Evaluating the impact of non-verbal communication features on user engagement and perceived anthropomorphism in Human-Robot Interaction

Investigating the Engagement and Anthropomorphic Effects of Facial Expressions vs. Body Movements in Huam-Social Robot Interaction

Jay Moon
Master of Design in Design for Interaction,
Carnegie Mellon University
Pittsburgh, PA, United States
jaymoon@andrew.cmu.edu

Nik Kim
Master of Science in Computational Design,
Carnegie Mellon University
Pittsburgh, PA, United States
nnkim@andrew.cmu.edu

Abstract— This study explores which non verbal communication features—facial expressions or body movements—most effectively enhances user engagement in human-robot interaction scenarios and anthropomorphic perception of robots. Using a within-subjects design, participants interacted with a robot under three conditions: verbal-only (control), verbal with facial expressions, and verbal with expressive body movement. Participants completed engagement and anthropomorphism surveys after each condition. Our results suggest that both facial expressions and body movements significantly increased perceived human-likeness compared to the control condition, with facial expressions having the strongest effect. In terms of engagement, both experimental groups outperformed the control group, although there was no significant difference between facial expressions and body movement. A moderate positive correlation was also found between perceived anthropomorphism and engagement. However, crucially, increasing anthropomorphism does not necessarily lead to a proportional increase in engagement—and the reverse is also true. These findings offer insights in prioritizing robot's socially expressive features for their cost-effective development. (*Abstract*)

Keywords — Human-Robot Interaction (HRI), Anthropomorphism, Social Robots, Robot Expressiveness, User Engagement, Perceived Human-Likeness, Interaction Design

I. INTRODUCTION

As social robots become more common in homes, schools, and public spaces, designing interactions that feel engaging and natural becomes essential. Among various modalities, facial expressions and body movements are considered two key elements that contribute to the richness and dynamics of human-robot interaction. However, we identified a research gap that there has been little work comparing which modality—facial expressions or body movements—leads to greater user engagement and perceived anthropomorphism under the same conditions.. Realizing expressive features in robots often come with mechanical complexity and cost. This study investigates which non-verbal communication features—facial expressions or body movement—plays a more significant role in shaping user engagement and the robot's perceived human-likeness. By uncovering these relationships, our goal is to offer insights in developing modalities that are more engaging yet cost-effective.

II. BACKGROUND

In the field of Human-Robot Interaction (HRI), anthropomorphism—the tendency to attribute human-like qualities to non-human entities—has been a central focus for improving user acceptance, trust, and engagement. As social robots become more integrated into everyday settings such as homes, classrooms, and healthcare environments, understanding how different non-verbal communication features contribute to user engagement and robot's perceived human-likeness is increasingly important.

Previous studies have shown that expressive features such as facial expressions, body movements, vocal tone, and gesture timing significantly impact how users interpret a robot's emotional state, intelligence, and social presence [1], [2]. For example, expressive humanoid robots like Furhat and Pepper have demonstrated that even subtle eye or mouth movements can enhance social warmth and engagement [3]. Meanwhile, animation-inspired robotic movement—such as Disney's expressive bipedal robot—has shown that dynamic, human-like motion can elicit strong emotional responses, even in robots with non-human form factors [4].

To assess the perceived human-likeness of a robot, many researchers have adopted the Godspeed Questionnaire Series [5], a widely accepted tool in HRI studies. The Godspeed scales assess dimensions such as Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety, using semantic differential scales (e.g., artificial–lifelike, mechanical–organic). For instance, the anthropomorphism subscale helps identify whether a robot is viewed as machine-like or human-like across different expressive modalities. In our study, we used Likert-scale questions inspired by the Godspeed framework to evaluate participants' perceived human-likeness and engagement in a more task-specific context.

Despite the growing body of research on robot's capabilities in expressive non-verbal communication , few studies have directly compared facial expressions and body movements side by side within the same interaction framework [6]. Most studies either focus on one modality or combine both in complex humanoid robots, which makes it difficult to isolate their individual impact. Additionally, many expressive robots rely on costly mechanical actuators for facial or full-body expressions, raising concerns about scalability and accessibility in real-world deployment.

Our study addresses this gap by directly comparing three interaction conditions—verbal-only (control), verbal with facial expressions, and verbal with expressive body movements—using a commercially available social robot with limited but expressive capabilities. By evaluating participants' engagement and perceived anthropomorphism across these conditions, our goal is to determine which expressive modality has the strongest impact. More importantly, we aim to explore the relation between user engagement and perceived anthropomorphism. Our findings offer insights into prioritizing socially expressive elements in robot design for cost-effective development by clarifying the relationship between user engagement and the anthropomorphic perceptions they evoke.

III. METHODS

3.1 Participants

Thirteen participants (N = 13) were recruited from the university community. Participants ranged in age from 25 to 44 and came from diverse academic backgrounds, including design, engineering, HCI and computer science. Most reported limited to moderate prior experience interacting with social robots. All participants gave informed consent before participating in the study.

3.2 Study Design

This study used a **within-subjects design**, where each participant experienced all three interaction conditions with a robot in a randomized order:

a) ***Condition A** : Verbal-only interaction (Control condition)*
b) ***Condition B**: Verbal + Facial Expressions*
c) ***Condition C**: Verbal + Expressive Body Movement*

To minimize order effects and learning bias, we applied a Latin Square counterbalancing method. This ensured that each condition appeared in each position (first, second, third) across participants.

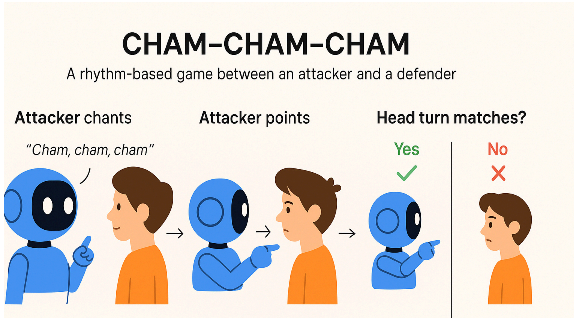

Fig. 1. How to play Cham-Cham-Cham (The image is generated by Chat GPT)


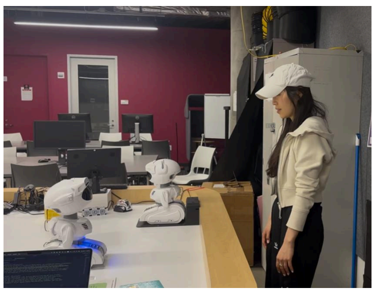

Fig. 2. Misty (left), Fig. 3. Misty and a participant (right)

3.3 Procedure

Participants engaged in a short interaction using a culturally familiar Korean game called **Cham-Cham-Cham**(Fig.1), which involves turn-taking gestures and rapid decision-making. In the original version of the game, the **attacker** rhythmically chants "Cham, Cham, Cham" and then quickly points in a direction (typically left or right), while the **defender** simultaneously turns their head in a direction. If the directions match, the attacker wins the round; if not, the round continues or resets.

In our implementation, the **robot (Misty)** played the role of the attacker. Due to Misty's hardware limitations, we adapted the pointing gesture using Misty's head direction instead of arm movement. Similarly, rather than having participants only turn their heads, we asked them to physically move to the left or right to make their responses easier for the robot to detect.

This game was chosen for its simplicity, visual clarity, and its playfulness, and because it allowed us to design minimally expressive interactions that still felt dynamic within Misty's physical capabilities. Each participant completed the game once for each condition, resulting in three sessions per person. In every session, Misty initiated the interaction, and participants played five rounds.

After each condition, participants completed a short survey assessing their **engagement** and **perception of human-likeness** based on the interaction they had just experienced. Upon completing all three conditions, participants were asked to compare the conditions and reflect on their preferences through ranking and open-ended responses.

3.4 Robot Platform

The interaction was implemented using the **Misty social robot**, a robot with basic body movement (e.g., head tilt, arm rotation) and a digital display for facial expressions. Misty's behaviors were programmed using the Misty II Javascript SDK, allowing for consistent expressive responses across all participants.

3.5 Measures and Survey Design

The user survey was developed through two rounds of pilot testing and a review session with the course instructor. The final version was structured to capture both quantitative and qualitative data on user engagement, perceived anthropomorphism, and interaction preference across three robot conditions. The survey consisted of four main components.

A. Demographics

Participants first completed a brief demographic form, providing age group, gender, academic background, and prior experience with social or expressive robots.

B. Pre-Interaction Questionnaire

Prior to experiencing any robot interaction, participants were asked to report their comfort level with social robots, familiarity with expressive robot behaviors (e.g., facial or body movement), and expectations for robot expressiveness. Responses were collected using 5-point Likert scales.

C. Condition-Specific Measures (Repeated for Conditions A, B, and C)

Following each interaction condition, participants responded to a set of standardized questions:

- ***Engagement*** *was measured using Likert-scale items adapted from prior HRI studies, including:*
  - *"I lost track of time while playing the game."*
  - *"I found myself emotionally reacting to the robot's behavior."*
  - *"I didn't think about anything else during the interaction."*
    In addition, an open-ended prompt asked participants to describe what aspects of the robot's behavior made the interaction more or less engaging.

- ***Anthropomorphism*** *was assessed using Likert-style items inspired by the Godspeed Questionnaire [5]:*
  - *"How human-like did the robot feel?"*
  - *"The robot seemed artificial or mechanical."*
  - *"The robot's expressions felt natural."*
    An indirect measure was included using a categorical question:
    "Which of the following best describes how the robot felt to you?"
    with options: (1) Machine or device, (2) Toy, (3) Pet or companion animal, (4) Human.
    Open-ended questions asked participants to reflect on moments when the robot felt particularly human-like or mechanical.

D. Post-Interaction Preference

After completing all three conditions, participants were asked to rank the conditions in order of preference and provide an open-ended explanation for their choices.

This survey design enabled the systematic comparison of engagement and anthropomorphism across conditions while allowing participants to reflect qualitatively on their experiences.

IV. RESULTS

We present results in two ways. First, in descriptive statistics and second in inferential statistics. Through descriptive statistics, which aim to summarize and describe the collected dataset without making conclusions about a larger population. Second, we will report inferential statistics, which allow us to draw conclusions or make inferences about the broader population based on the sample data. Specifically, we conducted hypothesis testing using repeated ANOVA analysis and correlation testing using Pearson r correlation.

4.1 Demographics and Logistics

We conducted a user test with a total of 13 participants, one male and twelve female. In pre-survey we questioned their prior experiences with research robots; one reported no experience, eight responded 1-5 times of former interaction, one reported 6-10 times and three answered as more than 10 times. Considering the distribution, our participants had moderate experience with research robot platforms.

In terms of experiment logistics, each participant had to play three rounds of the game paired with a different modality of Misty and within each round, participants played five times of the game. Considering the fairness of the game play, the order of modality that participants interact with was randomized.

After each round, participants were asked to fill out the questionnaire asking the degree of the robot's anthropomorphic effect. After finishing all three rounds, participants were lastly asked to rank Misty's engagement level and anthropomorphic effect to analyze the correlation between two variables.

4.2 Descriptive statistics

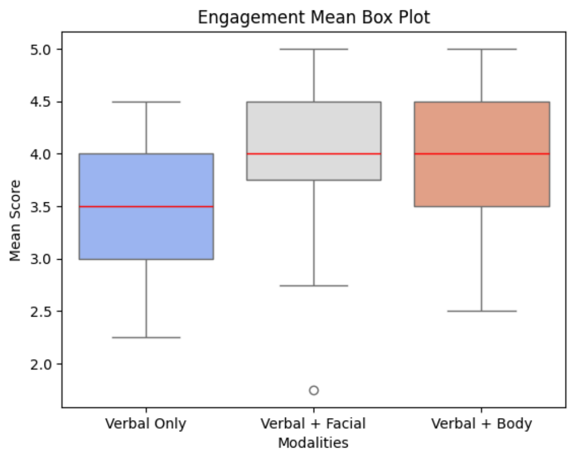

Fig. 4.1. Box plot showing engagement means per modalities.

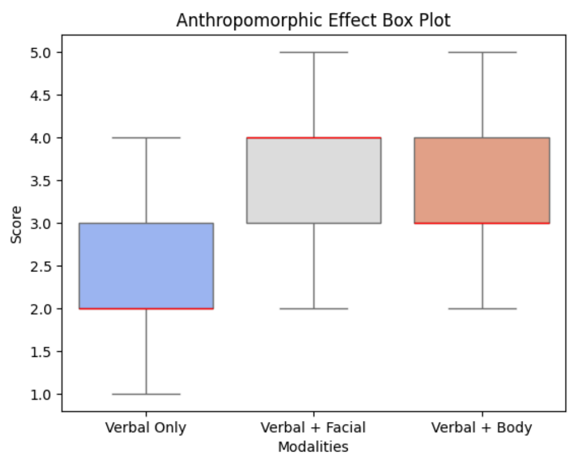

Fig. 4.2. Box plot showing anthropomorphic effect per modalities.

Fig 4.1 and Fig 4.2 show box plots of engagement level and anthropomorphic effect respectively. First in terms of engagement means, the control *verbal only* group showed median ≈ 3.5, IQR from about 3.0 to 4.0, whiskers ~ 2.25 to 4.5. Variable group verbal + facial expression showed median = 4.0, IQR ≈ 3.75–4.5, whiskers ~ 1.75 to 5.0. Lastly, another variable group verbal + body movement resulted in median = 4.0, IQR ≈ 3.5–4.5, whiskers ~ 2.5 to 5.0. To summarize, the median was the **same** in two variable groups, although the verbal + body movement group showed a bit wider IQR and whiskers compared to the verbal + facial expression group indicating more variances in the mean score. Intuitively, adding facial or body expressions on top of verbal expressions showed an increase in mean engagement score.

Fig 4.2 shows anthropomorphic effects observed in each group. Control verbal only group showed median 2.0, IQR from about 2.0 to 3.0, whiskers ~ 2.0 to 4.0. Compare group verbal + facial expression showed median = 4.0, IQR ≈ 3.0 to 4.0, whiskers ~ 3.0 to 5.0. Lastly another compare group verbal + body movement showed median = 3.0, IQR ≈ 3.0 to 4.0, whiskers ~ 2.0 to 5.0. To summarize, the median was **highest in verbal + facial** group in terms of anthropomorphic effect.

While asking participants to score the anthropomorphic effect in scale, we also asked them how close Misty was to a machine or device, a toy, a pet or companion animal, and a human. Following heat map shows the relationship between anthropomorphic score they gave and similarity keyword they answered.

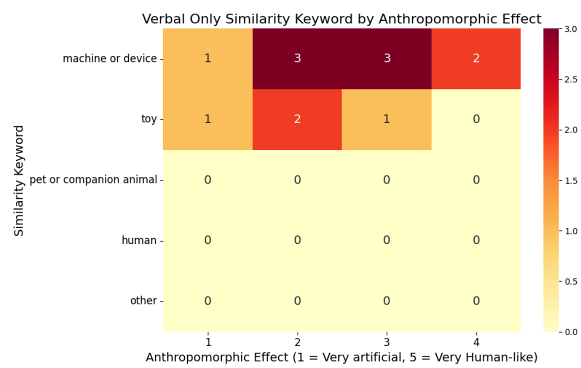

Fig. 4.3. Heat map showing anthropomorphic effect to similarity keyword in verbal only group.

Fig. 4.3 shows how participants mapped anthropomorphic to similarity keywords. 70% participants reported Misty was machine or device-like and 30% reported she as a toy. In terms of anthropomorphic effect, 70% reported 2 or 3, 15% gave 1 and 3 respectively. In verbal only control groups the most common mapping was similarity keyword machine or device and toy to 2 or 3 anthropomorphic effect.

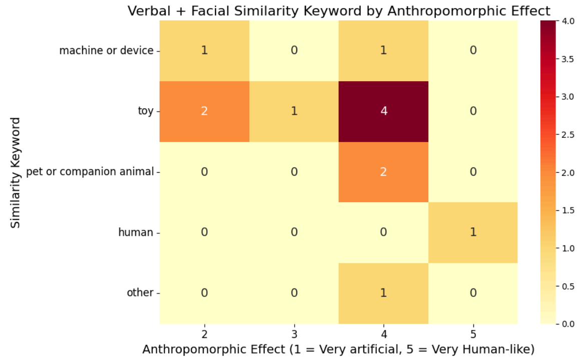

Fig. 4.4. Heat map showing anthropomorphic effect to similarity keyword in verbal + facial group.

Fig 4.4 shows the same mapping from similarity keyword to anthropomorphic effect in experimental group verbal + facial. Now the distribution varies compared to the earlier control group. Here, a toy keyword mapped into 4 of anthropomorphic effects are the most common. Only the 'other' respond illustrated similarity keywords to *in between toy and pet or companion animal.* In this group 70% of the participants reported Misty's anthropomorphic effect as 4 or 5.

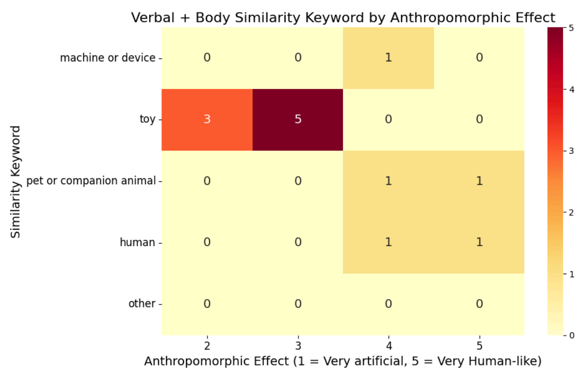

Fig. 4.5. Heat map showing anthropomorphic effect to similarity keyword in verbal + facial group.

Fig 4.5 shows the same mapping from similarity keyword to anthropomorphic effect in experimental group verbal + body movement. Similar to the earlier group, the distribution varies compared to the control group, showing higher anthropomorphic effect. Here, a toy keyword mapped into 3 of anthropomorphic effects are the most common. In this group, almost 70% of the participants reported Misty's anthropomorphic effect as 3 or 4.

4.3 Inferential statistics

Utilizing the sample data, we perform inferential statistics. The process of validating the independent variables effect on dependent variable was as below:

1. Perform repeated Analysis of Variance (ANOVA) to find there are significant differences in group means at least two of the groups.
2. Perform pairwise t-tests to examine which conditions differ from each other. Here we compare the control group to two experimental groups.
3. Perform correlation analysis to see if there is any correlation between dependent variables – engagement and anthropomorphic effect.

A. Repeated ANOVA

Repeated ANOVA is a type of hypothetical test that tries to verify the effect of the independent variable by rejecting the null hypothesis with predefined significance value p. For this test, we set the p threshold as 0.05, meaning there is a 5% chance that the null hypothesis is rejected not because of the main effect. Particularly we conducted repeated ANOVA since the same participants are measured multiple times under different conditions. We ran repeated ANOVA using open-source python statistical package *pingouin.*

Engagement

A one-way repeated measures ANOVA was conducted to examine the effect of modality on engagement. The result revealed a **non-significant** main effect of modality on the engagement, $F(2, 24) = 2.93$, $p = .073$, partial $\eta^2 = .085$.

TABLE I. EFFECT OF MODALITY ON ENGAGEMENT

Source	SS	df	MS	F	p	η²
Modality	2.93	2	1.01	2.93	.073	.085
Error	8.24	24	0.34			

Anthropomorphic Effect

A one-way repeated measures ANOVA was conducted to examine the effect of modality on anthropomorphic effect. The result revealed a **significant** main effect of modality on anthropomorphic effect, $F(2, 24) = 2.93$, $p = .001512$, partial $\eta^2 = .192$.

TABLE II. EFFECT OF MODALITY ON ANTHROPOMORPHIC EFFECT

Source	SS	df	MS	F	p	η²
Modality	8.36	2	4.18	8.62	.002	.192
Error	11.6	24	0.49			

B. Post hoc analysis; Pairwise t-tests

From repeated ANOVA tests, we found out there is no significant difference in engagement depending on modalities but a significant difference exists in anthropomorphic effect with respect to modalities. Based on this result, we conduct pairwise t-tests to examine detailed comparisons between facial expressions to baselines and body movement to baselines. Paired t-test is a method used to test whether the mean difference exists between pairs of measurements. We ran *pairwise t-tests* using open-source python statistical package *pingouin.* To correct p-values against the risk of false positives, we used one-step Bonferroni correction.

A paired-samples t-test revealed a significant difference between verbal only and verbal + facial expression, $t(12) = 4.07$, $p = .005$, $g = 1.08$.

A paired-samples t-test revealed a significant difference between verbal only and verbal + body movement, $t(12) = 3.39$, $p = .016$, $g = 0.82$.

TABLE III. PAIRED COMPARISON ON ANTHROPOMORPHIC EFFECT

Pairs	t(12)	p-corr	Hedges' g	BF10
Verbal vs Facial	4.07	.005	1.08	27.26
Verbal vs Body	3.39	.016	0.82	9.72

C. Correlation

Lastly, we checked the correlation between reported engagement mean and anthropomorphic effect using Pearson r correlation. Pearson correlation coefficient (r) is a way of measuring a linear correlation. The value ranges (-1, 1), measures the strength and negative, positive, or none relation between two variables. We conducted Pearson r test using *stats* module from *scipy* package.

- *$0 < r < 1$: Positive correlation*
- *$R = 0$: No correlation*
- *$-1 < r < 0$: Negative correlation*

There was a significant positive correlation between participant's reported engagement level and their perceived anthropomorphic effect, $r(37) = 0.38$, $p = 0.017$.

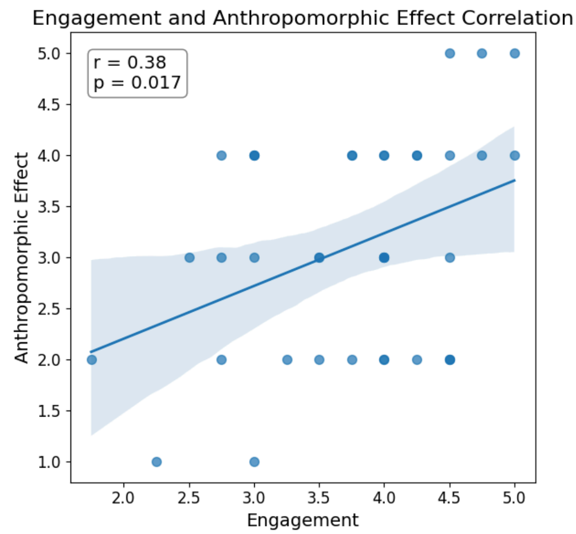

Fig. 4.6. Regression plot showing correlation between engagement and anthropomorphic effect.

V. DISCUSSION

This study aimed to examine which expressive modality—facial expressions or body movements—has a greater effect on enhancing engagement and perceived anthropomorphism in human-robot interaction (HRI). Our findings offer several insights for the design and evaluation of socially assistive robots, particularly those with limited expressive capability.

A. Engagement and Expressive Modality

Although descriptive statistics suggested that both facial and body expressive conditions increased engagement compared to the verbal-only control, the difference was not statistically significant in our inferential analysis. One possible explanation is that verbal feedback alone may already provide a sufficient baseline for engagement, particularly in short, structured interactions. Moreover, participants in our study received verbal prompts and behavioral cues with consistent timing, which may have masked the relative impact of expressive behaviors. Engagement, as a behaviorally-driven measure, may also be more sensitive to repetition or novelty loss, which was controlled using randomized condition orders but may still influence participant focus and motivation.

B. Anthropomorphic Perception and Expressive Modality

In contrast, the anthropomorphic effect showed a statistically significant difference across conditions, with both facial expressions and body movements rated as more human-like than verbal-only interaction. Facial expressions in particular led to the highest anthropomorphic ratings, suggesting that this modality plays a stronger role in shaping social perception. This aligns with previous findings in HRI literature that non-verbal social cues—especially facial animations—can elicit stronger feelings of human-likeness. Importantly, these effects are perceptual rather than behavioral and appear less susceptible to fatigue or learning effects within a short-term study.

C. Correlation Between Engagement and Anthropomorphism

We also observed a moderate positive correlation between participants' engagement scores and their anthropomorphism ratings. This implies that robots perceived as more human-like are also more engaging, but it does not suggest a perfect one-to-one relationship. In other words, higher anthropomorphism may support but does not guarantee increased engagement. This distinction is important for practical robot design: some expressive features may carry significant technical or cost burdens while offering only marginal gains in behavioral engagement.

D. Design Implications

Our results support the idea of a "just enough" anthropomorphism strategy—where designers prioritize expressive features that maximize perceived social presence without over engineering or inflating production costs. For example, facial expressiveness may yield high anthropomorphic perception and serve as a valuable feature to retain in low-cost or mobile platforms, whereas complex body articulation might offer diminishing returns depending on context and interaction goals.

E. Limitations and Future Work

This study has several limitations that point toward promising directions for future research. Despite the use of randomized condition orders, we observed an **order effect**. Some participants rated the control condition highly when it was presented first; however, encountering the same condition in later rounds resulted in lower ratings. As participants saw the same prompts repeatedly, they began to feel fatigued. They also got used to Misty's behavior, which made it feel less novel, and this may have led to decreased engagement.

There were also **technical limitations** related to the robot's expressive capabilities. Misty's movements were constrained and lacked fine-grained nuance. Additionally, only a single adult female voice was available, and it sounded stiff and robotic, which reduced the naturalness of the interaction.

Interestingly, **verbal prompts appeared to have a stronger impact on engagement than anticipated**. Challenge-oriented phrases (e.g., playful teasing) elicited heightened responses, suggesting that verbal delivery contextualized to competitive game settings, such as making fun of human players, teasing or challenging them, significantly influenced user engagement.

Lastly, we suggest several future research directions. First, researchers could investigate verbal effects deeper by conducting comparative studies on different prompting styles—such as positive reinforcement versus challenge-based language—and their impact on user engagement and emotional response. Moreover, employing **robots with a broader range of expressive capabilities** , and **designing experiments that minimize fatigue and bias**, will further deepen understanding of how to design effective social cues in HRI.

VI. CONCLUSION

This study investigated the impact of non-verbal communication modalities—facial expressions and body movements—on user engagement and perceived anthropomorphism during short-term interactions with a social robot. While both modalities enhanced engagement descriptively, only the perceived anthropomorphic effect showed significant improvement, especially in the facial expression condition. A positive correlation between engagement and anthropomorphism suggests that perceived human-likeness can support better interaction quality but may not be the sole determinant of engagement.

These findings suggest that expressive design in social robots should consider cost-effective expressivity, emphasizing features like facial animation that yield strong perceptual impacts with relatively low hardware complexity. Future work could explore long-term interaction scenarios, personalized expressions, or adaptive robot behavior to further deepen user connection and expand on these initial findings.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Professor Zackory Erickson and Professor Michelle Zhou for their guidance and feedback throughout the project. We also thank our teaching assistant, Diana Frias Franco, for her support during the study process. Finally, we are grateful to all the participants who volunteered their time for the user study.

REFERENCES

[1] C. Breazeal, "Social interactions in HRI: The robot view," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 34, no. 2, pp. 181–186, May 2004.

[2] K. Fischer, "How people talk with robots: Designing dialogue to reduce user uncertainty," AI & Society, vol. 26, no. 3, pp. 255–265, 2011.

[3] Furhat Robotics. “Furhat SDK Documentation.” [Online]. Available: https://docs.furhat.io

[4] Disney Research, “How Disney packed big emotion into a little robot,” IEEE Spectrum, 2024. [Online]. Available: https://spectrum.ieee.org/disney-robot

[5] C. Bartneck, D. Kulic, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," International Journal of Social Robotics, vol. 1, no. 1, pp. 71–81, 2009.

[6] M. Hu et al., “ELEGNT: Expressive and Functional Movement Design for Non-Anthropomorphic Robots,” arXiv preprint arXiv:2501.12493, 2025