

Design Implications for Effective Robot Gaze Behaviors in Multiparty Interactions

Maria Teresa Parreira, Sarah Gillet
KTH Royal Institute of Technology
parreira@kth.se, sgillet@kth.se

Marynel Vázquez
Yale University
marynel.vazquez@yale.edu

Iolanda Leite
KTH Royal Institute of Technology
iolanda@kth.se

Abstract—Human-robot non-verbal communication has been a growing focus of research, as we realize its importance to achieve interaction goals (e.g. modulating turn-taking) and manage human perception of the interaction. Consequently, the development of models for robot non-verbal behavior, such as gaze, should be informed by studies of human reaction and perception to that behavior. Here, we look at data from two studies where two humans interact describing words to a robot. The robot tries to balance participation of the two players through a combination of gaze aversion, looking at the listener and looking at the speaker. We analyze how momentary gaze patterns reflect in the participant’s turn length and perception of the robot, as well as in the participation imbalance. Our findings may be used as recommendations towards crafting robot gaze behaviors in multiparty interactions.

Index Terms—multiparty interaction, gaze, non-verbal behavior, social robotics

I. INTRODUCTION

Non-verbal communication is an important component of social interactions [1]. In particular, eye gaze has been extensively studied in human-human communication and is an emerging topic in Human-Robot Interaction (HRI) [2].

In human interactions, gaze direction and timing can skew the perception of interactants [3]. Gaze can additionally regulate conversational dynamics. For instance, gaze aversion – the act of temporarily looking away from an interactant’s face – plays a role in modulating intimacy [4], signaling of cognitive effort [3] and managing turn-taking [5].

Previous literature has provided evidence that robot and human gaze behaviors elicit different responses from people [6]. Further, Yu et al. [7] reported that humans spent on average less time looking at the face of the other interactant if they are human than if they are an artificial agent. These differences motivate studies that provide an adequate characterization of human response to robot’s gaze.

In HRI, recent work has explored how gaze can help balance participation in multiparty interactions [8, 9, 10]. In [9, 10], two human participants interacted with a Furhat robot in a word guessing game. Figure 1 illustrates the collaborative setting, where the two human participants take turns. At a given time, one human is the speaker, the other is an



Fig. 1: Overview of the scenario. The robot Furhat is looking at the participant on the right.

active listener, and the robot is the addressee of both. The robot tries to balance participation of the two players through different gaze policies combining the behaviors of *looking at the speaker*, *looking at the listener* and *averting gaze*. The authors found that participation unevenness [11], number of turns taken and amount of active participation vary as a function of the gaze conditions tested for the robot.

While the results from [9, 10] are encouraging, they provide a macro-level analysis of the effect of gaze patterns on the quality of the interaction. In contrast, the present work investigates the effect of different gaze behaviors on the human participants on a more fine-grained level.

This paper aims to shed light on which momentary robot behaviors influence human conversational patterns and perception of a robot. Using the data from [9, 10], we find evidence that specific gaze behaviors, but also the effect of combining different behaviors, play a role in shaping the length of interventions, participation imbalance and human impression of the robot. From this analysis, we formulate recommendations for crafting robot gaze behaviors in multiparty interactions and different interaction goals, such as modulating turn length and human perception of the interaction.

II. RELATED WORK

Substantial amount of research has studied human-robot non-verbal communication, in particular gaze behavior [2]. For example, prior work studied the gaze patterns exhibited by the speaker (human or robot) and how these effect conversational dynamics and roles [3, 12]. For multiparty interactions, Wang et al. [13] evaluated human perception of the actions taken by a virtual agent acting as a listener. Oertel et al. [14] characterized and built a model of attentive listener gaze behavior in a robot. Three-party human interaction gazes are analyzed in [15] and

This work was partially funded by the Swedish Research Council (no. 2017-05189), the Swedish Foundation for Strategic Research (FFL18-0199), the Jacobs Foundation (no. 2017 1261 06) and the Vinnova Competence Center for Trustworthy Edge Computing Systems and Applications at KTH.

used to develop gaze models in a humanoid robot. Whereas these studies characterize human response to gaze to inform the modeling of robot gaze behavior, we evaluate micro-level dynamics that emerge from the use of such models.

Previous work has investigated human impression of interactants according to their gaze behaviors. In HRI, gaze patterns can modulate human impression of the interaction [16]. Virtual agents and robots exhibiting higher amounts of mutual gaze have been found to be more effective at building rapport with humans [17, 18]. However, the timing of gaze behavior is important, as longer durations of mutual gaze can have a negative social effect [18]. Andrist and colleagues [19] showed that robot gaze aversions can be used to manage the conversational floor and improve human perception of robots. Closer to our work, a pilot study compared the gaze aversion ratio (GAR) – ratio of time gazing away from the interaction partner to the gaze cycle time – with a user’s interaction experience [20]. Results indicated that a small GAR has a negative effect on the perception of the interaction.

In order to assess the implications of robot gaze patterns in conversational dynamics and human perception of the robot, we conduct our analysis with data from two studies [9, 10] that use different combinations of gaze behaviors from a robot to attempt to balance participation of two human subjects. The authors find evidence that, through the robot’s gaze behavior, it is possible to regulate participation unevenness, number of turns taken and active participation.

III. METHODOLOGY

This work aims to examine human response to gaze patterns from a robot, and how these gaze patterns influence human impression of the robot. To this end, we analyze the data from two in-between subject studies [9, 10].

A. Dataset

In the studies from [9, 10], two human participants with different proficiency levels of Swedish (one native, one learner) interact with a Furhat robot while playing a variant of the game *With Other Words* [9] in Swedish. In the game, the human players describe words which the robot tries to guess. This experimental setting encourages participants to collaborate, establishing a dynamic of turn-taking. These studies provided evidence that robot gaze can help in balancing human participation in the game.

The studies include four different conditions which vary the gaze behavior of the robot. In the *Heuristic Condition 1* (HC1, originally control in [9]), the robot only looks at the active speaker and performs gaze aversion. *Heuristic Condition 2* (HC2, originally experimental condition in [9]) tested the effect of a hand-crafted heuristic behavior, where the robot looks at the speaker or the listener, but never performs gaze aversion. In the *Learning Condition 1* (LC1) and *Learning Condition 2* (LC2), interaction data from the previous study [9] was used to train a gaze policy for Furhat via imitation learning (LC1, originally IL in [10]) and via batch reinforcement learning (LC2, originally RL in [10]). The learned gaze policies include

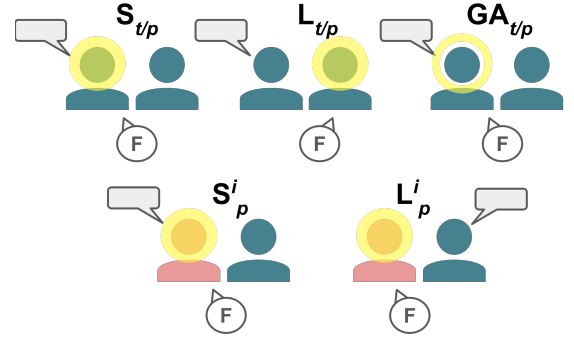


Fig. 2: Schematic view of gaze behaviors and measures considered. Furhat represented with F .

three types of behaviors: *look at speaker*, *look at listener*, and *perform gaze aversion*.

The dataset used for this study comprises 51 interaction episodes collected in the four different conditions (15 HC1, 12 HC2, 12 LC1, 12 LC2) with 15-20 min of length, in a total of 6014 conversational turns. Available data from the studies includes: video footage, action commands sent to the robot, participants’ voice activation as well as game information. The dataset further provides participants’ perception of the interaction from a post-interaction questionnaire.

A total of 102 participants are included across [9, 10], with ages between 67 and 18 ($M = 30$, $SD = 11$). In terms of gender, there were 50 female, 50 male and 2 people who rather not say. Also, 58 participants reported not having interacted with a robot before while 44 did, and a total of 22 participants reported having contact with robots at work.

B. Research Questions

We set to answer the following research questions:

RQ1 *How is the turn length, i.e. the time one participant is holding the floor, affected by the robots gaze behaviors (look at speaker/listener, perform gaze aversion)?*

Because [9] found that gaze behaviors could balance participation, we also wanted to investigate:

RQ2 *To what extent is participation unevenness influenced by the precise robot gaze behaviors?*

Lastly, because different gaze behaviors have different signaling functions[3], we explored:

RQ3 *Is human impression of the robot (discomfort, warmth) modulated by the robot’s gaze behaviors?*

C. Data Preparation

We analyze participants’ interventions during the interaction with data collected from the Voice Activation Detection (VAD) method used in [9, 10]. We make the following assumptions: any robot action command that led to action durations smaller than 0.5 seconds was considered as not executed by the Furhat robot and is therefore not used for the analysis; likewise, any participant intervention shorter than 0.5 seconds is considered backchannelling and not a conversational turn.

D. Measures

To analyze the effects of the robot's gaze behaviors at a fine granular level, we study how momentary gaze behavior patterns demonstrated by Furhat reflect on human behavior (i.e. holding the floor and participation behavior) and perception of the robot (in terms of warmth and discomfort). We characterize the distribution of robot behaviors taking into account the **relative time performing gaze aversion (GA)**, **relative time looking at the speaker (S)**, and **relative time looking at the listener (L)**.

We compute the distributions of the robot's gaze behaviors within different horizons of the interactions. We consider the distribution of behaviors in each turn t , but we also evaluate the collective effect of robot gaze patterns across several turns. For this, we divide each interaction ($M = 957s$, $SD = 30s$) into three **interaction phases** p , each comprising of 300s (the last phase varies in length). We look at $\mathbf{GA}_p$, $\mathbf{S}_p$ and $\mathbf{L}_p$, where the cumulative time performing each behavior per phase is normalized to the length of the phase. Note that the speaker changes during the phases. These measures therefore describe the general robot behavior towards the interacting dyad. Finally, we compute $\mathbf{S}_p^i$ and $\mathbf{L}_p^i$, which can be defined, respectively, as the cumulative time participant i is looked at by the robot while speaking/listening, normalized to the length of the phase. As opposed to $\mathbf{GA}_p$, $\mathbf{S}_p$ and $\mathbf{L}_p$, these two participant-centric measures allow us to extract the robot's behaviors as directed to the individual participant during one phase (Fig. 2). We evaluate how these distributions affect:

1) **Turn length (s)**: for each turn t , we consider $\mathbf{GA}_t$, $\mathbf{S}_t$ and $\mathbf{L}_t$, where the cumulative time performing each behavior is normalized to the duration of that turn.

2) **Participation unevenness [11]**: for each phase p , is defined by:

$$\text{uneven}_p = \sum_{i \in [1,2]} |\text{sp}^i - \bar{\text{sp}}| \quad (1)$$

where sp^i is the amount of time that participant i has spoken over the total amount of speech of the two human players. $\bar{\text{sp}}$ represents to the mean of the relative speech time of the players. In this case, $\bar{\text{sp}} = \frac{1}{2} \sum_{i \in [1,2]} \text{sp}^i$.

3) **Perception of the robot**: we measure impressions of the robot by analyzing the **Discomfort** and **Warmth** scale from the Robotic Social Attributes Scale (RoSAS) [21]. Note that the dataset only provides perceptions of the robot for a subset of the data ($LC1$ and $LC2$). We compare these impressions with $\mathbf{GA}_p$, $\mathbf{S}_p$ and $\mathbf{L}_p$, $\mathbf{S}_p^i$ and $\mathbf{L}_p^i$.

Finally, we consider self-reported **extroversion** and **proficiency of the language learner**, as these are expected to influence the behavior of the participants in the task of playing the language game [9, 10].

IV. RESULTS

A. Length of turns

We analyzed the effect of $\mathbf{GA}_t$, $\mathbf{L}_t$, $\mathbf{S}_t$ and the *Condition* on the *Length of the Turn* through a four-way ANCOVA while controlling for the *Proficiency* of the language learner and

TABLE I: Analysis of deviance table for turn length, a double dot indicates the interaction between the factors. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

	Chisq	Df	Pr(>Chisq)	
extroversion	4.03	1	0.044	*
proficiency	5.33	4	0.255	
condition	13.11	3	0.004	**
$\mathbf{S}_t$	67.87	1	<2.2e-16	***
$\mathbf{GA}_t$	35.80	1	2.182e-09	***
$\mathbf{L}_t$	0.09	1	0.763	
$\mathbf{S}_t:\mathbf{GA}_t$	374.58	1	<2.2e-16	***
$\mathbf{S}_t:\mathbf{L}_t$	254.04	1	<2.2e-16	***
$\mathbf{GA}_t:\mathbf{L}_t$	9.33	1	0.002	**
$\mathbf{S}_t:\mathbf{GA}_t:\mathbf{L}_t$	24.53	1	7.320e-07	***

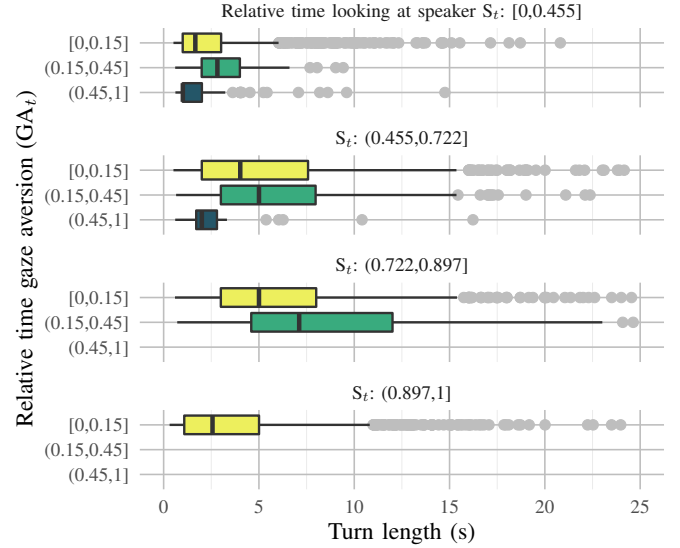


Fig. 3: The effect of increasing $\mathbf{S}_t$ (top to bottom, large grids) and $\mathbf{GA}_t$ (top to bottom, small grids) on turn length.

Extroversion. The analysis showed a main effect of *Condition*, $\mathbf{GA}_t$ and $\mathbf{S}_t$ on the turn length. The covariate *Extroversion* was significantly related to the turn length. The summary of the results can be found in Table I. Fig. 3 illustrates how turn length varies according to $\mathbf{GA}_t$ and $\mathbf{S}_t$. Increasing the time spent looking at the speaker ($\mathbf{S}_t$) to a high to moderate amount gives rise to longer turns, but very high amounts of gaze aversion ($\mathbf{GA}_t$) (> 0.45) will shorten turns.

B. Participation unevenness

We analyzed the effect of $\mathbf{GA}_p$, $\mathbf{L}_p$, $\mathbf{S}_p$, the *Condition* and the *Interaction Phase* on the *Unevenness* in participation through a five-way ANCOVA while controlling for the *Proficiency* of the language learner. The analysis showed a main effect of $\mathbf{S}_p$ on participation *Unevenness*, $F(1, 137) = 4.433$, $p = 0.037$, indicating that the more the robot focused on gazing to the speaker, the higher the imbalance. Further, we found a trend to significance for $\mathbf{L}_p$ ($F(1, 137) = 3.051$, $p = 0.082$) and *Condition* ($F(3, 137) = 2.155$, $p = 0.096$). The covariate *Proficiency* was significantly related to the participation unevenness.

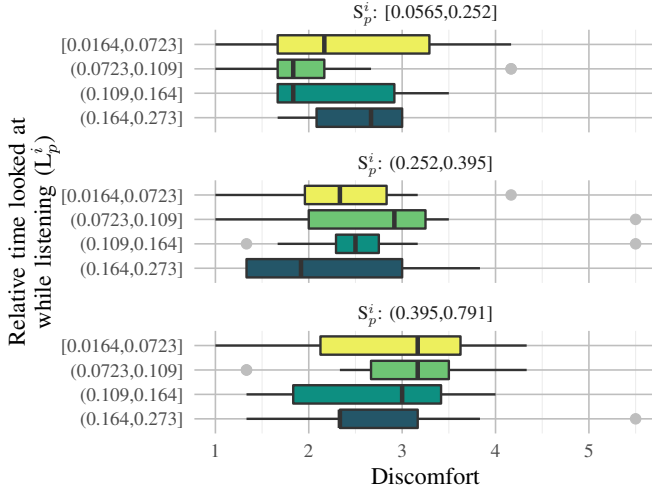


Fig. 4: The effect of increasing S_p^i (top to bottom, large grids) and increasing L_p^i (top to bottom, small grids) on reported discomfort.

C. Robot perception

1) *Discomfort*: We analyzed the effect of GA_p , L_p , S_p , S_p^i , L_p^i , the *Condition* and the *Interaction Phase* on the perceived *Discomfort* through a five-way ANCOVA while controlling for the *Proficiency* of the language learner and *Extroversion*. The analysis showed a main effect of S_p^i , $F(1, 104) = 4.020$, $p = 0.048$, and L_p^i , $F(1, 104) = 4.442$, $p = 0.037$, on perceived discomfort. It indicates that discomfort increases the more time the robot spends gazing at the person while they are speaking and decreases the more the robot looks at the person while they are listening. Further, we found significant interactions between S_p^i and L_p^i ($F(1, 104) = 4.238$, $p = 0.042$), S_p^i and S_p ($F(1, 104) = 4.068$, $p = 0.046$), L_p^i and S_p ($F(1, 104) = 4.174$, $p = 0.044$), as well as S_p^i , L_p^i and S_p ($F(1, 104) = 4.278$, $p = 0.041$). Figure 4 shows how L_p^i and S_p^i interact according to reported discomfort. For lower levels of S_p^i (top large grid), reported discomfort is low, but it increases if the robot is looking more at the listener while the participant is talking (higher L_p^i , dark blue). A reverse effect can be observed for higher levels of S_p^i (bottom large grid), where higher L_p^i actually lowers reported discomfort.

2) *Warmth*: We analyzed the effect of GA_p , L_p , S_p , S_p^i , L_p^i , *Condition* and *Interaction Phase* on perceived *Warmth* through a five-way ANCOVA while controlling for *Proficiency* of the language learner and *Extroversion*. No significant main effects were found on *Warmth*. *Proficiency* was significantly related to *Warmth*, $F(4, 104) = 12.353$, $p < 0.001$.

V. DISCUSSION

Our results suggest that important aspects of conversational dynamics can be predicted by a combination of the robot's gaze behaviors. Longer turns by human participants are associated with more time being looked at by the robot while speaking. This may be explained by perceived attentiveness of the robot [3]. When analyzing the interaction of S_t and GA_t ,

longer turns seem to emerge when a high proportion of the time is spent with the robot looking at the speaker combined with moderate gaze aversion (< 0.45).

Interestingly, interactions where the robot looks more at the speaker appear to be more imbalanced in terms of human participation. These results are consistent in that longer turns may provide fewer opportunities to balance participation. The aforementioned effects of the distribution of gaze behaviors in conversational dynamics are an interesting finding which may be used to inform the development of robot non-verbal behavior models.

We also find that human impression of the robot appears to be related to the distribution of the robot's gaze behavior. Reported discomfort increases if the participant is looked at more while speaking. This is in line with prior work that found that maximizing mutual attentiveness can have negative social effects [17]. Figure 4 considers the interaction between time being looked at while speaking (S_p^i) with the time being looked at while listening. We note that higher discomfort is associated with higher S_p^i . The lowest discomfort was obtained when both S_p^i and L_p^i are moderate to low, which may indicate that Furhat's gaze on the participant may lead to a higher feeling of uneasiness. In sum, there is a contrasting trend: more time attending to the speaker leads to *longer turns*, but *more imbalanced participation* and *discomfort*. This trend should be taken into account when implementing robot gaze behavior.

A. Limitations

In this work, we analyzed human response to robot gaze patterns. However, non-verbal communication is a complex problem, even in the controlled, experimental setting used in [9, 10]. We made the assumption the participants were always looking at the robot. In reality, participants interact with the robot and among themselves. First insights into participant's reaction in terms of their non-verbal behavior in [9] were discussed by Weldon et al. [22]. Additionally, we do not analyze head position and movements, which have been shown to be important in turn management [23, 24].

VI. CONCLUSION

This work evaluates how momentary robot gaze patterns in a multiparty interaction reflect in the participant's turn length and perception of the robot, as well as in the interaction imbalance. When crafting gaze behaviors, these findings may be taken into account according to the goal of the interaction. If the goal is to elicit longer interventions from the human interactants, more time looking at the speaker may be a contribution factor. In contrast, this guideline may be counterproductive if the goal is to create interactions which are ranked positively for human impression, since a longer time being looked at while speaking leads to more reports of discomfort. Likewise, if the aim is to balance participation, gazing a lot at the speaker is unadvised. More studies of this kind in different interaction settings will be important to provide generalizable insights into human reaction to robot gaze.

REFERENCES

- [1] J. K. Burgoon and A. E. Bacue, "Nonverbal communication skills." 2003.
- [2] H. Admoni and B. Scassellati, "Social Eye Gaze in Human-Robot Interaction: A Review," *Journal of Human-Robot Interaction*, vol. 6, no. 1, p. 25, 3 2017. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3109975>
- [3] M. Argyle and C. Mark, *Gaze and mutual gaze*. Cambridge University Press, 1976.
- [4] A. Abele, "Functions of gaze in social interaction: Communication and monitoring," *Journal of Nonverbal Behavior*, vol. 10, pp. 83–101, 06 1986.
- [5] A. Kendon, "Some functions of gaze-direction in social interaction," *Acta Psychologica*, vol. 26, pp. 22–63, 1967. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0001691867900054>
- [6] H. Admoni and B. Scassellati, "Robot gaze is different from human gaze: Evidence that robot gaze does not cue reflexive attention," in *Proceedings of HRI '12 Gaze in Human-Robot Interaction Workshop*, March 2012.
- [7] C. Yu, P. Schermerhorn, and M. Scheutz, "Adaptive eye gaze patterns in interactions with human and artificial agents," *ACM Trans. Interact. Intell. Syst.*, vol. 1, no. 2, jan 2012. [Online]. Available: <https://doi.org/10.1145/2070719.2070726>
- [8] G. Skantze, "Predicting and Regulating Participation Equality in Human-robot Conversations: Effects of Age and Gender," *ACM/IEEE International Conference on Human-Robot Interaction*, vol. Part F1271, pp. 196–204, 2017.
- [9] S. Gillet, R. Cumbal, A. Pereira, J. Lopes, O. Engwall, and I. Leite, "Robot gaze can mediate participation imbalance in groups with different skill levels," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 303–311. [Online]. Available: <https://doi.org/10.1145/3434073.3444670>
- [10] S. Gillet, M. T. Parreira, M. Vázquez, and Leite, "Learning gaze behaviors for balancing participation in group human-robot interactions," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022.
- [11] H. Tennent, S. Shen, and M. Jung, "Micbot: A Peripheral Robotic Object to Shape Conversational Dynamics and Team Performance," *ACM/IEEE International Conference on Human-Robot Interaction*, vol. 2019-March, pp. 133–142, 2019.
- [12] B. Mutlu, J. Forlizzi, and J. Hodgins, "A storytelling robot: Modeling and evaluation of human-like gaze behavior," *Proceedings of the 2006 6th IEEE-RAS International Conference on Humanoid Robots, HUMANOIDS*, pp. 518–523, 2006.
- [13] Z. Wang, J. Lee, and S. Marsella, "Multi-party, multi-role comprehensive listening behavior," *Autonomous Agents and Multi-Agent Systems*, vol. 27, p. 218–234, 09 2013.
- [14] C. Oertel, P. Jonell, D. Kontogiorgos, K. F. Mora, J.-M. Odobez, and J. Gustafson, "Towards an engagement-aware attentive artificial listener for multi-party interactions," *Frontiers in Robotics and AI*, vol. 8, p. 189, 2021. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frobt.2021.555913>
- [15] T. Shintani, C. T. Ishi, and H. Ishiguro, "Analysis of role-based gaze behaviors and gaze aversions, and implementation of robot's gaze control for multi-party dialogue," in *Proceedings of the 9th International Conference on Human-Agent Interaction*, ser. HAI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 332–336. [Online]. Available: <https://doi.org/10.1145/3472307.3484653>
- [16] A. Fukayama, T. Ohno, N. Mukawa, M. Sawaki, and N. Hagita, "Messages embedded in gaze of interface agents — impression management with agent's gaze," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '02. New York, NY, USA: Association for Computing Machinery, 2002, p. 41–48. [Online]. Available: <https://doi.org/10.1145/503376.503385>
- [17] N. Wang and J. Gratch, "Don't just stare at me!" in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 1241–1250. [Online]. Available: <https://doi.org/10.1145/1753326.1753513>
- [18] Y. Zhang, J. Beskow, and H. Kjellström, "Look but don't stare: Mutual gaze interaction in social robots," in *Social Robotics*, A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eysse, and H. He, Eds. Cham: Springer International Publishing, 2017, pp. 556–566.
- [19] S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu, "Conversational gaze aversion for humanlike robots," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 25–32. [Online]. Available: <https://doi.org/10.1145/2559636.2559666>
- [20] M. Koller, D. Bauer, J. de Pagter, G. Papagni, and M. Vincze, "A pilot study on determining the relation between gaze aversion and interaction experience," in *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '19. IEEE Press, 2019, p. 644–645.
- [21] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas): Development and validation," in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017, pp. 254–262.
- [22] C. F. Weldon, S. Gillet, R. Cumbal, and I. Leite, "Exploring non-verbal gaze behavior in groups mediated by an adaptive robot," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 357–361.
- [23] K. Jokinen, M. Nishida, and S. Yamamoto, "Eye-gaze experiments for conversation monitoring," 01 2009, pp. 303–308.
- [24] K. Jokinen, K. Harada, M. Nishida, and S. Yamamoto, "Turn-alignment using eye-gaze and speech in conversational interaction," 01 2010, pp. 2018–2021.