (BF_{Incl} = 0.393).

Preference

When asked to share their preferred robot gesture, 13 participants (65%) expressed that they preferred non-anthropomorphic iconic gestures, citing that they found this type of gesture to be clear and intuitive. 4 participants stated that they preferred deictic iconic gestures (20%), and 3 participants stated that they preferred abstract deictic gestures (15%), explaining that they believe the robot's use of abstract deictic gestures was intended to inform them the referent is currently not visible for them. Additionally, 12 participants (60%) explicitly mentioned that they disliked anthropomorphic iconic gestures, explaining that they found these gestures to be confusing, unhelpful, and distracting.

Discussion

Our first hypothesis **(H1)** was that more precise gestures would achieve greater task performance. Our results partially support this hypothesis. While the most precise gesture (deictic iconic) did indeed facilitate greatest task performance across all task performance metrics, there was not a strictly linear relationship between precision of gesture type and task performance. One explanation for this is that for highly imprecise gestures, users merely relied on different information channels to maintain performance. Indeed, some participants noted during their interviews that they chose to rely on robot speech more when abstract deictic gestures were used, since it was easy to interpret that the gesture itself simply instructed participants to look among non-visible objects.

Our second hypothesis **(H2)** was that more anthropomorphic gestures would achieve greater subjective perceptions. This was not supported. No effects were found for perceived intelligence, perceived competence, perceived agency, or helpfulness; deictic-iconic gestures (only the second most humanlike gesture) were viewed as most understandable, and second most preferred, the anthropomorphic iconic gestures were viewed as *least* understandable and *least* preferred, and the explicitly non-anthropomorphic iconic gestures were in fact viewed as the most preferred. **Overall then, these results clearly refute our hypotheses and instead suggest that deictic iconic gestures hold unique benefits to human-robot interaction, and that non-anthropomorphic iconic gestures, while not as highly performing, are dramatically preferred by participants.**

Conclusion

In this work, we presented four novel AR-based gestures: **abstract deictic, anthropomorphic iconic, non-anthropomorphic iconic, and deictic iconic gestures**. Our results show that deictic iconic gestures may be an ideal gesture for referring to non-visible entities on objective terms, but also show that non-anthropomorphic iconic gestures enjoy significant user preference as well. While the reason for this disconnect is unclear and should be examined in future work, these results overall suggest that deictic iconic gestures could be reasonably adopted as a new standard for referring to non-visible objects. Moreover, this work opens up a number of avenues for future work, which could explore an even wider array of gestures, compare and evaluate physical and AR-based gestures, in more challenging and complex experimental contexts, and as deployed into real-world in-the-wild tasks. Overall, this work represents a significant step forward for enabling effective human-robot interactions in realistic large-scale open-world domains.

References

[1] M. J. Matarić and B. Scassellati, "Socially assistive robotics," *Springer handbook of robotics*, 2016.
[2] K. S. Welfare, M. R. Hallowell, J. A. Shah, and L. D. Riek, "Consider the human work experience when integrating robotics in the workplace," in *HRI*. IEEE, 2019.
[3] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, F. Nori *et al.*, "Integration of action and language knowledge: A roadmap for developmental robotics," *IEEE Transactions on Autonomous Mental Development*, 2010.
[4] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, H. Yan *et al.*, "Human conversation as a system framework: Designing embodied conversational agents," *Embodied conversational agents*, 2000.
[5] D. McNeill, "Hand and mind1," *Advances in Visual Semiotics*, 1992.
[6] D. J. Lincoln, "Drama in the classroom: How and why marketing educators can use nonverbal communication and enthusiasm to build student rapport," *Marketing Education Review*, 2008.
[7] J. C. Lester, J. L. Voerman, S. G. Towns, and C. B. Callaway, "Deictic believability: Coordinated gesture, locomotion, and speech in lifelike pedagogical agents," *Applied Artificial Intelligence*, 1999.
[8] A. Kobsa, J. Allgayer, C. Reddig, N. Reithinger, D. Schmauks, K. Harbusch, and W. Wahlster, "Combining deictic gestures and natural language for referent identification," in *COLING*, 1986.
[9] H. Holle and T. C. Gunter, "The role of iconic gestures in speech disambiguation: Erp evidence," *J. cognitive neuroscience*, 2007.

[10] A. Kendon, *Gesture: Visible action as utterance*. Cambridge University Press, 2004.

[11] N. Dargue, N. Sweller, and M. P. Jones, "When our hands help us understand: A meta-analysis into the effects of gesture on comprehension." *Psychological Bulletin*, 2019.

[12] S. Goldin-Meadow, "Talking and thinking with our hands," *Current directions in psychological science*, 2006.

[13] I. Masson-Carro, M. Goudbeek, and E. Krahmer, "How what we see and what we know influence iconic gesture production," *Journal of nonverbal behavior*, 2017.

[14] A. G. Brooks and C. Breazeal, "Working with robots and objects: Revisiting deictic reference for achieving spatial common ground," in *HRI*, 2006.

[15] N. Gasteiger, T. Zhu, E. Broadbent, J. Lim, B. A. MacDonald, and H. S. Ahn, "Social robots in hospital settings: An initial exploration of the services provided, interaction style and in the field evaluation," in *Companion of the ACM/IEEE Intl. Conference on HRI*, 2024.

[16] Z. Han, Y. Zhu, A. Phan, F. S. Garza, A. Castro, and T. Williams, "Crossing reality: Comparing physical and virtual robot deixis," in *HRI*, 2023.

[17] A. Huang, A. Ranucci, A. Stogsdill, G. Clark, K. Schott, M. Higger, Z. Han, and T. Williams, "(gestures vaguely): The effects of robots' use of abstract pointing gestures in large-scale environments," in *HRI*, 2024.

[18] H. Admoni, T. Weng, B. Hayes, and B. Scassellati, "Robot nonverbal behavior improves task performance in difficult collaborations," in *11th ACM/IEEE Intl. Conference on HRI*, 2016.

[19] Z. Han, E. Phillips, and H. A. Yanco, "The need for verbal robot explanations and how people would like a robot to explain itself," *ACM Transactions on HRI*, 2021.

[20] M. Salem, S. Kopp, I. Wachsmuth, K. Rohlfing, and F. Joublin, "Generation and evaluation of communicative robot gesture," *Intl. Journal of Social Robotics*, 2012.

[21] A. Sauppé and B. Mutlu, "Robot deictics: How gesture and context shape referential communication," in *HRI*, 2014.

[22] W. Ali and A. B. Williams, "Evaluating the effectiveness of nonverbal communication in human-robot interaction," in *Comp. HRI*, 2020.

[23] M. Lohse, R. Rothuis, J. Gallego-Pérez, D. E. Karreman, and V. Evers, "Robot gestures make difficult tasks easier: the impact of gestures on perceived workload and task performance," in *CHI*, 2014.

[24] J. de Wit, A. Brandse, E. Krahmer, and P. Vogt, "Varied human-like gestures for social robots: Investigating the effects on children's engagement and language learning," in *HRI*, 2020.

[25] J. Li, W. Ju, and C. Nass, "Observer perception of dominance and mirroring behavior in human-robot relationships," in *HRI*, 2015.

[26] L. D. Riek, P. C. Paul, and P. Robinson, "When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry," *Journal on Multimodal User Interfaces*, 2010.

[27] A. Deshmukh, B. Craenen, A. Vinciarelli, and M. E. Foster, "Shaping robot gestures to shape users' perception: the effect of amplitude and speed on godspeed ratings," in *HAI*, 2018.

[28] T. S. Tadele, T. de Vries, and S. Stramigioli, "The safety of domestic robotics: A survey of various safety-related publications," *IEEE robotics & automation magazine*, 2014.

[29] A. Zacharaki, I. Kostavelis, A. Gasteratos, and I. Dokas, "Safety bounds in human robot interaction: A survey," *Safety science*, 2020.

[30] R. S. Lunding, M. S. Lunding, T. Feuchtner, M. G. Petersen, K. Grønbæk, and R. Suzuki, "Robovisar: Immersive authoring of condition-based ar robot visualisations," in *HRI*, 2024.

[31] S.-K. Ong, A. Yew, N. K. Thanigaivel, and A. Y. Nee, "Augmented reality-assisted robot programming system for industrial applications," *Robotics and Computer-Integrated Manufacturing*, 2020.

[32] A. Rotsidis, A. Theodorou, J. J. Bryson, and R. H. Wortham, "Improving robot transparency: An investigation with mobile augmented reality," in *RO-MAN*, 2019.

[33] A. E. Solyman, K. M. Ibrahem, M. R. Atia, H. I. Saleh, and M. R. Roman, "Perceptive augmented reality-based interface for robot task planning and visualization," *Int. J. Innov. Comput. Inf. Control*, 2020.

[34] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in *HRI*, 2018.

[35] Y. Cao, T. Wang, X. Qian, P. S. Rao, M. Wadhawan, K. Huo, and K. Ramani, "Ghostar: A time-space editor for embodied authoring of human-robot collaborative task with augmented reality," in *UIST*, 2019.

[36] H. Wang, T. Do, and Z. Han, "To understand indicators of robots' vision capabilities," in *VAM-HRI*, 2024.

[37] T. Groechel, Z. Shi, R. Pakkar, and M. J. Matarić, "Using socially expressive mixed reality arms for enhancing low-expressivity robots," in *RO-MAN*, 2019.

[38] A. M. Villanueva, Z. Liu, Z. Zhu, X. Du, J. Huang, K. A. Peppler, and K. Ramani, "Robotar: An augmented reality compatible teleconsulting robotics toolkit for augmented makerspace experiences," in *CHI*, 2021.

[39] T. Williams, L. Hirshfield, N. Tran, T. Grant, and N. Woodward, "Using augmented reality to better study human-robot interaction," in *VAMR*, 2020.

[40] N. Tran, T. Grant, T. Phung, L. Hirshfield, C. Wickens, and T. Williams, "Now look here! mixed reality improves robot communication without cognitive overload," in *VAMR*, 2023.

[41] L. Brown, J. Hamilton, Z. Han, A. Phan, T. Phung, E. Hansen, N. Tran, and T. Williams, "Best of both worlds? combining different forms of mixed reality deictic gestures," *ACM Transactions on HRI*, 2023.

[42] A. Lücking, T. Pfeiffer, and H. Rieser, "Pointing and reference reconsidered," *Journal of Pragmatics*, 2015.

[43] T. Williams and M. Scheutz, "A domain-independent model of open-world reference resolution." in *CogSci*, 2015.

[44] T. Williams, S. Acharya, S. Schreitter, and M. Scheutz, "Situated open world reference resolution for human-robot dialogue," in *11th ACM/IEEE International Conference on HRI*. IEEE, 2016.

[45] M. Higger, P. Rygina, L. Daigler, L. F. Bezerra, Z. Han, and T. Williams, "Toward open-world human-robot interaction: What types of gestures are used in task-based open-world referential communication?" in *SemDial*, 2023.

[46] K. Mahadevan, M. Sousa, A. Tang, and T. Grossman, ""grip-that-there": An investigation of explicit and implicit task allocation techniques for human-robot collaboration," in *CHI*, 2021.

[47] M. Salem, K. Rohlfing, S. Kopp, and F. Joublin, "A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction," in *RO-MAN*, 2011.

[48] P. Bremner and U. Leonards, "Iconic gestures for robot avatars, recognition and integration with speech," *Front. Psych.*, 2016.

[49] B. R. Duffy, "Anthropomorphism and the social robot," *Robotics and autonomous systems*, 2003.

[50] J. Fink, "Anthropomorphism and human likeness in the design of robots and human-robot interaction," in *Int'l Conf. Soc. Rob.*, 2012.

[51] A. Waytz, J. Heafner, and N. Epley, "The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle," *Journal of experimental social psychology*, vol. 52, pp. 113–117, 2014.

[52] L. Onnasch and C. L. Hildebrandt, "Impact of anthropomorphic robot design on trust and attention in industrial human-robot interaction," *ACM Transactions on Human-Robot Interaction (THRI)*, 2021.

[53] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*, 1988.

[54] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proc.of the human factors and ergonomics society annual meeting*, 2006.

[55] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *Int'l J. Soc. Rob.*, 2009.

[56] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas) development and validation," in *HRI*, 2017.

[57] J. Korman, A. Harrison, M. McCurry, and G. Trafton, "Beyond programming: can robots' norm-violating actions elicit mental state attributions?" in *HRI*, 2019.

[58] C. Frazier-Young, M. McCurry, K. Zish, and G. Trafton, "Perceived agency changes performance and moral trust in robots," in *Proc. of the Annual Meeting of the Cognitive Science Society*, 2022.

[59] E.-J. Wagenmakers, J. Love, M. Marsman, T. Jamil, A. Ly, J. Verhagen, R. Selker, Q. F. Gronau, D. Dropmann, B. Boutin *et al.*, "Bayesian inference for psychology. part ii: Example applications with jasp," *Psychonomic bulletin & review*, 2018.

[60] JASP Team, "JASP (Version 0.19.1.0)[Computer software]," 2024. [Online]. Available: https://jasp-stats.org/

[61] D. van den Bergh, E.-J. Wagenmakers, and F. Aust, "Bayesian repeated-measures anova: An updated methodology implemented in jasp," 2022.