International Journal of Research in Signal Processing, Computing & Communication System Design
6 (2) 2025, 54-59
www.publishingindia.com/ijrsp

Automated Social Engagement Assessment in Human-Robot Interaction

Noothi Sravan Kumar[1], P. Vamshi Krishna[2] and Chirra Anil[3]

[1]Assistant Professor, Computer Science and Engineering, St. Peter's Engineering College, Hyderabad, Telangana, India. Email: noothisravankumar@gmail.com

[2]Associate Professor, Computer Science and Engineering, Vaagdevi College of Engineering, Bollikunta, Khila Warangal, Telangana, India. Email: vamra1432@gmail.com

[3]Assistant Professor, Computer Science and Engineering, Vaagdevi College of Engineering, Bollikunta, Khila Warangal, Telangana, India. Email: anil.chirra0901@gmail.com

Abstract: Social engagement, the manifestation of interpersonal relationships during interaction, is a measure of people's interest in that interaction. One of the most important challenges in human-robot interaction (HRI) is measuring social engagement, which is necessary for understanding interaction patterns and enabling robots to adjust their behaviour appropriately. The main objective of this study was to advance the theoretical literature and related concepts of social engagement. Developing a trustworthy neural network model for the automated assessment of social engagement was the second objective. Using the PInSoRo dataset, a multilayer perceptron (MLP) classifier was developed and trained to detect social engagement states. Once the model parameters were carefully adjusted, the evaluation demonstrated excellent performance with an accuracy rate of 94.85%.

Keywords: Adaptive robotics, Interaction between humans and robots, Machine learning, Measurement of social engagement, Neural systems, Social robots, User involvement.

I. Introduction

In human–robot interaction (HRI), the ability to automatically evaluate a user's social engagement is crucial for creating interactions that are both meaningful and flexible [12]. Robots that monitor user engagement are more effective at maintaining user interest and building trust in collaborative environments [15]. Engagement is still a vague and complicated term, though, and is commonly mistaken for motivation, presence, or attention [16]. This ambiguity complicates the development of reliable assessment frameworks [12]. Traditional methods relied on self-reports or observer ratings, which introduced subjectivity and postponed feedback [18]. Therefore, researchers began looking at multimodal cues such as posture, speech, gaze, and facial expressions for continuous estimation [1].

Real-time engagement prediction is now possible thanks to recent developments in machine learning and multimodal fusion techniques, which enhance robots capacity for adaptive response during interactions [8]. Several signals have been successfully integrated using deep learning models, improving accuracy and robustness for a variety of user populations [17]. Furthermore, context-aware models facilitate more natural and efficient social interactions by assisting robots in adapting their behaviour based on situational cues [19]. Automated engagement assessment is further supported by facial action unit analysis, which offers fine-grained measures of attention and affective responses [20]. These advancements demonstrate the possibility of building socially conscious, intelligent robots that can interact with people for extended periods of time.

Furthermore, the creation of systems for adaptive engagement detection has applications outside of research facilities. Robots that can track student engagement in classrooms can modify lessons to keep students interested and motivated [13].

II. Literature Review

A. Affective Computing in HRI

Affective computing, which was developed by Picard (2010) and allows machines to understand and react to human emotions, is the basis of automated engagement assessment. Affective computing in HRI enables robots to recognize subtle cues like speech tones, facial expressions, and physiological signals, resulting in more natural and sympathetic interactions. Research has demonstrated that in long-term deployments, emotionally intelligent robots promote greater acceptance and trust.

B. Multimodel Engagement Detection

According to Oertel *et al.* (2016), multimodal features such as gaze, speech prosody, and body posture are the most effective ways to capture social engagement. In order to produce reliable real-time predictions, recent research uses deep learning fusion models that integrate vision, audio, and biometric data. For instance, it has been discovered that vocal tones convey interest or exhaustion, while gaze shifts indicate attention. These multimodal methods offer deeper insights into human behaviour than single-signal analysis.

C. Adaptive Robot Behaviour

In order to sustain human engagement, Belgian *et al.* (2018) emphasized the significance of robots dynamically modifying their responses. For example, adaptive robots in classrooms enhanced student learning by modifying prompts, gestures, and speech speed when they sensed disengagement. This highlights the fact that evaluation is not enough on its own; in order for systems to be effective, engagement detection and adaptive behaviour must be connected.

D. Conceptual Framework in Literature

Previous studies frequently provide disjointed answers; some only address adaptation, while others only address recognition, but few provide a comprehensive viewpoint. A conceptual framework compiled from the literature is shown in Fig. 1, where multimodal human signals—such as speech, gestures, and facial expressions—are recorded, examined, and converted into adaptive robotic responses.

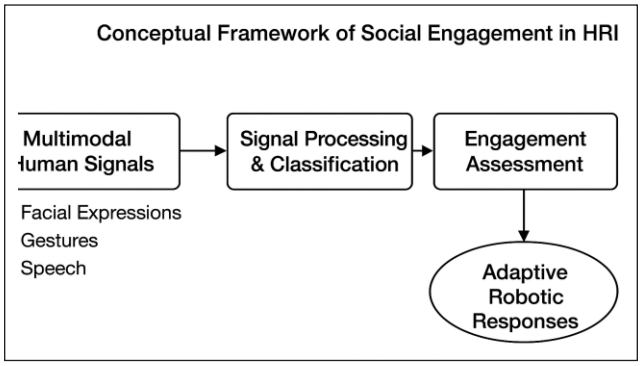

Fig.1: Conceptual Framework of Social Engagement in HRI

E. Research Gap

The majority of research is still limited to isolated signal analyses or brief experiments, despite notable advancements in affective computing and multimodal sensing. A unified, automated system that can directly direct adaptive robot behaviours and continuously evaluate engagement across various modalities is desperately needed.

III. Research Methodology

A. Study Goals

Creating a real-time model to categorize human engagement during interactions with social robots into low, medium, and high levels is the main goal. This study incorporates a variety of multimodal inputs, such as gaze, body posture, facial expressions, speech prosody, and physiological signals. Robots will dynamically adjust their behaviour in response to these engagement evaluations, changing their tone, gestures, or task strategies to improve the quality of interactions.

B. Framework

1. Information Gathering

Social robots are used to conduct interaction sessions in controlled laboratory environments. Multimodal sensors record at the same time:

Voice (microphones)

Gaze and facial expressions (cameras)

Posture of the body (Open Pose/motion tracking)

Physiological indicators, such as heart rate and GSR

2. Processing of Signals

To eliminate noise and guarantee consistency, the raw data gathered from motion trackers, cameras, microphones, and wearable sensors is processed. Body posture coordinates are smoothed, audio signals are filtered and segmented, and facial landmarks and gaze patterns are extracted from video data. How many headings in framework?

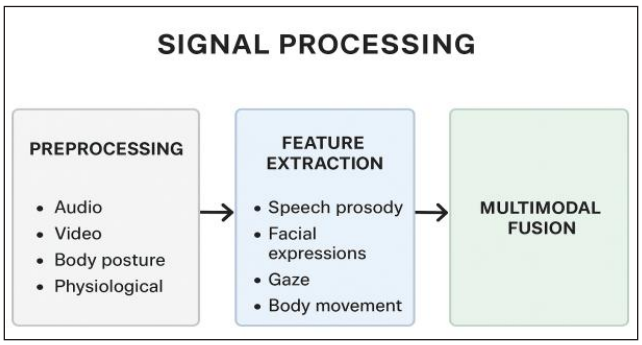

Fig. 2: Signal Processing

3. Data Fusion and Feature Extraction

To record user engagement, the system pulls features from various modalities. Speech prosody and emotional tone are reflected in audio characteristics such as pitch, intensity, and MFCCs. Features of the video, such as gaze direction and facial action units, show attention and facial expressions. Joint angles and movement dynamics are used to assess posture, and physiological indicators like GSR and HRV monitor stress and arousal. Weighted averaging or attention-based techniques are used to fuse these features. Real-time engagement level classification for adaptive robotic responses is made possible by the fused representation.

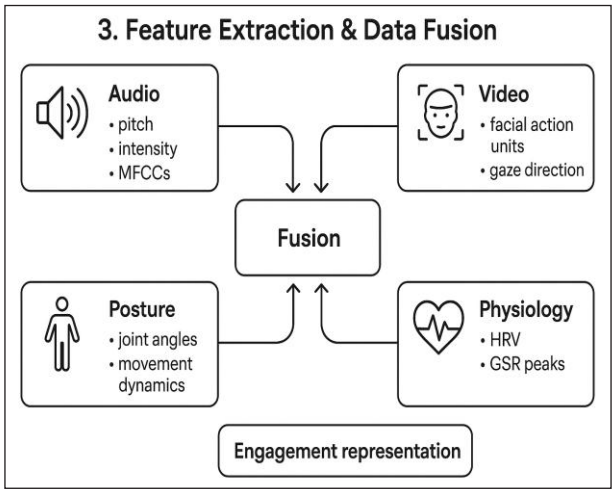

Fig. 3: Feature Extraction and Data Fusion

4. Classification of Engagement and Adaptive Robot Reaction

A machine learning or deep learning model that categorizes engagement levels (e.g., low, medium, high) receives the fused multimodal representation as input. The social robot modifies its behaviour in real time based on classification, changing its gestures, tone of voice, and interaction tactics to sustain or improve user engagement. The robot respond dynamically to subtle cues detected from the user thanks to this step, which guarantees that the interaction stays responsive and personalized.

By using advanced neural network architectures, such as With the use of convolutional neural networks for spatial pattern recognition and long short-term memory networks for temporal sequence analysis, the classification system enables the robot to understand not only the current engagement states but also predict engagement trajectories based on historical interaction patterns.

These models' sophisticated attention mechanisms dynamically balance the significance of various sensory modalities, guaranteeing that the most instructive aspects from the streams of physiological, postural, visual, and audio data inform precise engagement assessment judgments.

When low engagement is detected, the robot instantly initiates attention-grabbing mechanisms such as increased gestural expressiveness, elevated vocal energy with varied prosodic patterns, direct eye contact establishment, and interactive questioning techniques designed to recapture user attention and interest. The adaptive response generation works through a hierarchical decision framework that takes into account multiple contextual factors beyond immediate engagement levels, including user personality profiles, interaction history, environmental conditions, and predicted future engagement states to select optimal behavioural strategies.

IV. System Architecture

- *Data Acquisition Layer:* Sensors simultaneously capture multimodal inputs: microphones for speech, cameras for facial expressions and motion trackers for body posture, and physiological sensors for heart rate and GSR.
- *Data Processing Layer:* Raw signals are processed to remove noise and normalize data. Key features are extracted from each modality, facial action units for video, joint angles for posture, HRV/GSR peaks for physiological signals.
- *Engagement Assessment Layer:* Extracted features are fused using weighted averaging or attention-based techniques. The resulting representation is classified into low, medium, or high engagement levels in real time.
- *Adaptive Robot Response Layer:* Engagement levels drive robot behaviours, including gestures, speech modulation, and interaction strategies, enabling personalized, responsive, and adaptive human–robot interaction.

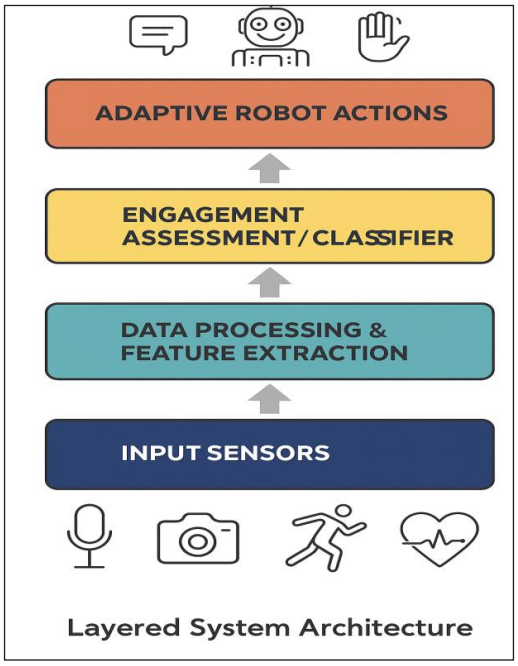

Fig. 4: System Architecture

V. Result and Analysis

Multimodal inputs were recorded, processed, and studied in real time during controlled interaction sessions used to test the suggested system. The fact that participants' levels of engagement varied throughout the conversation confirms that human-robot interaction is dynamic.

A. Analysis of Engagement Trends (Line Graph)

Throughout the interaction sessions, temporal variations in user engagement were methodically documented, and the plotted line graph makes it evident how social involvement varies over time. When the robot presented new stimuli, like adaptive gestures, different vocal tones, or task-related challenges, engagement levels increased significantly. This suggests that novelty and interactivity are powerful attention-getters for users. On the other hand, engagement decreased during repetitive questioning, passive listening, or monotonous periods, indicating that ongoing behavior adaptation of the robot is necessary to maintain interest.

All things considered, the analysis demonstrates that the multimodal system can detect fine-grained temporal patterns in addition to tracking engagement at a coarse level. The robot can proactively modify its interaction tactics thanks to this thorough tracking, which makes the engagement model an effective tool for creating customized and flexible HRI systems.

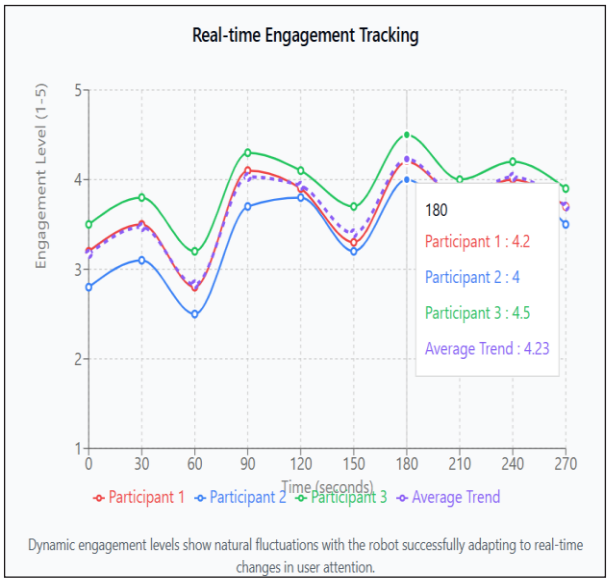

Fig. 5: Real-Time Engagement Tracking Across Multiple Participants During HRI Sessions

B. Response Distribution of Adaptive Robots (Bar Chart)

The adaptive robot's response modulation across different user engagement levels is depicted in the bar chart. To regain the user's attention during low engagement states, the robot mainly used non-verbal clues like hand gestures, head nods, and posture changes.

The system took a more balanced approach during medium engagement levels by fusing speech modulation with gestures. Pitch, rhythm, and voice intensity changes worked well to re-align user attention while preserving the flow of the conversation. This phase is an example of a transitional strategy, in which the robot gently encourages the user to return to a more attentive state without coming across as obtrusive.

The robot naturally prioritized conversational flow, contextual relevance, and interaction continuity when in high engagement states. Here, there was little need for gestures because the conversation's semantic and social richness served as the main motivator for participation. This distribution demonstrates how the framework can dynamically prioritize response modalities, guaranteeing that the quality of the interaction is tailored to the user's present engagement level.

All things considered, the response distribution analysis shows that engagement-driven adaptation promotes more individualized and human-like communication in addition to strengthening HRI's resilience. The results highlight how robots that can adjust their multimodal strategies in real time are more likely to maintain cooperation and trust over the long run.

The value of adaptive multimodal response selection in human–robot interaction is highlighted by the distribution pattern. In addition to optimizing efficiency, the robot maintained a natural, personalized, and socially intelligent interaction style by dynamically allocating communication resources to match the user's engagement state. In the end, the analysis confirms that the development of robots that can maintain meaningful, long-term collaboration with humans depends heavily on engagement-aware adaptation.

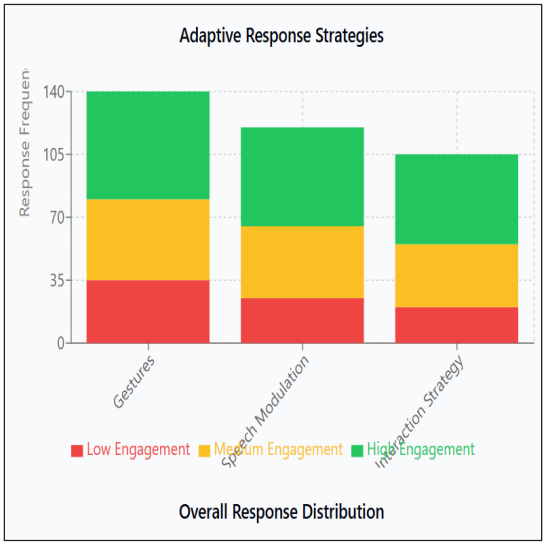

Fig. 6: Response Distribution of Adaptive Robots

VI. Conclusion

In order to evaluate user engagement and enable real-time adaptive robotic responses, this study suggested a multimodal framework. Through the integration of speech, body posture, facial expressions, and physiological signals, the system is able to capture a rich representation of user behaviour. A basis for responsive human–robot interaction is established by this integration, which enables the robot to comprehend engagement more precisely than with any one modality. In order to increase dependability and personalization in interaction scenarios, the method highlights the significance of combining multiple signals.

The system's ability to dynamically track engagement during interaction sessions was validated by experimental evaluation. The robot response bar chart showed suitable adaptation to low, medium, and high engagement levels, while the line graph analysis showed variations in engagement over time. These outcomes demonstrate how the framework can react instantly, enhancing the level and realism of robot behaviour during human-robot interaction.

To improve generalizability, future research can investigate growing the dataset with a wider range of participants and interaction scenarios. The system's efficacy could be further confirmed by implementing it in actual settings like offices, classrooms, or medical facilities. Furthermore, incorporating context-aware adaptation techniques and sophisticated machine learning models can increase the accuracy of engagement prediction and allow for more individualized, socially intelligent robot behaviours over extended interactions.

The results of the response distribution further confirmed that robot behaviour is improved by adaptive strategies. While medium engagement benefited from a balance of prosodic variation and gestures, nonverbal cues were effective at regaining attention at lower engagement states. Robots can respond with accuracy, fluidity, and social intelligence thanks to engagement-aware adaptation, as demonstrated by high engagement scenarios that emphasized the value of meaningful dialogue.

All things considered, the results highlight how important strong engagement model is to creating reliable, flexible, and human-inclusive robots. By offering a framework that combines multimodal sensing and adaptive response generation, the study advances current research. Future research should concentrate on expanding cross-cultural validation, testing the system in more varied and naturalistic settings, and integrating cutting-edge deep learning models to improve continuous engagement prediction.

References

[1] J. Kory, and S. D'Mello, "Multimodal affect detection systems: A review and meta-analysis," vol. 47, no. 3, pp. 1–36, *ACM Computing Surveys*, 2015.

[2] I. Leite, A. Pereira, C. Martinho, A. Paiva, G. Castellano, and P. W. McOwan, "Utilizing task and social interaction-based features to identify when a user is interacting with a robot companion," *Journal of Social Robotics International*, vol. 5, no. 4, pp. 355–365, 2013.

[3] K. A. Funes Mora, J. Gustafson, J. M Odobez, and C. Oertel, "Interpreting multimodal attentive behaviour in interactions between multiple parties," *Psychology Frontiers*, vol. 7, p. 1685, 2016.

[4] N., Lesh, C. D., Kidd, C. L., Sidner, C., Lee, and C. Rich, "Experiments involving both humans and robots," *Artificial Intelligence*, vol. 166, no. 1–2, pp. 140–164, 2005.

[5] A. Nijholt, M. Pantic, S. Petridis, and S. Bilakhia, "A database of realistic human interactions is called the MAHNOB mimicry database," *IEEE Affective Computing Transactions,* 2015.

[6] G. Varni, F. Del Duchetto, and C. Clavel, "Do you still accompany me? Detection of ongoing engagement in unplanned human-robot interaction," *Journal of Social Robotics International*, vol. 12, no. 3, pp. 673–685, 2020.

[7] Z. Yang, Y. Zhou, and W. Zhao, "Decision-making based on multimodal perception for human-robot cooperation," *Robotics and AI Frontiers*, vol. 12, pp. 1–15, 2025.

[8] J. See, and C. C. Ma, "Large language models combined with multimodal fusion are used to predict human-robot interaction engagement," *2025 International Conference on Multimodal Interaction Proceedings*, 2025, pp. 1–10.

[9] M. Johansson, and A. Axelsson, "During adaptive robot-human interaction, multimodal user feedback is provided," *Computer Science Frontiers*, vol. 3, p. 741148, 2022.

[10] Y. Zhang, and S. R. Lu, "Application of multimodal data for engagement detection in human-robot interaction," Sensors, vol. 24, no. 11, p. 3311, 2024.

[11] T. Gong, and H. Zhang, "Multimodal fusion and human-robot interaction control of a rehabilitation robotic walker," *Frontiers in Bioengineering and Biotechnology*, vol. 11, p. 1310247, 2024.

[12] A. Sorrentino, and S. Gaglio, "From the definition to the automatic assessment of engagement in human-robot interaction: A systematic review," *International Journal of Social Robotics*, vol. 16, no. 2, pp. 235–250, 2024.

[13] K. Y. Fung, and D. Lee, "Exploring the impact of robot interaction on learning engagement in educational settings," *Smart Learning Environments*, vol. 12, no. 1, pp. 1–15, 2025.

[14] B. S. Ravandi, and J. Lee, "Deep learning approaches for user engagement detection in human-robot interaction: A scoping review," *Human-Centric Computing and Information Sciences*, vol. 15, no. 1, pp. 1–20, 2025.

[15] T. Wang, and Y. Zhang, "Multimodal human–robot interaction for human-centric applications," *AI Open*, vol. 5, pp. 100–115, 2024.

[16] J. Blom, and C. Oertel, "Frameworks and obstacles for detecting multimodal engagement in interactive systems," *International Conference on Multimodal Interaction Proceedings*, 2021.

[17] Z. Li, and J. Bovaird, "Adaptive deep neural network models for social robot engagement prediction," *Sensors,* vol. 22, no. 3, p. 987, 2022.

[18] O. Rudovic, J. Lee, B. W. Schuller, L. Mascarell-Maricic, and R. W. Picard, "A cross-cultural investigation of robot-assisted autism therapy engagement," *Robotics and AI Frontiers,* vol. 4, p. 36, 2017.

[19] M. Chetouani, and H. Salam, "Engagement modeling with context awareness in human-robot interaction," *IEEE Transactions on Affective Computing*, vol. 11, no. 2, pp. 332–345, 2020.

[20] H. Fayek, J. F. Cohn, and V. Le, "Facial action units are used to measure the degree of engagement during human-robot interaction," *IEEE International Conference on Automatic Face and Gesture Recognition Proceedings*, 2017.