Good Looking: How Gaze Patterns affect Users' Perceptions of an Interactive Social Robot

Vidya Somashekarappa[1], Christine Howes[2], and Asad Sayeed[3]

Abstract— **Gaze automation in social robots is pivotal for enhancing Human-Robot Interaction (HRI) by promoting engagement, intuition, and effectiveness in communication. This paper investigates whether different gaze patterns from a Furhat robot can lead to more effective, natural and engaging interactions. Our results indicate that gaze manipulations based on gaze patterns from human-human interaction positively impact user perceptions compared to the neutral and random conditions. Participants rate the anthropomorphism and animacy of the robot in the experimental condition. The findings contribute to understanding the impact of robot gaze on user perceptions and engagement, offering insights for the design and improvement of interactive social robots.**

I. INTRODUCTION

Gaze automation in a social robot is a crucial aspect of Human-Robot Interaction (HRI) that aims to make the robot more engaging, intuitive, and effective in its communication with humans. The ability of a robot to control its gaze —-where it looks, how it looks, and when it looks- is essential for establishing a natural and meaningful interaction with users [1], [2], [3]. The main purpose of this research is to assess different patterns of gaze in robots in order to contribute to a more effective, natural, and engaging interaction. **How does human-like behavior of a robot influence people's perception of it?**

Maintaining appropriate eye contact helps the robot establish a connection with users, fostering a sense of engagement, trust, and rapport by directing attention towards users or relevant objects, signaling engagement and facilitating more focused interactions. Gaze behavior serves as a powerful nonverbal communication tool, allowing robots to convey intentions, emotions, and social cues [4], [5], and using these gaze cues to signal turn-taking in conversations, facilitating more natural and intuitive interactions. Gaze automation enables robots to provide visual feedback, such as nodding or looking towards objects, to confirm understanding or indicate agreement and enables better task performance by focusing attention on critical elements, reducing distractions, and improving task efficiency [6].

The main objectives and contributions of the paper are:

1) Implementing different gaze patterns in a social robot in an interactive setting
2) Experimental evaluation of robot gaze patterns in human-robot interaction
3) Analysis of how different gaze patterns may be predictive of people's engagement in an interaction with a social robot

In this experimental study, we investigate gaze patterns as a potential continuous metric to gauge people's perceptions of robots. Our specific research question are i) Is there any correlation between robot gaze patterns participants' perception of its friendliness, cooperativeness and sociability? ii) Does the robot's gaze pattern significantly influence users assessments of the robot's usability and overall interaction experience? iii) If so, how do specific gaze patterns indicate the emotional engagement of the user during the interaction?

II. CORRESPONDING WORK

Research has shown that a robot consistently maintaining the user's mutual attention is viewed as more genuine [7]. Similarly, leveraging the speaker's visual attention through gaze-tracking positively impacts understanding speech, aiding in the prediction, clarification, and resolution of spoken references [8], [9]. Mutlu et al., delved into the significance of mutual attention within collaborative scenarios, where both humans and robots engage in coordinated tasks within a shared environment [10]. Additionally, they explored gaze strategies employed to structure dialogs and define distinct roles of speakers and listeners in human-robot interactions.

The computational frameworks developed for recognizing and facilitating engagement, focus on "the method by which multiple participants establish, sustain, and conclude their perceived bond during shared activities" [11]. This research highlights the influence of targeted and mutual gaze on the relational dynamics between interacting parties but doesn't address the role of gaze in clarifying speech or using gaze to prompt responses.

While certain investigations have focused on general modelling approaches to integrate various modes and dialog reasoning for multi-modal interfaces, embodied conversational agents, and human-robot interactions, they often overlook key concepts such as joint attention, mutual attention, or engagement [12]. The current paper builds upon insights from existing literature but advances the field by incorporating different gaze aspects into a unified and innovative modelling framework. The model for our experimental condition is based on detailed observations of gaze behaviors in human-human interactions, made from the manual annotation and

*This work is supported by the Swedish Research Council

[1]Vidya Somashekarappa, PhD Fellow at CLASP, University of Gothenburg, Sweden vidya.somashekarappa@gu.se

[2]Christine Howes, Professor at CLASP, University of Gothenburg, Sweden asad.sayeed@gu.se

[3]Asad Sayeed, Associate Professor at CLASP, University of Gothenburg, Sweden christine.howes@gu.se

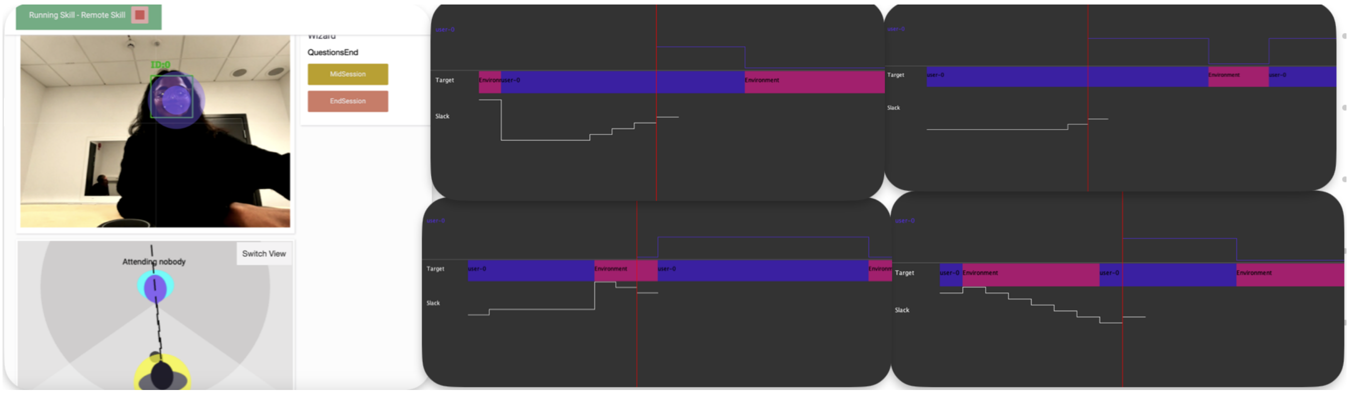

Fig. 1. Event Attention Interface (EAI) showing the gaze model state transitions of the robot in the interaction space. *Blue* block indicates gaze on the user and *pink* block in the surrounding environment. The *blue* horizontal line depicts the user's gaze on the robot and the *white* horizontal line is the beginning and the end of speech turn of the robot.

analysis on the Good Housekeeping Institute (GHI) corpus [13].

Humans employ diverse communication modalities to convey information, while selecting channels based on their efficiency and communicative potential [14]. To disambiguate, individuals rely on their counterparts' ability to integrate information from these channels for clarity. Solely relying on verbal expressions can introduce uncertainties, potentially disrupting mutual understanding [15]. Consequently, listeners often integrate a speaker's verbal content with accompanying gestures and eye movements to derive a clearer interpretation before seeking clarification [16]. Notably, when using gaze during speech, the visual cue typically precedes speech by approximately 800-1000 milliseconds, while individuals tend to focus on the referenced object about 200 milliseconds after auditory input [17].

In collaborative efforts, individuals strive to direct their partners' attention either to specific objects in their environment or to themselves. In addition to verbal and pointing cues, they commonly use gaze and a combination of these communication methods to achieve this goal. Furthermore, individuals track their partner's gaze to establish a shared frame of reference, leading to synchronized focus and mutual attention on a particular object. This interaction signifies active engagement in the joint activity and the ability to discern indicated items. When both parties maintain this focused interaction, they engage in mutual gaze [18]. Both gaze mechanisms, aimed at fostering joint attention, play a crucial role in preserving shared understanding. In our robotic application, the system draws attention to the speaker based on the demands of the conversational act such as maintaining attention while the speaker is looking at the robot or looking away during a long utterance to avoid discomfort. Effective communication relies on skillful mechanisms that regulate the roles of speakers and listeners [19]. In this context, the orientation of gaze becomes a pivotal cue in either facilitating or constraining these role transitions. Generally, speakers divert their gaze from the audience, signaling their desire to maintain control of the conversation, whereas they redirect their focus toward another participant to relinquish the floor [20]. If a statement does not conclude with a gaze directed towards another individual, the transition between speakers may be prolonged.

Despite achieving good results in replicating human-like gaze behaviors in robots, a prevalent constraint is their predominantly reactive nature. Although certain systems devise gaze behavior plans for forthcoming utterances at the onset of speech (e.g., [21]), these plans lack incremental updates and do not significantly influence the ongoing gaze behavior. Another common limitation is the static nature of many systems, utilizing fixed duration for gaze shifts. For instance, in [22], the robot's gaze remained fixed on the relevant target for 1-5 seconds during interactions before transitioning to the target with the lowest priority. In contrast, Human-Human Interaction (HHI) entails extensive planning. Studies indicate that gaze behavior is intricately coordinated with the underlying speech plan [23]. The duration of planning determines whether a swift glance suffices or necessitates head movement for a more extended gaze.

In summary, prior research has explored gaze models for robots, but it has often overlooked the consideration of the human partner's eye movements. The predominant gaze behavior of existing robots is designed in response to users' speech, or gaze behavior is accomplished through head movements. However, head movements are limited to approximating coarse gaze direction and cannot effectively convey nuanced eye movements. Given the identified gaps in existing research, the present study addresses this gap by directing attention toward the intricate interplay between human and robot gaze dynamics. Our focus is to discern how incorporating a more comprehensive understanding of human eye movements into the design of robotic gaze behaviors could enrich the overall human-robot interaction experience. This research is aimed to contribute valuable insights into refining robotic gaze models, moving beyond traditional constraints, and fostering a more natural and intuitive communication between humans and robots.

Fig. 2. *Session in progress*. Experiment setting showing cameras placed behind the users and Furhat during the interaction. Images to the left show various positioning of gaze movements of the robot in accordance with the conversation flow (consent has been provided my the participants to use the images from the experiment for publishing purposes with faces being swapped/anonymized.)

III. Human to Robot interaction setup

This section describes the implementation of our system that enables real-time gaze interaction with a social robot head. Furhat[1], is an anthropomorphic robot equipped with a *Software Development Kit (SDK)*, which provides tools tailored conception, deployment, and analysis of applications. It features a biomimetic neck design facilitating lifelike head movements, comprehensive control over facial expressions, gestures, and ambient lighting. The platform supports customization, enabling adjustments to facial characteristics, ethnicity, gender, multilingual modality, and even species, with adaptable faces securely attached through magnetic mechanisms.

The primary components of Furhat's programming infrastructure encompasses development of skills using *Kotlin API*, with integration into python for object detection. The skill framework constitutes an advanced layer building upon rudimentary I/O capabilities, allowing *Natural Language Understanding* (NLU), dialog management, multimodal utterances, interaction logging, and the incorporation of *Graphical User Interfaces (GUI)*. The facial behavior is controlled by a 3D face model which is similar to that of virtual agents.

The robot can execute rapid gaze shifts through digital animation and incorporates physical servos in its neck to mimic head movements. Convincing neck and eye gaze behavior are crucial for our specific task, where users are expected to assess the robot's visual behavior. The robot achieves precision in looking at various parts of the lab by standing in the fixed position, and is calibrated to ensure accurate gaze shifts towards specific locations.

A. Interaction session

The experiment is a within-subject design where the participants interacted with the robot in 3 consecutive sessions. The social robot displayed three variations of gaze behavior: neutral, experimental and random (see section IV-B for details). The participants had zero exposure to the robot prior to the experiment. The social interaction sessions lasted about 30-40 minutes. Each session was programmed to take up to 10 minutes where the robot asked questions and the user was requested to briefly answer at their own pace while the robot maintained expressive gaze movements throughout the experiment. Finally, at the end of each session two questionnaires were provided to measure user engagement and perception of the robot.

B. Participants

21 participants between the age of 25 to 48 with the average age of 36.5 years were recruited (M=13; F=7; Non-Binary=1). They were either first or second language English speakers, with a minimum of undergraduate education. Prior to the main session we conducted a pilot study on 5 individuals. At the beginning of the session, participants were presented with information about the study and provided their informed consent. The study was approved by the Swedish Ethical Review Authority[2].

IV. Experiment and evaluation

The participants were seated approximately 150 cm away from the robot. The user and Furhat, were centrally aligned. We adjusted the participants' sitting height to guarantee that their eyes are approximately the same level as the robot's eyes. Our baseline condition is using continuous eye contact as similar to gaze behavior always used in existing robotic systems: the robot attends to users' face when they are facing towards the robot. The gaze shift is generated via eye movement. The robot's neck moved to track the user during the experiment along with their gaze. Furhat's blink was consistent across all experiments.

Prior to the study, the experimenter explained the purpose and procedure of the study to the participants. The entire experiment lasted for around 30 minutes per participant. During each session, the participants had three interactions with the robot (one in each of the gaze behavior conditions; neutral, experimental and random, with the order randomized between participants). In each interaction, the participants were asked to answer six unique questions. After each interaction, each participant filled in a questionnaire to assess their perceptions of the robot's behavior. The Perception

[1]https://furhatrobotics.com/

[2]`https://etikprovningsmyndigheten.se/`, 2023-03044-01.

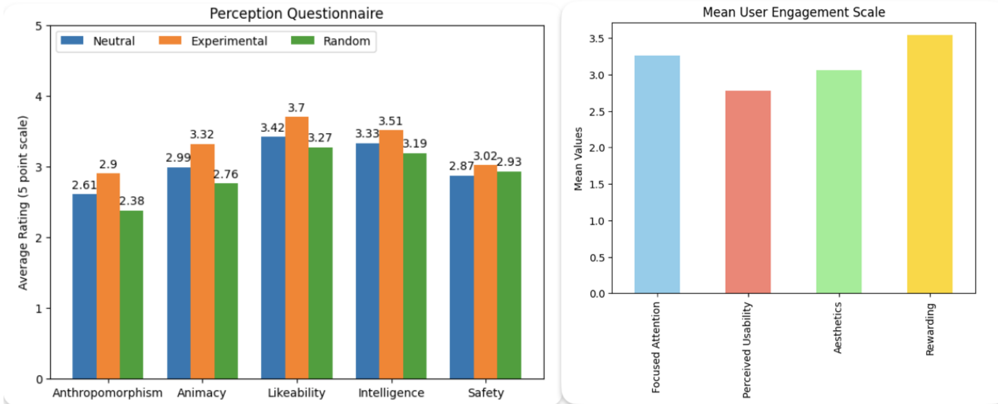

Fig. 3. Questionnaire Analysis. *Plot 1*: Perception Questionnaire, *Plot 2*: User Engagement Scale

Questionnaire consists of questions to assess users' perceptions of the robot's anthropomorphism, animacy, likeability, intelligence and safety, with each of these factors based on 3 to 5 sub-questions rated on a five point Likert scale (e.g. the subscales related to anthropomorphism included items such as rating the robot on a scale that ranged from 'fake' (1) to 'natural' (5) and 'artificial' (1) to 'lifelike' (5).

After all three interactions, the subject filled in a user engagement scale questionnaire about their overall experience and feedback. This questionnaire consists of 12 statements to be assessed in a 5 point Likert scale from "strongly disagree" (1) to "strongly agree" (5), with each statement relating to one of four factors. The four factors are 'Focused Attention (FA)', 'Perceived Usability (PU)', 'Aesthetic Appeal (AE)', and 'Reward Factor (RF)' [24]. Examples of items relating to perceived usability for example, are 'I felt frustrated while talking to Furhat.' and 'I found Furhat confusing to use.´

Furhat, first greeted the participants and asked them open ended unique questions during each session which were consistent across subjects. There were a total of 18 questions and the robot answered these questions briefly as well while the interactant was asked to observe the naturalistic behavior of the robot.

A. Measures

Users' subjective understanding of each interaction was assessed using the responses to the perception questionnaires. We further assessed users overall experience by analysing the responses to the user engagement scale. For both of these scales we use only average figures aggregated over the sub-items for each factor.

Figure 3 (Plot-1) represents the analysis of a Perception Questionnaire comparing ratings across different attributes ('Anthropomorphism', 'Animacy', 'Likeability', 'Intelligence', 'Safety') under the three experimental conditions: 'Neutral', 'Experimental', and 'Random'.

B. Event Attention Interface

In the *neutral* condition the system tracks the face with the users movement and blinks at regular intervals similar to the other conditions but does not react to the user's gaze. Instead the robot's attention was directed towards the user's face position. A wait key was initialized in all conditions in case of delayed response from the user. In case of no reply, the robot repeated the question and if the user did not want to share or answer during the conversation, by default Furhat moved on the next question after a bridging sentence. This paradigm was to make sure the conversation flow did not affect the perception of the robot since the main focus in to access the non-verbal cues.

In the *experimental* condition, the gaze patterns were dynamically programmed so as to follow the verbal cues with semantic and pragmatic information, in line with evidence from gaze research in human-human interaction. The completion of the syntactic unit by the user was a cue to break mutual gaze, while the end of sentence completion of the robot was a cue to move gaze back to the listener which are cues for turn-yielding and turn-holding [25]. Gaze aversions are used to reduce cognitive effort and modulate intimacy, hence the robot looked away during the speaking turn [26]. During a longer speaking turn, the robot looked away in the environment for one second before (randomly initialized) to avoid eerie mutual attention (figure 1, block 4) [13]. In order to avoid overlap, if the participants began speaking before or after turn completion, Furhat paused until the user finished their turn.

In contrast, the *random* condition was designed to randomly trigger the gaze of the robot without any predetermined contextual padding. In this case we implemented the same basic gaze movements as in the experimental condition, but the initiation of the gaze behaviors was not directly tied to the interactive context. For example, gaze could be directed towards the environment (and away from the speaker) during a users speech, towards the speaker for an indefinite length of time or the robot could appear to look away in the middle of a sentence.

In order to understand and visualize the dynamic gaze events in random and experimental conditions we utilize the *Event Attention Interface (EAI)*. Figure 1 shows the platform of the robots interface, indicating the attention on the user to the left. The graphical representation depicts the generated gaze behavior of Furhat during interactions and also the status of users gaze focus on or away from the robot. During the transitions, the robot maintained mutual attention towards the user while they glanced away frequently. Similarly, the robot actively broke the eye contact when the user looked steadily at the robot.

V. Questionnaire Analysis

Participants' feedback on the overall interaction experience and the impression of the robot is presented in figure 3.

Attributes Comparison: The graph in figure 3 plot 1 shows the average ratings for each attribute across the three experimental conditions.

As can be seen in figure 3, the *experimental* condition, based on human gaze behavior, consistently leads to higher ratings across all attributes, suggesting that the experimental natural gaze manipulations positively impact user perceptions. The *random* condition generally falls below the *experimental* condition but shows comparable ratings to the *neutral* condition. This might indicate that random gaze factors have a less pronounced impact on perceived attributes.

We ran a series of Generalized Linear Mixed Models (GLMMs) with each of the five attributes as dependent variable, gaze pattern type (neutral, experimental or random) as independent variable, participant ID as a within-subject factor and age and gender as random effects. Post hoc pairwise comparisons were carried out in the case of significant effects to identify which differences were significant.[3]

For *anthropomorphism*, there was a significant main effect of gaze pattern type ($F_{2,60} = 5.681, p = 0.006$). Post hoc pairwise analyses showed that the rating for anthropomorphism was significantly higher in the experimental than random condition ($t = 3.362_{60}, p = 0.001$) and marginally higher than the neutral condition ($t = 1.895_{60}, p = 0.063$). Neutral and random conditions were not significantly different from eah other ($t = 1.46_{60}, p = 0.148$).

For *animacy*, there was a significant main effect of gaze pattern type ($F_{2,60} = 15.666, p < 0.001$). Pairwise tests showed that animacy was rated significantly higher in the experimental condition than the neutral condition ($t = 3.283_{60}, p = 0.002$) and the random condition ($t = 5.568_{60}, p < 0.001$). Furthermore, the neutral gaze pattern was rated as significantly more animated than the random gaze pattern ($t = 2.284_{60}, p = 0.026$).

There was also a significant main effect of gaze pattern type on *likeability* ($F_{2,60} = 8.183, p < 0.001$). Likeability was rated significantly higher in the experimental condition than both random ($t = 3.990_{60}, p < 0.001$) and neutral ($t = 2.572_{60}, p = 0.013$). Likeability was not rated significantly differently in the random and neutral conditions ($t = 1.419_{60}, p = 0.161$).

The same pattern of effects was found for the ratings of *intelligence*, with a significant main effect of gaze pattern type ($F_{2,60} = 8.183, p < 0.001$), a significant difference between the experimental and random ($t = 3.883_{60}, p < 0.001$) and experimental and neutral ($t = 2.142_{60}, p = 0.036$) conditions and no significant difference between random and neutral gaze patterns ($t = 1.691_{60}, p = 0.096$).

There was no significant main effect of gaze pattern on participants' perceptions of *safety* ($F_{2,60} = 1.485, p = 0.235$).

With the exception of *safety*, therefore, user's perceptions of the robot were higher for all our measured attributes in the experimental condition than in the other two conditions, highlighting the importance of gaze behavior in users perceptions of a social robot in interaction.

User Engagement Scale: Utilizing a structured User Engagement Scale (fig 3, plot 2), four key aspects were examined: 'Focused Attention,' 'Perceived Usability,' 'Aesthetics,' and 'Rewarding.' Each aspect was assessed based on user-provided ratings, and the mean values were calculated to discern the overall perception of users. Each of these contain three sub-statements. Participants expressed a moderate level of engagement in terms of focused attention (Mean Value = 3.19). Perceived usability and aesthetics yielded a moderate mean value of approximately 2.84 & 3.0. This indicates a generally satisfactory level of usability, hence requires potential enhancements in the user experience. The aspect of 'Rewarding' exhibited a relatively higher mean value of approximately 3.5 where participants found the interaction to be rewarding, indicating a positive and fulfilling experience. Note that as we only asked our subjects these questions once after all three of their interactions, further between-subject experiments are needed to assess whether these engagement factors are also positively impacted by more human-like gaze behavior in the robot.

VI. Discussion

This study highlights the importance of human-like gaze behavior in positively influencing user perceptions, especially in the context of an anthropomorphic social robot. Tying the gaze behavior to the context of the interaction in progress, by looking away and returning the gaze to the interlocutor in line with how humans tend to do this in relation to turn-taking interaction and not continually gazing at one's interlocutor as is often the case in social robots, positively impacts on participants' perceptions of anthropomorphism, animacy, likeability and intelligence, but has no impact on participants' perceptions of safety. Perhaps surprisingly there was no difference in participants' perceptions between the neutral and random conditions, except in the animacy case where the neutral gaze behavior (which does not vary as much) was rated as *more* animated than the random gaze behavior. This suggests that care must be taken when implementing particular behaviors in social robots as bad algorithms might actually be worse than doing nothing.

One factor we have not yet addressed in this research is the question of how people's perceptions of robots evolve over repeated interactions, which requires the creation of measurement methods suitable for extended evaluations. Currently, assessing people's views of robots heavily relies on questionnaires and interviews, which come with inherent

[3] All statistical analyses were run using SPSS 28. The models use a linear model with a normal distribution.

limitations [27], [28]. Firstly, these tools only reflect an individual's viewpoint at a particular instance, making it challenging to link shifts in perception to specific interaction moments. Secondly, for an accurate longitudinal assessment, multiple evaluations are necessary. Yet, repeatedly completing questionnaires disrupts the natural interaction with the robot, potentially reducing engagement and task performance. Lastly, relying on self-reported measures introduces biases; individuals might recall previous responses, leading to response fatigue or inadvertently revealing experimental objectives. In future work we will address these questions by analysing the videos of the interaction to try to discover if there are behavioral cues from the users (e.g. smiling, verbal and non-verbal feedback) which are correlated with their reported perceptions of the interactions.

The study could also benefit from qualitative data to complement quantitative ratings, providing deeper insights into users' subjective experiences. Therefore, to effectively analyze how perceptions of robots change over time and connect these changes to specific robot actions, there's a need for more subtle and continuous assessment methods.

Developers can leverage the insights gained from this analysis to prioritize elements such as gaze that enhance anthropomorphism and animacy in similar systems. In conclusion, the analysis provides a nuanced understanding of how different experimental conditions impact user perceptions across multiple attributes, offering valuable insights for both researchers and practitioners in the field.

ACKNOWLEDGMENT

The research reported in this paper was supported by grant 2014-39 from the Swedish Research Council, which funds the Centre for Linguistic Theory and Studies in Probability (CLASP) in the Department of Philosophy, Linguistics, and Theory of Science at the University of Gothenburg, Sweden, and is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 101077927)

REFERENCES

[1] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell, "A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception," in *Computer graphics forum*, vol. 34, no. 6. Wiley Online Library, 2015, pp. 299–326.

[2] 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, 2004.

[3] E. B. Onyeulo and V. Gandhi, "What makes a social robot good at interacting with humans?" *Information*, vol. 11, no. 1, p. 43, 2020.

[4] P. Aliasghari, A. Taheri, A. Meghdari, and E. Maghsoodi, "Implementing a gaze control system on a social robot in multi-person interactions," *SN Applied Sciences*, vol. 2, pp. 1–13, 2020.

[5] C. Liu, C. T. Ishi, H. Ishiguro, and N. Hagita, "Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 285–292.

[6] V. Somashekarappa, A. Sayeed, and C. Howes, "Neural network implementation of gaze-target prediction for human-robot interaction," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2023, pp. 2238–2244.

[7] G. Hoffman and C. Breazeal, "Collaboration in human-robot teams," in *AIAA 1st intelligent systems technical conference*, 2004, p. 6434.

[8] Z. Prasov, *Eye gaze for reference resolution in multimodal conversational interfaces*. Michigan State University. Computer Science, 2011.

[9] Z. Degutyte and A. Astell, "The role of eye gaze in regulating turn taking in conversations: a systematized review of methods and findings," *Frontiers in Psychology*, vol. 12, p. 616471, 2021.

[10] B. Mutlu, N. Roy, and S. Šabanović, "Cognitive human–robot interaction," *Springer handbook of robotics*, pp. 1907–1934, 2016.

[11] C. Rich, B. Ponsler, A. Holroyd, and C. L. Sidner, "Recognizing engagement in human-robot interaction," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 375–382.

[12] T. Wang, P. Zheng, S. Li, and L. Wang, "Multimodal human–robot interaction for human-centric smart manufacturing: A survey," *Advanced Intelligent Systems*, p. 2300359, 2023.

[13] V. Somashekarappa, C. Howes, and A. Sayeed, "A deep gaze into social and referential interaction," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 43, no. 43, 2021.

[14] R. S. Hessels, "How does gaze to faces support face-to-face interaction? a review and perspective," *Psychonomic bulletin & review*, vol. 27, no. 5, pp. 856–881, 2020.

[15] J. Cassell *et al.*, "Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents," *Embodied conversational agents*, vol. 1, 2000.

[16] M. Gullberg and K. Holmqvist, "Keeping an eye on gestures: Visual perception of gestures in face-to-face communication," *Pragmatics & Cognition*, vol. 7, no. 1, pp. 35–63, 1999.

[17] E. Heyselaar, D. Peeters, and P. Hagoort, "Do we predict upcoming speech content in naturalistic environments?" *Language, Cognition and Neuroscience*, vol. 36, no. 4, pp. 440–461, 2021.

[18] H. Mitterer and E. Reinisch, "Visual speech influences speech perception immediately but not automatically," *Attention, Perception, & Psychophysics*, vol. 79, pp. 660–678, 2017.

[19] P. M. Brown, "Mechanisms for listener-adaptation in language production: Limiting the role of the model of the listener," *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman*, vol. 105, pp. 117–142, 1991.

[20] I. Mason, "Gaze, positioning and identity in interpreter-mediated dialogues," *Coordinating participation in dialogue interpreting*, vol. 102, p. 177, 2012.

[21] R. Ishii, K. Otsuka, S. Kumano, and J. Yamato, "Prediction of who will be the next speaker and when using gaze behavior in multiparty meetings," *ACM Transactions on Interactive Intelligent Systems (TIIS)*, vol. 6, no. 1, pp. 1–31, 2016.

[22] D. Das, M. G. Rashed, Y. Kobayashi, and Y. Kuno, "Supporting human–robot interaction based on the level of visual focus of attention," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 6, pp. 664–675, 2015.

[23] J. Holler and K. H. Kendrick, "Unaddressed participants' gaze in multi-person interaction: optimizing recipiency," *Frontiers in psychology*, vol. 6, no. 98, pp. 1–14, 2015.

[24] H. L. O'Brien, P. Cairns, and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form," *International Journal of Human-Computer Studies*, vol. 112, pp. 28–39, 2018.

[25] G. Skantze, "Turn-taking in conversational systems and human-robot interaction: a review," *Computer Speech & Language*, vol. 67, p. 101178, 2021.

[26] R. Jaber, "Towards designing better speech agent interaction: Using eye gaze for interaction," Ph.D. dissertation, Department of Computer and Systems Sciences, Stockholm University, 2023.

[27] J. Fink, V. Bauwens, O. Mubin, F. Kaplan, and P. Dillenbourg, "People's perception of domestic service robots: same household, same opinion?" in *Social Robotics: Third International Conference, ICSR 2011, Amsterdam, The Netherlands, November 24-25, 2011. Proceedings 3*. Springer, 2011, pp. 204–213.

[28] M. Flensborg Damholdt, C. Vestergaard, M. Nørskov, R. Hakli, S. Larsen, and J. Seibt, "Towards a new scale for assessing attitudes towards social robots: The attitudes towards social robots scale (asor)," *Interaction Studies*, vol. 21, no. 1, pp. 24–56, 2020.