Article

Enhancing Human-Robot Companionship: Comparing Verbal and Non-Verbal Communication Cues in Humanoid Robots

Syed Munawar Saeed Kazmi[1,*]

[1] School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK; munawarsaeedkazmi1@gmail.com

[*] Correspondence: munawarsaeedkazmi1@gmail.com

Academic Editor: *To be assigned*

Received: — Revised: — Accepted: — Published: —

Citation: *Kazmi, S.M.S. Enhancing Human-Robot Companionship: Comparing Verbal and Non-Verbal Communication Cues in Humanoid Robots. Robotics* **2025**, *14, x. https://doi.org/10.3390/xxxxx*

Abstract: Background: As humanoid robots enter domestic, healthcare, and educational settings, effective communication of need states to human partners has become a central design concern. This study investigates how verbal and non-verbal communication modalities influence trust-building and social perception in companionship contexts. Methods: Using the JD Humanoid Robot (EZ-Robot Inc.), a within-subjects counterbalanced experiment with 20 participants compared non-verbal gestural cues (pointing sequences) and explicit verbal requests (spoken battery-low alert) in conveying a low-battery need state. Perceived clarity, competence, warmth, and discomfort were assessed using the Robotic Social Attributes Scale (RoSAS). Results: Verbal communication yielded significantly higher clarity ratings ($M = 5.8$, $SD = 0.9$) than non-verbal gestures ($M = 4.9$, $SD = 1.1$; $t(19) = 2.61$, $p = 0.017$, $d = 0.58$). Engagement scores were statistically comparable across conditions ($p = 0.12$). Thematic analysis identified two themes: trust through mimicry (non-verbal) and clarity through speech (verbal), with participants independently proposing hybrid strategies. Conclusions: Hybrid verbal-gestural strategies are recommended for companion robot design to simultaneously optimise clarity and warmth. Findings are exploratory; effect-size estimates are provided to power future replication studies (estimated $N \approx 55$).

Keywords: *human-robot interaction; companionship; non-verbal communication; verbal communication; trust-building; RoSAS; humanoid robots; social robotics*

1. Introduction

Human-robot interaction (HRI) has emerged as a multidisciplinary research domain at the intersection of robotics, cognitive science, and social psychology, with companionship increasingly recognised as a compelling application domain [1]. With the increasing use of robotic systems in home, healthcare, and education environments, the ability of robots to communicate their internal states (needs, intentions, and limitations) in a natural and comprehensible way to human collaborators is a key design challenge [3].

An excellent analogy to the process of domestication of animals, particularly dogs, is the development of non-linguistic communication as a means of sustaining deep affiliative bonds [2]. Dogs communicate needs to other dogs via postural cues, gaze redirection and directed locomotion, which are all understood by humans with a high degree of accuracy without common language. To play similar companionship roles, humanoid robots need to exhibit similar communicative competence in both gestures and speech. Although significant strides have been made in robot motion planning and speech synthesis, there has been surprisingly little empirical research that directly compares the two modes of communication, verbal and non-verbal, in the context of expressing needs and forming companions between robots. Holthaus et al. [3] offer a systematic classification of communicative robot signals, but they state that the effectiveness of gestures versus speech in building trust of the user is under-studied. Broadbent [6] has concluded that the nature of human reactions to robots is a manifestation of deep-seated intuitions of social agency in general, and that communicative modality may influence not only comprehension, but also perceived social attributes. In this study, we try to fill this gap by using a non-verbal gestural sequence and an explicit verbal utterance as tools to communicate a low-battery need state in the JD Humanoid Robot [4] and comparing the two. Operationalization of trust and social perception is done by means of the RoSAS [5]. The following research questions are asked:

- RQ1: How well can non-verbal gestural cues express robot intent, compared to vocal communication?
- RQ2: Do there exist differences between the effect of the verbal and non-verbal modality on perceived social attributes (competence, warmth, discomfort) of the RoSAS?

It is hypothesized that the accuracy of intention identification and clarity will be higher with verbal communication, and that verbal and non-verbal gestures will produce similar warmth and engagement. Results are presented as exploratory and effect-size estimates are given for the planning of further, larger-scale studies.

2. Related Work

2.1. Communication Modalities in HRI

In Dautenhahn's [1] work, effective communication is said to be a prerequisite for the integration of a social robot, as both verbal and non-verbal communication contributes to the users' evaluation of robot intelligence and trustworthiness. Holthaus et al. [3] give a formal taxonomy of communicative robot signals, which can be classified according to channel (visual, auditory), intentionality, and communicative function. They argue that their framework implies that when there is no prior salience of the object, the pointing gesture is inherently ambiguous in its reference. This forecast is directly tested in the current study.

Human mirror neuron responses are found to be triggered by movements similar to biological motion, as shown by neurocognitive evidence in Cross et al. [7] that suggests that human-like gestures are effective communicators because of an underlying perceptual mechanism. In contrast, a verbal communication mode that is text-to-speech (TTS) is informationally explicit but lacks biological motion cues that support intuitive social understanding.

2.2. Trust, Social Attributes, and Companionship

Broadbent [6] draws a conclusion from a large review of the work on HRI that human responses to robots are influenced by the mind, agency, and social role attributions made to robots. Carpinella et al. [5] operationalize these attributions with the validated

RoSAS, an 18-item scale with three orthogonal factors: Competence, Warmth and Discomfort. The scale is validated for different robot platforms and interaction settings.

Malle and Ullman [8] suggest a multidimensional model, which separates the performance-based trust (competence attributions) from values-based trust (warmth and ethical alignment). They suggest that the basis of performance-based trust is likely to be explicit information provision, whereas the basis of values-based trust is likely to be the perceived emotional resonance of non-verbal mimicry.

2.3. Non-Verbal and Animal-Inspired Communication

Larson and Bradley [2] report on dogs' communication of need states by gaze, locomotor approach, and object-directed pointing, which do not require shared language. Koay et al. [9] leverage this insight in their approach to narrative prototyping, which involves building ecologically valid HRI scenarios with a scripted series of behavioural sequences. The present study follows this approach, creating a standardised gestural narration to tell a story of low battery.

3. Materials and Methods

3.1. Research Design

A within-subjects counterbalanced design was employed, with all 20 participants exposed to both the non-verbal and verbal conditions. Presentation order was counterbalanced to control for order and learning effects, maximising statistical power given the constrained sample size. Ethical approval was granted by the University of Hertfordshire Ethics Committee (reference: cSPECS/CL/UH/05098). All participants provided written informed consent prior to participation.

Regarding the use of AI tools: during the preparation of this manuscript, the author used Claude (Anthropic, claude.ai) for the purposes of manuscript drafting and language refinement. The experimental data, statistical analysis, study design, and core intellectual content are entirely the author's own. The author has reviewed and edited all AI-assisted output and takes full responsibility for the content of this publication.

3.2. Participants

Twenty participants (10 female, 10 male; M_age = 23.4 years, SD = 3.2, range = 18-35) were recruited from the University of Hertfordshire via opportunity sampling. Exclusion criteria included prior professional experience in robotics or HRI research. No financial incentive was provided. The sample size is consistent with pilot and prototyping norms in HRI research [9,10] and is treated accordingly in all interpretations.

3.3. Apparatus

The JD Humanoid Robot (EZ-Robot Inc.) served as the experimental platform: a programmable biped humanoid equipped with servo-actuated arms, a head, and a built-in TTS module. A standard AC charger was placed at a fixed, visible position within the experimental space. All robot behaviours were fully scripted and standardised across sessions to eliminate between-session variability.

3.4. Experimental Conditions

Both conditions conveyed the same semantic content: the robot's battery was low and it required connection to the charger. Conditions differed only in communicative modality.

Condition A (Non-Verbal Gestural Phase). The robot: (1) waved to initiate joint attention; (2) locomoted towards the charger; (3) raised and extended its arm to point at the charger; (4) paused for approximately two seconds with a subtle head tilt; (5) repointed. No speech was produced. The sequence paralleled referential pointing behaviours documented in human-canine communication [2].

Condition B (Verbal Communication Phase). The robot waved to initiate joint attention, then produced the TTS utterance: "I'm running low on battery — could you please plug me into the charger?" No additional gestures were produced beyond the initial wave.

3.5. Measures

The RoSAS [5] was administered following each condition: 18 items rated on a seven-point Likert scale (1 = not at all, 7 = extremely), yielding Competence, Warmth, and Discomfort subscale scores (Table 1). A single-item clarity rating (1-7) assessed perceived intent clarity, and a forced-choice question recorded intent identification accuracy. Post-interaction semi-structured interviews were transcribed and analysed thematically.

Table 1. Robotic Social Attributes Scale (RoSAS) Categories and Constituent Items [5].

Competence	Warmth	Discomfort
Reliable	Organic	Awkward
Competent	Sociable	Scary
Knowledgeable	Emotional	Strange
Interactive	Compassionate	Awful
Responsive	Happy	Dangerous
Capable	Feeling	Aggressive

Competence and Warmth items index positive social attributes; Discomfort items index negative attributes. Each item rated 1-7 per condition.

3.6. Procedure

Sessions were conducted individually in a controlled laboratory setting. Participants received a standardised briefing explaining that they would observe a robot in two scenarios and complete questionnaires after each. The briefing was intentionally non-specific regarding communicative content to avoid demand characteristics. Following both conditions, a brief semi-structured interview explored impressions of the robot's communication style. Total session duration was approximately 25 minutes.

3.7. Statistical Analysis

Paired-samples t-tests compared within-participant RoSAS subscale and clarity scores across conditions. Cohen's d was computed as a measure of effect size. Intent-identification accuracy was compared using McNemar's test for paired proportions. Thematic analysis of interview transcripts followed Braun and Clarke's [11] six-phase procedure. Two independent coders achieved satisfactory inter-rater agreement (Cohen's kappa = 0.74) prior to consensus discussion.

4. Results

4.1. Intent Identification Accuracy

In the non-verbal condition, 16 of 20 participants (80%) correctly identified the robot's intent as a battery-charging request. In the verbal condition, 19 of 20 participants (95%) correctly identified the intent. McNemar's test indicated that this difference did not reach conventional significance ($p = 0.25$), consistent with ceiling performance in the verbal condition and the small sample size. All three non-verbal misidentifications were characterised by referential ambiguity — confusion about the specific object being indicated — rather than total communicative failure. Figure 1 summarises accuracy by condition.

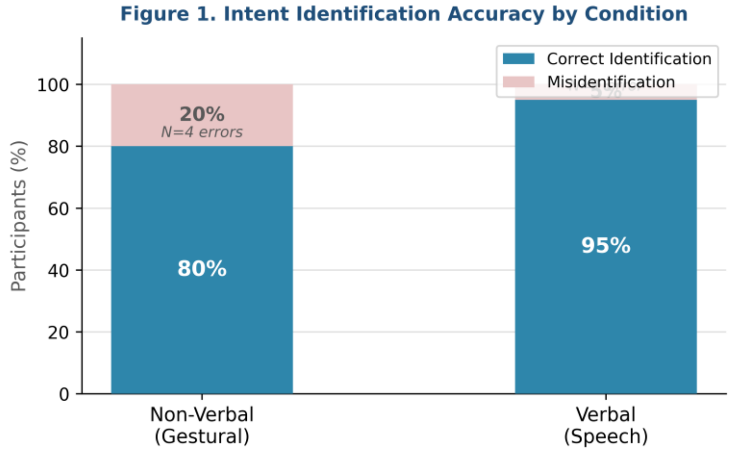

Figure 1. Intent identification accuracy by condition. Blue = correct identification; red = misidentification ($N = 20$ per condition).

4.2. RoSAS Clarity and Engagement Ratings

Paired-samples t-tests revealed a statistically significant difference in clarity ratings between the non-verbal ($M = 4.9$, $SD = 1.1$) and verbal conditions ($M = 5.8$, $SD = 0.9$; $t(19) = 2.61$, $p = 0.017$, $d = 0.58$), representing a medium-to-large effect. Engagement scores trended in the same direction (NV: $M = 4.7$, $SD = 1.0$; V: $M = 5.2$, $SD = 0.8$; $t(19) = 1.60$, $p = 0.120$, $d = 0.36$) but did not reach significance, indicating broadly comparable engagement across modalities. Results are summarised in Table 2.

Table 2. Summary of Quantitative Results Across Conditions.

Measure	NV: M (SD)	V: M (SD)	t(19)	p-Value
Clarity	4.9 (1.1)	5.8 (0.9)	2.61	**0.017** *
Engagement	4.7 (1.0)	5.2 (0.8)	1.60	0.120
Intent Identification (%)	80% (16/20)	95% (19/20)	—	—

NV = Non-Verbal condition; V = Verbal condition. * *$p < 0.05$. Paired-samples t-tests, $df = 19$.*

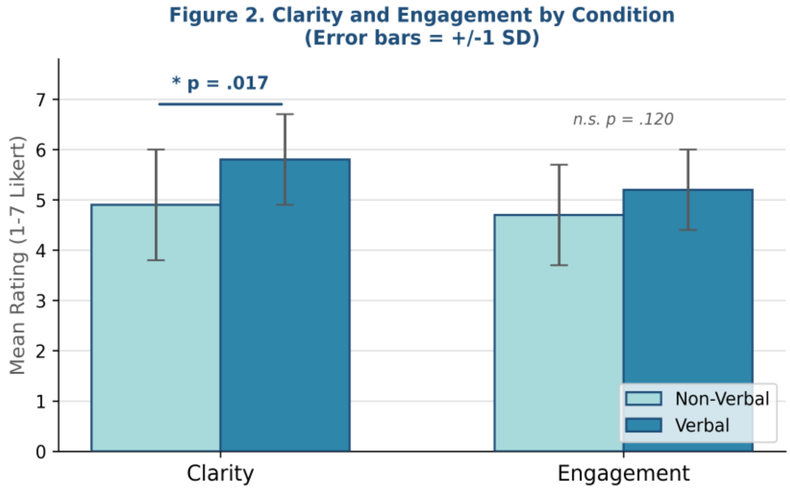

Figure 2. Mean clarity and engagement ratings by condition. Error bars = ±1 SD. Significance marker: * p = 0.017.

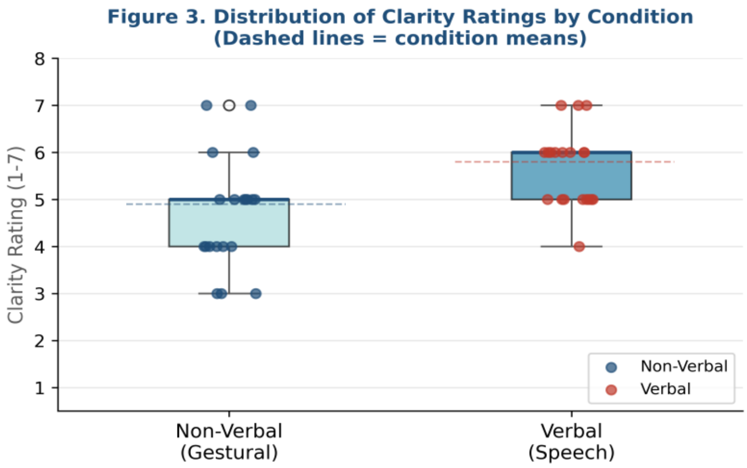

Figure 3. Distribution of clarity ratings by condition with individual data points overlaid (N = 20). Dashed lines indicate condition means.

4.3. RoSAS Subscale Profiles

Both conditions generated predominantly positive Competence and Warmth profiles, with low Discomfort scores across both modalities, indicating that neither communicative strategy generated appreciable threat perception. Warmth ratings were marginally higher in the non-verbal condition, consistent with biological motion warmth-attribution [7], though this difference did not reach significance. Figure 4 presents the radar profile comparison.

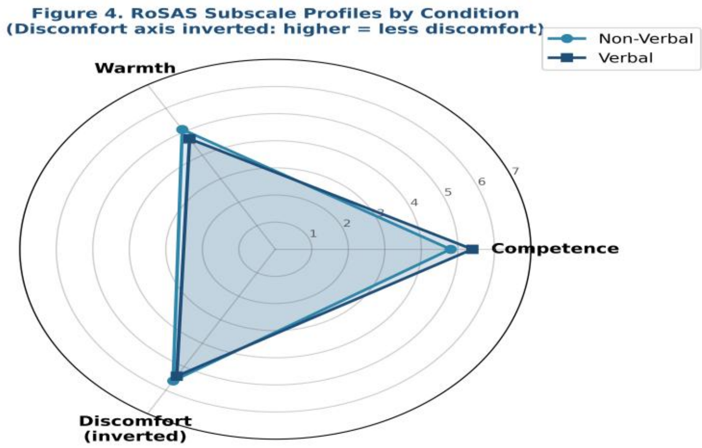

Figure 4. RoSAS subscale radar profiles by condition. Discomfort axis inverted (higher = less discomfort). Both conditions produced positive Competence and Warmth profiles.

4.4. Qualitative Themes

Theme 1: Trust Through Mimicry (Non-Verbal Condition). Participants who rated the non-verbal phase positively described the robot's gestures as "human-like" and attributed cognitive agency to the contemplative pause: "It was like it was thinking about how to tell me." This generated trust grounded in perceived shared intentionality rather than information transfer accuracy.

Theme 2: Clarity Through Speech (Verbal Condition). Participants consistently cited directness as a positive feature: "It just told me what it needed, no guessing." Minor criticisms of TTS voice quality ("the voice sounded a bit robotic, not warm") suggest that TTS affect may moderate verbal communicative effectiveness in longer-term interactions.

Theme 3: Complementarity of Modalities (Cross-Cutting). Across both conditions, multiple participants spontaneously proposed that combining gestural and verbal communication would be optimal: "If it pointed and then said what it wanted, that would be perfect." This unsolicited convergence constitutes ecologically valid support for the study's primary design recommendation.

5. Discussion

5.1. Verbal Superiority in Clarity, Gestural Parity in Engagement

The finding that verbal communication significantly outperformed non-verbal gestures in clarity ($d = 0.58$) is consistent with Holthaus et al.'s [3] prediction that directive gestures carry inherent referential ambiguity when object salience is not pre-established by context. All three non-verbal misidentifications involved failed joint attention to the charger object, suggesting that gesture-based communication effectiveness is context-contingent in ways that verbal communication is not.

The absence of a significant engagement difference ($p = 0.12$) is theoretically important. Engagement may be determined by broader interaction-scenario features held constant across conditions, rather than communicative modality per se. This supports the hypothesis that non-verbal cues achieve equivalent engagement despite lower clarity — a finding with independent value for companionship applications where sustained interaction quality matters as much as momentary accuracy.

5.2. Implications for Hybrid Communication Design

The Malle and Ullman [8] trust framework helps interpret the qualitative convergence on hybrid strategies: verbal communication primarily supports performance-based trust through accurate information transfer, while non-verbal gesture supports values-based trust through perceived warmth and intentionality. A hybrid strategy addressing both dimensions should produce robots that are both competent and warm — the combination most conducive to long-term companionship.

This recommendation aligns with the communicative layering principle in Holthaus et al.'s [3] typology, wherein multimodal redundancy reduces interpretation load. Companion robots should default to paired gesture-speech outputs when communicating need states, reserving purely gestural communication for contexts where speech would be disruptive.

5.3. Limitations

Several limitations constrain the generalisability of these findings. The sample ($n = 20$) is appropriate for an exploratory pilot study but insufficient for definitive inference; an estimated $N = 55$ would be required for 80% power to detect $d = 0.58$ at alpha = 0.05, two-tailed. The experiment was conducted in a single controlled laboratory with a single robot platform, and the low-battery scenario represents a narrow slice of companion robot communicative demands. Thematic findings should be treated as hypothesis-generating rather than confirmatory.

6. Conclusions

This study provides empirical pilot evidence that verbal communication significantly outperforms non-verbal gesture in clarity of intent conveyance ($t(19) = 2.61$, $p = 0.017$, $d = 0.58$), while engagement levels remain statistically comparable across modalities. Qualitative evidence consistently supports hybrid verbal-gestural communication strategies that exploit the clarity advantages of speech and the warmth-attributing advantages of human-like gesture simultaneously.

These findings extend Holthaus et al.'s [3] communicative typology with condition-controlled empirical estimates of modality-specific effectiveness in a companionship context, and offer practical design guidance for socially adept humanoid robot development. Future work should replicate these findings with larger and more diverse samples, across multiple robot platforms, and in longitudinal paradigms capturing the evolution of trust over time.

Author Contributions: Conceptualization, S.M.S.K.; methodology, S.M.S.K.; software, S.M.S.K.; validation, S.M.S.K.; formal analysis, S.M.S.K.; investigation, S.M.S.K.; resources, S.M.S.K.; data curation, S.M.S.K.; writing — original draft preparation, S.M.S.K.; writing — review and editing, S.M.S.K.; visualization, S.M.S.K. The author has read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by the author.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the University of Hertfordshire Ethics Committee (reference: cSPECS/CL/UH/05098).

Informed Consent Statement: Informed consent was obtained from all participants prior to participation in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon reasonable request.

Acknowledgments: The author thanks Prof. Patrick Holthaus (University of Hertfordshire) for supervision of the original MSc thesis research on which this study is based, and for his guidance on HRI research methodology. During the preparation of this manuscript, the author used Claude (Anthropic, claude.ai) for the purposes of manuscript drafting and language refinement. The author has reviewed and edited the output and takes full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HRI	Human-Robot Interaction
RoSAS	Robotic Social Attributes Scale
TTS	Text-to-Speech
NV	Non-Verbal
V	Verbal
SD	Standard Deviation
M	Mean

References

1. Dautenhahn, K. Socially intelligent robots: Dimensions of human-robot interaction. Philos. Trans. R. Soc. B 2007, 362, 679-704. https://doi.org/10.1098/rstb.2006.2004
2. Larson, G.; Bradley, D.G. How much is that in dog years? The advent of canine population genomics. PLOS Genet. 2014, 10, e1004093. https://doi.org/10.1371/journal.pgen.1004093
3. Holthaus, P.; Schulz, T.; Lakatos, G.; Soma, R. Communicative robot signals: Presenting a new typology for human-robot interaction. In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, Stockholm, Sweden, 13-16 March 2023; pp. 132-141.
4. EZ-Robot Inc. JD Humanoid Robot Kit. Available online: https://www.ez-robot.com (accessed on 1 January 2024).
5. Carpinella, C.M.; Wyman, A.B.; Perez, M.A.; Stroessner, S.J. The Robotic Social Attributes Scale (RoSAS): Development and validation. In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6-9 March 2017; pp. 254-262.
6. Broadbent, E. Interactions with robots: The truths we reveal about ourselves. Annu. Rev. Psychol. 2017, 68, 627-652. https://doi.org/10.1146/annurev-psych-010416-043958
7. Cross, E.S.; Hortensius, R.; Wykowska, A. From social brains to social robots: Applying neurocognitive insights to human-robot interaction. Philos. Trans. R. Soc. B 2019, 374, 20180024. https://doi.org/10.1098/rstb.2018.0024
8. Malle, B.F.; Ullman, D. A multidimensional conception and measure of human-robot trust. In Trust in Human-Robot Interaction; Nam, C.S., Lyons, J.B., Eds.; Academic Press: Cambridge, MA, USA, 2021; pp. 3-25.
9. Koay, K.L.; Syrdal, D.S.; Dautenhahn, K.; Walters, M.L. A narrative approach to human-robot interaction prototyping for companion robots. Paladyn J. Behav. Robot. 2020, 11, 66-85. https://doi.org/10.1515/pjbr-2020-0005
10. Salem, M.; Lakatos, G.; Amirabdollahian, F.; Dautenhahn, K. Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, USA, 2-5 March 2015; pp. 141-148.

11. Braun, V.; Clarke, V. Using thematic analysis in psychology. Qual. Res. Psychol. 2006, 3, 77-101. https://doi.org/10.1191/1478088706qp063oa