# Look but Don't Stare: Mutual Gaze Interaction in Social Robots

Yanxia Zhang[1], Jonas Beskow[2], and Hedvig Kjellström[2]

[1] Delft University of Technology, The Netherlands
[2] KTH Royal Institute of Technology, Sweden
Yanxia.Zhang@tudelft.nl
{beskow, hedvig}@kth.se

**Abstract.** Mutual gaze is a powerful cue for communicating social attention and intention. A plethora of studies have demonstrated the fundamental roles of mutual gaze in establishing communicative links between humans, and enabling non-verbal communication of social attention and intention. The amount of mutual gaze between two partners regulates human-human interaction and is a sign of social engagement. This paper investigates whether implementing mutual gaze in robotic systems can achieve social effects, thus to improve human robot interaction. Based on insights from existing human face-to-face interaction studies, we implemented an interactive mutual gaze model in an embodied agent, the social robot head Furhat. We evaluated the mutual gaze prototype with 24 participants in three applications. Our results show that our mutual gaze model improves social connectedness between robots and users.

## 1 Introduction

Many robotic applications require users to interact and collaborate with robots, for example, tutoring systems, rehabilitation, and information guidance. With the advancement of sensing technologies, there is an increasing interest in research to integrate natural user interfaces with intelligent systems. Previous work has demonstrated the effectiveness of using natural human actions such as hand gestures and gaze as inputs for interactive systems [19]. In addition to integrating human actions, robot interaction can be made more appealing and engaging by enabling artificial agents to interpret humans' communication signals and to generate appropriate responses [18].

Recently, there has been significant progress in enabling natural human-robot communication via different interfaces, such as speech, touch, and natural human gestures. Adapting a robot's gaze behaviour to real-time humans' gaze in social interactions has also gained much interests in interactive robotic systems. Human-like eye movements in a robot provide high communicative value, and thus can be a powerful resource for facilitating natural HRI [4]. Humans are highly sensitive to their social partners' gaze signals and adapt their behaviours accordingly [5,18]. One vital gaze behaviour is *mutual gaze*; a natural behaviour when two people maintain eye contact between each other (for communication, signalling, and social coordination). Without eye-contact, people do not perceive that they are fully immersed in a conversation [2]. Mutual gaze is also a nonverbal cue that regulates intimacy and a sign of affiliation level [3]. Although gaze models have previously been implemented in robotic systems to convey non-verbal behaviours of robots [12], their systems are typically not reactive in real-time. Furthermore, the majority of prior research focused on designing gaze behaviours that make virtual agents and robots look more natural and realistic [7]. Whether gaze behaviour in robots can achieve social effects and how this influences interaction is less known.

This paper contributes a novel interactive gaze model that enables a social robot to establish mutual gaze with their users. We evaluated this model with 24 users in three applications to understand how adding gaze into the interaction can affect people's perception of the robot and how mutual gaze regulates social attention. Our results showed that users indeed perceived the gaze attention behaviour that we implemented in the embodied agent. Integrating mutual gaze resulted in positive social effect, improved fluency in interactive applications, and drew more user attention.

## 2 Related Work

Inherently, eye gaze is a salient stimulus that conveys social cues and has multiple functions in human-human communication [2,3]. People actively use their gaze to regulate social attention and to influence the receiver. For instance, humans use eye contact to seek for confirmation of their progress in collaborative tasks. Gazing away from the interaction partner indicates distractions or lost of interests. Besides gaze direction, the duration that we look at each other in the eyes (mutual gaze) is another crucial signal. The longer that two people stare at each other can increase their arousal state, e.g. the gaze duration can be interpreted as aggressiveness. Therefore, people would look at others' facial areas briefly and then return to the eye region. Furthermore, mutual gaze regulates human perception of their social interaction, such as intimacy, regulator of conversations, and trust.

Although gaze models have been extensively studied in virtual agents [12], only a small number of research works investigated the use of gaze input in HRI. Gaze and joint attention for grounding is studied in a shared tabletop collaborative scenario [8]. In their work, a NAO robot actively moved its head to achieve joint visual attention and to signal turn-taking to facilitate the grounding process. Yoshikawa *et al.* proposed two gaze models that enable robots to follow user's gaze continuously and to avert the gaze when users look at the robot's face [17]. They found that robots with responsive gaze control give users a stronger feeling of being looked at, and it can increase the users' perception of the agent's social presence. Andrist *et al.* employed gaze aversion in conversational robots, and their results showed that robots can use gaze aversions to appear more intentional and thoughtful [1]. However, their work did not use humans' gaze as an input, but their gaze aversions were responsive to human speech.

In sum, previous research investigated gaze models for robots without taking into account of the human partner's eye movements. The majority of existing robots' gaze behaviour respond to users' speech [1,14], or the gaze behaviour is achieved via head movements [10,16,8,14]. Nonetheless, head movements can only approximate coarse gaze direction and cannot convey subtle eye movements such as saccades.

## 3 Mutual Gaze Model Implementation

This section describes the implementation of our system that enables real-time mutual gaze interaction with a social robot head. We use a Tobii EyeX Controller to capture gaze information and a social robot head as an output for user interaction. The EyeX controller tracks a user's eye ball positions and gaze at an optimal distance of 45-80 cm (see Figure 1). We set the eye tracker to operate at a frame rate above 30Hz on average. The social robot used in this work is designed by the Furhat robotics [9]. The robot head (referred as *Furhat* from here onward) projects a 3D-animation of a face model on a translucent mask. Furhat conveys a different physical affordance compared to the other social robots in the market, e.g., the Nao robot. It exhibits human-like gaze behaviour and facial expression via a back projection. The facial behaviour is controlled by a 3D face model which is similar to that of virtual agents (see Figure 1).

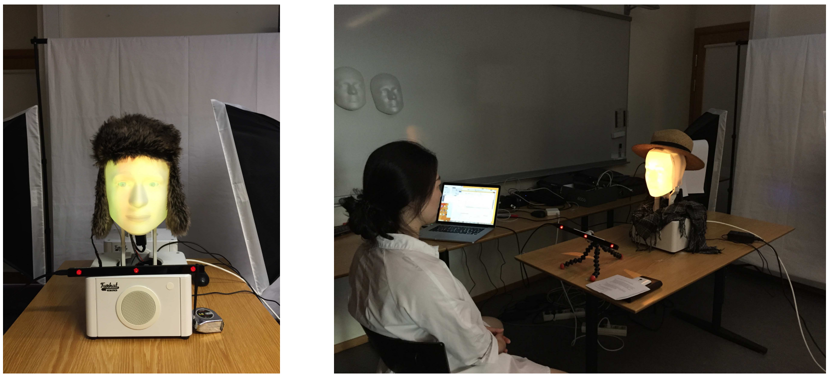

**Fig. 1.** The interactive gaze model is implemented on a talking social robot head Furhat. An eye tracker is placed in front to detect users' gaze.

To detect users' gaze towards the robot, we define it as when the user is fixated within the facial region of the robot's eye area. Due to the physiological properties of human eye movements, our eyes cannot stare at a single point for a prolonged duration. Therefore, we need to filter out fast saccadic eye movements. Our system tracks users' gaze continuously and takes the filtered gaze coordinates $(x,y)$ from the Tobii SDK. We denote the gaze region within the robot's eye and face area as $R$. We maintain a window frame $W = w_i$ where $i \in (1,2...,6)$, which keeps a record of whether the gaze falls within the predefined region. For each time stamp, $w_i$ is defined as 1 if $(x,y) \in R$, otherwise 0. The mean of $W$ represents users' gaze state. In our implementation, when the mean is greater than 0.66, user's state is classified as gaze *on* the robot, and when the mean is less than 0.66, user's state is classified as looking *away*.

We denote the robot and human's gaze status at any timestamp $t$ as two states: $A$ represents looking away and $F$ represents looking towards the other party. In the face-to-face situated interaction, the system can have four possible states (the first symbol denotes the robot status and the latter denotes the user's status): $\{AF, AA, FA, FF\}$; for example, AF denotes the state when the robot is looking away from the user and the user is looking at the robots. When the system initialises, it is in a default START state (see Figure 2). The transition from time $t_{i-1}$ to $t_i$ invokes when the current user's gaze status changes or the duration permitted in the current state has expired. Users' gaze status is determined from the mean of values in window frame $W$. Based on existing analysis from human interaction data [5], there is no direct mutual inversion state transition from from AF to FA and from AA to FF. For example, the transition has to go through from AA to FA (the robot looks at the user first) or from AA to AF (the user initiates eye contact with the robot).

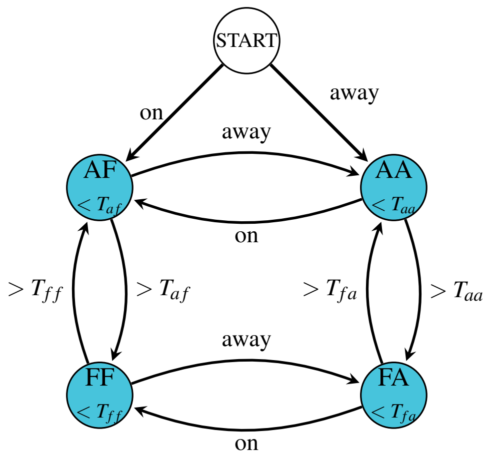


**Fig. 2.** State transition diagram for our interactive mutual gaze model. *On* denotes when user is detected as looking towards the robot. *Away* denotes when user is averting their gaze from Furhat.

The state transition is a deterministic process as shown in Figure 2. We model the duration of mutual gaze and gaze away with a Gaussian distribution based on findings in the previous work [7]. The transition model consists of the following variables:

- $T_{ff}$ represents the time before breaking mutual gaze. It informs the robot when to look away and how long the mutual gaze holds.
- $T_{af}$ is the variable that characterises how fast the robot reacts to human's eye contact. It determines how long the gaze aversion should be.
- We set $T_{aa}$ and $T_{fa}$ as fixed term in the current model. It represents how often the robot attempts to initiate eye contact with the user.

The system implements the state transition diagram in Figure 2. The eye movements of the robot are inferred from the system's current state and observations of user's gaze status. The Furhat system controls the robot eye movement via a built-in 3D gaze model, which triggers a gaze shift to a location in 3D space at a certain speed and direction. To generate naturalistic eye movements, the gaze shift amplitudes are set as less than

30 degrees [7]. Our system also injects small jittery eye movements as human eyes never stay still completely. Even when robots gaze at a fix location, the eyes still make tiny movement around the fixation point. Although existing literature suggested that users may avert their gaze away in different directions (e.g. up, down, side) based on conversation contexts and internal cognitive states [1]. Our system does not control the gaze behaviour based on semantic information in the dialogue but assign different aversion directions randomly.

## 4 Evaluation

### 4.1 Setup and Procedure

We conducted a user study to find out people's subjective experience and evaluate the effectiveness of our model in a face-to-face human-robot interaction setup. We recruited 24 participants (14 males and 10 females, with a mean age of 27 years SD 3.5 years) to take part in the study. An eye tracker was placed at a distance of 60 cm in front of a Furhat robot (as shown in Figure 1). The participants were seated approximately 130 cm away from the robot. The eye tracker, Furhat, and the user were centrally aligned.

Before each session started, the participant's head orientation was calibrated to achieve direct eye contact from the robot. We adjusted the participants' sitting height to guarantee that their eyes are on the same level as the robot's eyes. We performed user calibrations on a computer screen and tested the accuracy of our eye contact detection system. During the actual study, we removed the screen and put Furhat at the same distance as the screen plane.

We adopted a between-subject design with two different conditions varying between participants (see Table 1). Each 12 subjects experienced only one of the two conditions. Our baseline condition is using continuous eye contact as similar to gaze behaviour always used in existing robotic systems: the robot attends to users' face when they are facing towards the robot. We omitted random gaze as a condition in our study, because previous work showed that random gaze in humanoids can make users perceive them as unnatural and negatively impact user performance [13]. The gaze shift is generated via eye movement only. The robot's neck remains still in the experiment as our pilot study showed users considered fast movements with robot head overwhelming.

**Table 1.** Design parameters of gaze model in the default and mutual gaze condition

| Condition | Description |
|---|---|
| *Default* | attend to face continuously $T_{af} = 0$ |
| *Mutual gaze* | $T_{ff} = 1500\ ms$ ; $T_{af} = 2500\ ms$ |

Prior to the study, the experimenter explained the purpose and procedure of the study to the participants. The entire experiment lasted for 30 mins on average. During each session, the participants performed three tasks (a verbal task, a story telling task and a quiz game) and filled in a questionnaire after each task. After all tasks, the subject filled in a questionnaire about their overall experience and feedback. The task order was counterbalanced. The verbal task is similar to [11] by asking participants to talk about emotional and personal topics. The participants were asked to answer two questions during this task. Furhat first greeted the participants and asked them to describe the route that they took to the university that day and their first memories of the city. In the storytelling task, the participants were asked to listen to a story about the history of Stockholm. This task tested whether the person was looking and engaged in mutual gaze with Furhat while the robot was speaking. The quiz game aimed to test whether users involve in mutual gaze in a cognitive-demanding task and how fluent when users interact with Furhat in a continuous interaction. In this task, Furhat asked questions from different categories including physics, history, famous person, and culture. The task ended when the participants answered five questions correctly.

### 4.2 Measures

To understand the subjective feedback on how users perceive the interaction, we used a five-item social scale to measure users' impression as in [6]. It contains items about the perception of the robots in terms of friendliness, cooperativeness, sociability and warmth, and another item to measure how intelligent the participants perceive the robot. To measure trust, we used two items asking how trustworthy and honest the robot was perceived. Another factor we considered is the feeling of rapport towards the robots, which indicates how much users emotionally engaged in the interaction. We used the rapport scale adopted in [11,15]. We also asked about interaction experience with two items on how the participants perceive their overall experience as "enjoyable" and "boring". After each task, we asked about whether users felt Furhat recognised their attention and how they felt about the amount of attention given by Furhat. Additionally, in the verbal task, we asked whether users were nervous while speaking to Furhat. In the storytelling task, we asked how concentrate users were while listening. In the quiz task, we asked whether it was fun to play with Furhat and how fluent the interaction was. Each question was rated on a 7-point scale from 1 (strongly disagree) to 7 (strongly agree). Besides questionnaires, we collected eye movements for gaze analysis.

## 5 Results

### 5.1 Gaze Analysis

In the default condition, we only recorded the coordinates of the user's gaze, and the system does not react to the user's gaze. Instead, Furhat kept its gaze focused on the user's face position, which was captured during calibration. To get a better understanding of eye movement patterns in the mutual gaze condition, we plotted a recorded segment of the generated Furhat gaze behaviour states in Figure 3. The interaction events denoted four state transitions in the green line from the collected data. As described in previous sections, the status of user gaze focusing on or away from the Furhat is plotted as red segment in the top row in Figure 3. This example showed that our system enabled mutual gaze interactions between Furhat and users. We observed

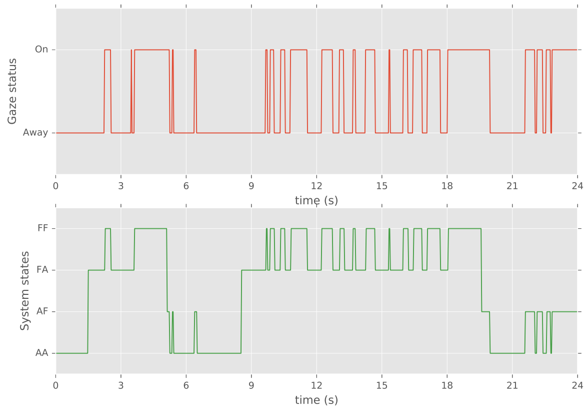


**Fig. 3.** Recorded interactions showing user gaze status on Furhat and gaze model state transitions. The four states are: *AA*: robot look away and user look away; *AF*: robot look away and user look at Furhat; *FA*: robot look at user and user look away; *FF*: mutual eye-to-eye contact.

frequent states switching between *FF* and *FA*. During these transitions, Furhat provided continuous attention towards the subject while he/she glanced away frequently. Later on, the subject continuously gazed at Furhat. This lead to a state transition from *FF* to *AF* when Furhat actively broke the eye contact with the subject.

Mutual attention is one of the major aspects to characterise rapport. We summarise how the participants allocate their visual attention from the collected data to verify our hypothesis, and also compare the results to our collected subjective measures. We define the visual attention as the amount of time that a user gaze away and on the facial areas of the Furhat. Table 2 illustrates the average time that users' visual attention were on and away from the Furhat in the three tasks. We performed an independent-samples t-test to compare the percentage of visual attention allocated on the default condition and the mutual gaze condition. We found a significant difference in the amount of visual attention in the quiz task between the two conditions. 60% of users focused more on the mutual gaze condition ($t(22) = -2.512, p = 0.020$), while only 39% of attention were on Furhat in the default condition.

**Table 2.** Average percentage of visual attention on the Furhat at each task in two conditions

| *Task* | *Gaze status* | *Default* $(m \pm std)$ | *Mutual Gaze* $(m \pm std)$ | $p-value$ |
|---|---|---|---|---|
| Verbal | Away | $65\% \pm 16\%$ | $54\% \pm 18\%$ | 0.117 |
| | On | $35\% \pm 16\%$ | $46\% \pm 18\%$ | |
| Story | Away | $35\% \pm 23\%$ | $33\% \pm 22\%$ | 0.825 |
| | On | $65\% \pm 23\%$ | $67\% \pm 22\%$ | |
| Quiz | Away | $61\% \pm 24\%$ | $40\% \pm 17\%$ | $0.020^{*}$ |
| | On | $39\% \pm 24\%$ | $60\% \pm 17\%$ | |

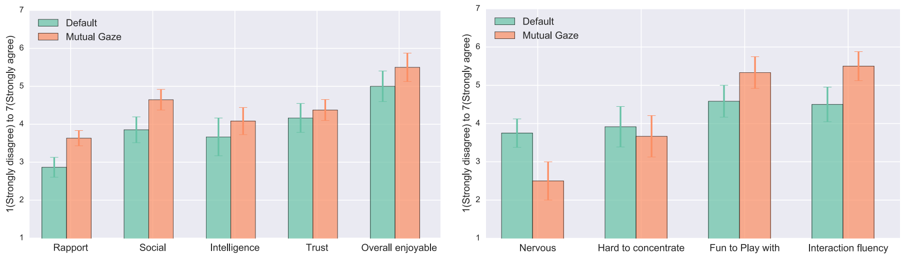


**Fig. 4.** Left: overall subjective feedback after the entire study. Right: user feedback on interaction experience. The error bars in all the following figures represent the standard error of the mean.

### 5.2 Questionnaire Analysis

Participants' feedback on the overall interaction experience and the impression of the robot is presented in Figure 4. The overall rating in five categories (*Rapport*, *Social*, *Intelligence*, *Trust*, *Overall enjoyable*) is the sum of all items in each category.

We performed a Mann-Whitney U test analysis on all items from the questionnaire rating feedback. Our results showed that users perceived that they established higher rapport in the mutual gaze condition compared to the default condition ($U$ = 35.5, $p$ = 0.017). There are no significant differences in the overall rating of the other four categories. There are significant differences in two items in the rapport rating. Users in the mutual gaze condition ($M$ = 4.0) found that they were able to engage Furhat better than the default ($M$ = 3.5) group ($U$ = 42.0, $p$ = 0.033). In addition, they perceived that Furhat and them could understand each other better ($U$ = 27.5, $p$ = 0.004) in the mutual gaze condition ($M$ = 4.0), but not in the default condition ($M$ = 2.0). Among the four items in the social category, the participants perceived Furhat to be significantly more social ($U$ = 40.0, $p$ = 0.029) in the mutual gaze condition ($M$ = 4.5) than the default condition ($M$ = 3.5).

Results of subjective feedback on the three tasks are shown in the right bar plot in Figure 4. In the verbal task, users felt significantly more nervous ($U$ = 43.0, $p$ = 0.043) while talking to Furhat in the default condition ($M$ = 4) but less nervous in the mutual gaze condition ($M$ = 2). In the storytelling task, we did not find a significant difference between the two conditions about whether users found it hard to listen to the story. For the interactive quiz application, users perceived the interaction to be significantly more fluent ($U$ = 42.0, $p$ = 0.035) in the mutual gaze ($M$ = 6) than the default condition ($M$ = 5). This result is consistent with the average interaction time in the quiz game task in the mutual gaze condition. The interaction time is significantly shorter ($t(22) = 5.045$, $p <$ 0.0005) for successfully solving the quiz task in the mutual gaze condition ($M$ = 120.5$s$, $Std$ = 43.0$s$) compared to the default condition ($M$ = 202.7$s$, $Std$ = 36.6$s$).

## 6 Discussion

Previous works mainly used head orientations as proxies of gaze attention in robotic systems [10,16,8,14]. Our results demonstrated that subtle mutual gaze interaction can provide implicit and positive social effects. Coarse head directions might not be efficient in expressing implicit non-verbal messages embedded in subtle eye movements.

Humans are highly sensitive to other's gaze. Communicating with eye gaze is subconscious behaviour that does not require learning. One major reason of the effectiveness of our gaze model is that we implemented it on a back-projected interface, which allowed the animation to simulate fine-grained eye movements. Eye movements are subtle and usually move at high frequency; to achieve natural mutual gaze interaction, it is challenging to proximate gaze attention with mechanical head movements. The average mutual gaze duration lasted approximately 1 second in this experiment. Our initial testing in the pilot study showed that head movement at high speed (e.g., robot hold two second mutual gaze with user and then look away) lead to negative user experience, in particular with mechanical neck used in the Furhat robot. In contrast, using visually animated gaze in our system successfully carried the affordances of rapid eye movement.

Our results showed that users paid significantly more attention on Furhat and solved the quiz questions much faster in the mutual gaze condition. The subjects further agreed that interactions were more fluent. One reason for the positive outcome could be that mutual gaze makes interaction smoother throughout the multiple question-answer rounds. Also, in the default condition, prolonged gaze on users could increase their arousal state. This might be another factor that influenced their perception of being pressured to provide answers and could lead to higher cognitive load. However, further studies are needed to verify this, i.e., using biosignals to measure cognitive load.

We learned that the subjects generally found the interaction with Furhat to be more fun in the mutual gaze condition. The participants felt that they were more connected with Furhat and perceived Furhat to be more social. These effects could be due to similar findings in human-human interaction that mutual eye contact is the key to developing social interaction [2,3]. Maintaining eye contact is a natural mimic behaviour to form a sense of trust. In contrast, the default condition where Furhat continuously looked at the user made the subjects more nervous during conversation. These results suggest that designers of human-like robot should consider generating appropriate gaze behaviours by taking into account of social gaze interaction. Failing to respect the process of gaze-based social interaction could potentially impede natural human-robot interaction. In addition, gaze is highly related to speech content. Prior to learning the effect of gaze and speech, this work is a first step to understand how gaze alone affects HRI. An extension for this work is to incorporate other modalities. In the future, we will also investigate data driven approaches to learn the gaze model from human interaction data.

## 7 Conclusion

This paper proposed a novel interactive gaze model that enables a robot to establish mutual gaze with users. We implemented a proof-of-concept of the gaze model, and we conducted a user study to evaluate the effectiveness of mutual gaze interaction with a social robot head. While the conventional method of conveying social attention is through robotic head movements, our work revealed that subtle gaze feedback can impact a user's perception of social connectedness with a robot. Our results further suggested that enabling mutual gaze can improve rapport between the robot and its user. Expressing full gaze attention can negatively influence users' perception of social interaction with a robot. This calls for further social robotic research to design appropriate gaze attention models for users.

## References


1. Andrist, S., Tan, X.Z., Gleicher, M., Mutlu, B.: Conversational gaze aversion for humanlike robots. In: Proc. HRI 2014. pp. 25–32. ACM (2014)
2. Argyle, M., Dean, J.: Eye-contact, distance and affiliation. Sociometry pp. 289–304 (1965)
3. Bailenson, J.N., Blascovich, J., Beall, A.C., Loomis, J.M.: Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. Presence 10(6), 583–598 (2001)
4. Breazeal, C., Fitzpatrick, P.: That certain look: Social amplification of animate vision. In: Proc. AAAI 2000 (2000)
5. Cappella, J.N., Pelachaud, C.: Rules for responsive robots: Using human interactions to build virtual interactions. Stability and change in relationships p. 325 (2002)
6. Hoffman, G., Birnbaum, G.E., Vanunu, K., Sass, O., Reis, H.T.: Robot responsiveness to human disclosure affects social impression and appeal. In: Proc. HRI 2014. ACM (2014)
7. Lee, S.P., Badler, J.B., Badler, N.I.: Eyes alive. ACM Trans. Graph. 21(3), 637–644 (Jul 2002)
8. Mehlmann, G., Häring, M., Janowski, K., Baur, T., Gebhard, P., André, E.: Exploring a model of gaze for grounding in multimodal hri. In: Proc. ICMI 2014. pp. 247–254. ACM (2014)
9. Moubayed, S.A., Skantze, G., Beskow, J.: The furhat back-projected humanoid head–lip reading, gaze and multi-party interaction. International Journal of Humanoid Robotics 10(01), 1350005 (2013)
10. Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., Ishiguro, H.: Conversational gaze mechanisms for humanlike robots. ACM Trans. Interact. Intell. Syst. 1(2), 12:1–12:33 (Jan 2012)
11. Riek, L.D., Paul, P.C., Robinson, P.: When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. Journal on Multimodal User Interfaces 3(1-2), 99–108 (2010)
12. Ruhland, K., Peters, C., Andrist, S., Badler, J., Badler, N., Gleicher, M., Mutlu, B., McDonnell, R.: A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In: Computer Graphics Forum. vol. 34, pp. 299–326. Wiley Online Library (2015)
13. Skantze, G., Hjalmarsson, A., Oertel, C.: Exploring the effects of gaze and pauses in situated human-robot interaction. In: Proceedings of the 14th Annual Meeting of Special Interest Group on Doscourse and Dialogue-SIGDial. pp. 375–383 (2013)
14. Skantze, G., Johansson, M., Beskow, J.: Exploring turn-taking cues in multi-party human-robot discussions about objects. In: Proc. ICMI 2015. pp. 67–74. ACM (2015)
15. Wang, N., Gratch, J.: Don't just stare at me! In: Proc. CHI 2010. pp. 1241–1250. ACM (2010)
16. Yonezawa, T., Yamazoe, H., Utsumi, A., Abe, S.: Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. In: Proc. ICMI 2007. pp. 140–145. ACM (2007)
17. Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., Miyamoto, T.: Responsive robot gaze to interaction partner. In: Robotics: Science and systems (2006)
18. Yu, C., Schermerhorn, P., Scheutz, M.: Adaptive eye gaze patterns in interactions with human and artificial agents. ACM Trans. Interact. Intell. Syst. 1(2), 13:1–13:25 (Jan 2012)
19. Zhang, Y., Stellmach, S., Sellen, A., Blake, A.: The costs and benefits of combining gaze and hand gestures for remote interaction. In: Proc INTERACT 2015. vol. 9298, pp. 570–577 (2015)