Research Article

A Proactive Approach of Robotic Framework for Making Eye Contact with Humans

Mohammed Moshiul Hoque, Yoshinori Kobayashi, and Yoshinori Kuno

Graduate School of Science & Engineering, Saitama University, Saitama 338-8570, Japan

Correspondence should be addressed to Mohammed Moshiul Hoque; moshiulh@yahoo.com

Making eye contact is a most important prerequisite function of humans to initiate a conversation with others. However, it is not an easy task for a robot to make eye contact with a human if they are not facing each other initially or the human is intensely engaged his/her task. If the robot would like to start communication with a particular person, it should turn its gaze to that person and make eye contact with him/her. However, such a turning action alone is not enough to set up an eye contact phenomenon in all cases. Therefore, the robot should perform some stronger actions in some situations so that it can attract the target person before meeting his/her gaze. In this paper, we proposed a conceptual model of eye contact for social robots consisting of two phases: capturing attention and ensuring the attention capture. Evaluation experiments with human participants reveal the effectiveness of the proposed model in four viewing situations, namely, central field of view, near peripheral field of view, far peripheral field of view, and out of field of view.

1. Introduction

Human-robot interaction (HRI) is an interdisciplinary research field aimed at improving the interaction between human beings and robots and developing robots that are capable of functioning effectively in real-world domains, working and collaborating with humans in their daily activities. For robots to be accepted into the real world, they must be capable of behaving in such a way that humans do with other humans. Although a number of significant challenges remained unsolved related to the social capabilities of robots, the robot that can proactively make eye contact with human is also an important research issue in the realm of natural HRI.

Eye contact is a phenomenon that occurs when two people cross their gaze (i.e., looking at each other) which plays an important role in initiating an interaction and in regulating face-to-face communication [1, 2]. Eye contact behaviour is the basis of and developmental precursor to more complex gaze behaviours such as joint visual attention [3]. It is also a component of turn-taking that sets the stage for language learning [4, 5]. Eye contact also results in better information recall of the conversation [6]. For any social interaction to be initiated and maintained, parties need to establish eye contact [7]. However, it is very difficult to establish such gaze behaviours for one person while the target person is not facing him/her or while the target people are intensely attending to their task. A robot that naturally makes eye contact with human is one of its major capabilities to be implemented in social robots. Capturing attention and ensuring attention capture are the two important prerequisites for making an eye contact episode between the human and the robot.

When the robot and the intended target human are not facing each other, the robot should use proactive approach for making eye contact with him/her. This proactive nature is an important capability for robots that should be explored in the realm of HRI. This approach enables robots to initiate communication with a particular human urgently, such as in the case of reporting an emergency. The robot's capability to make eye contact proactively can also be used as an invitation service. Providing robots with natural social skills that foster the impression of a more intelligent and intuitive interaction

ensures a high level of satisfaction for interacting humans [8]. Moreover, to cope with the collaborative environment with humans, the robot not only respond against their needs but also convey its own intention to them. In summary, the major issues in our research are (i) how can a robot use subtle cues (i.e., actions) to attract a human's attention (i.e., attention capture) if he/she is not facing the robot, in other words, if the robot cannot capture his/her eyes or whole face due to the spatial arrangements of the person and the robot and (ii) how robot ensures that the human is responding against its action and how it can tell when it has captured attention? To answer these issues we proposed a framework and we designed a robotic framework based on this that was confirmed to be effective in making eye contact with humans in experimental evaluation.

2. Hypotheses in Making Eye Contact

Humans usually turn their head or gaze first toward the person with whom they would like to communicate with because turning the head or gaze is considered as the most fundamental cue to capture someone's attention [9]. If the target human does not respond, she/he tries again with the same action or with more strong signals (e.g., waving hand, shaking head, moving body, or voice, and so on). Robots should use the same convention as humans in a natural HRI scenario.

When the situation involves delegating an action to an agent, it is especially important that the speaker have evidence about whether the action was successful [10]. Attention attraction can produce observable behavioural responses such as eye, head movements, or body orientation [11, 12]. Where a speaker or listener is looking is potentially a powerful cue about attention and intention in face-to-face communication [13]. Therefore, if the target person felt attracted by the robot, she/he will turn toward it, which will make face-to-face orientation (i.e., crossing gaze of each other). Psychological studies show, however, that this gaze crossing action alone may not be enough to make a successful eye contact event [14]. Yoshikawa et al. [15] also mentioned that simply staring is not always sufficient for a robot to make someone feel that they are being looked at. Thus, after crossing the gaze with the intended recipient the robot should interpret the human looking response and display gaze-awareness which is an important behaviour for humans to feel that the robot understands his/her attentional response. To display awareness explicitly the robot should use some actions (verbal or nonverbal). In this paper, we use eye blinking actions for the robot as its gaze-awareness function to ensure attention capture.

Based on the above discussion, we can hypothesize that robots should perform two tasks consecutively: (i) attention capture and (ii) ensuring attention capture in order to set up an eye contact event. Figure 1 illustrates the conceptual process of making proactive eye contact in terms of these tasks. To perform a successful eye contact episode, both a robot (R) and a human (H) need to show some explicit behaviours and need to respond appropriately to them by communicative behaviours in each phase.
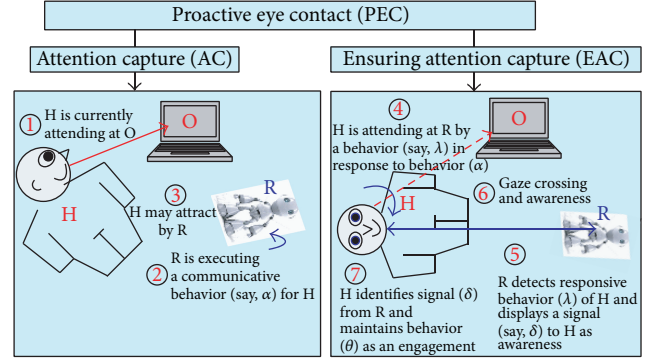

Figure 1: Prerequisites of proactive eye contact process.

In this work, we apply a set of behaviours of robot, such as α = {head_turning, head_shaking, reference_terms} in attention capture phase and δ = {face-detection, eye_blinks} in ensuring attention capture phase. We are also expecting human behaviours such as λ = {(head ∧ gaze ∧ body) turn_toward_robot} in attention attraction phase and θ = {keep_looking_toward_robot} in attention capture phase. After detecting current focus of attention of H, R sends actions to capture his/her attention toward it. If H is looking at R by turning his/her gaze, head, and/or body then face-to-face situation may occur which meets the first condition. The second phase is initiated by detecting H's response. After H's face detection, R displays its eye blink actions as an ensuring attention capture behaviour that satisfies the second condition.

2.1. DoA and Viewing Situations. Without information from the eyes, head direction and body orientation are also important indicators to determine someone's direction of attention (DoA). There are plenty of situations where determining the receivers' eyes direction is infeasible or almost impossible. In particular, when people are at greater distances or people are not facing each other, head orientation becomes a stronger cue than information from the eyes in determining direction of attention [14, 16]. That is why, if one wishes to know the attentional focus of another person, monitoring that person's head orientation is often a good substitute for monitoring his/her eye gaze. Because the head is a much larger visual stimulus than the eyes, the social signals associated with the facing direction of the head are still quite accessible in peripheral vision despite the lower visual acuity [17, 18]. Gaze is not the only cue that is used to determine the focus of another individual's direction of attention. The whole head, in particular the orientation in which it is directed, is a sufficient indicator of attention direction (and therefore interest). In some instances, the eyes are not visible and the only cue available for processing is the head direction. By the same reasoning, if the head is occluded or in shadow, the orientation of the body provides a sufficient cue for communication [19, 20].

Although, there may be various situations that we are facing in our daily lives, in this paper, we consider more general situations where a human and a robot are not facing

(a) Central field of view (CFOV)

(b) Near peripheral field of view (NPFOV)

(c) Far peripheral field of view (FPFOV)
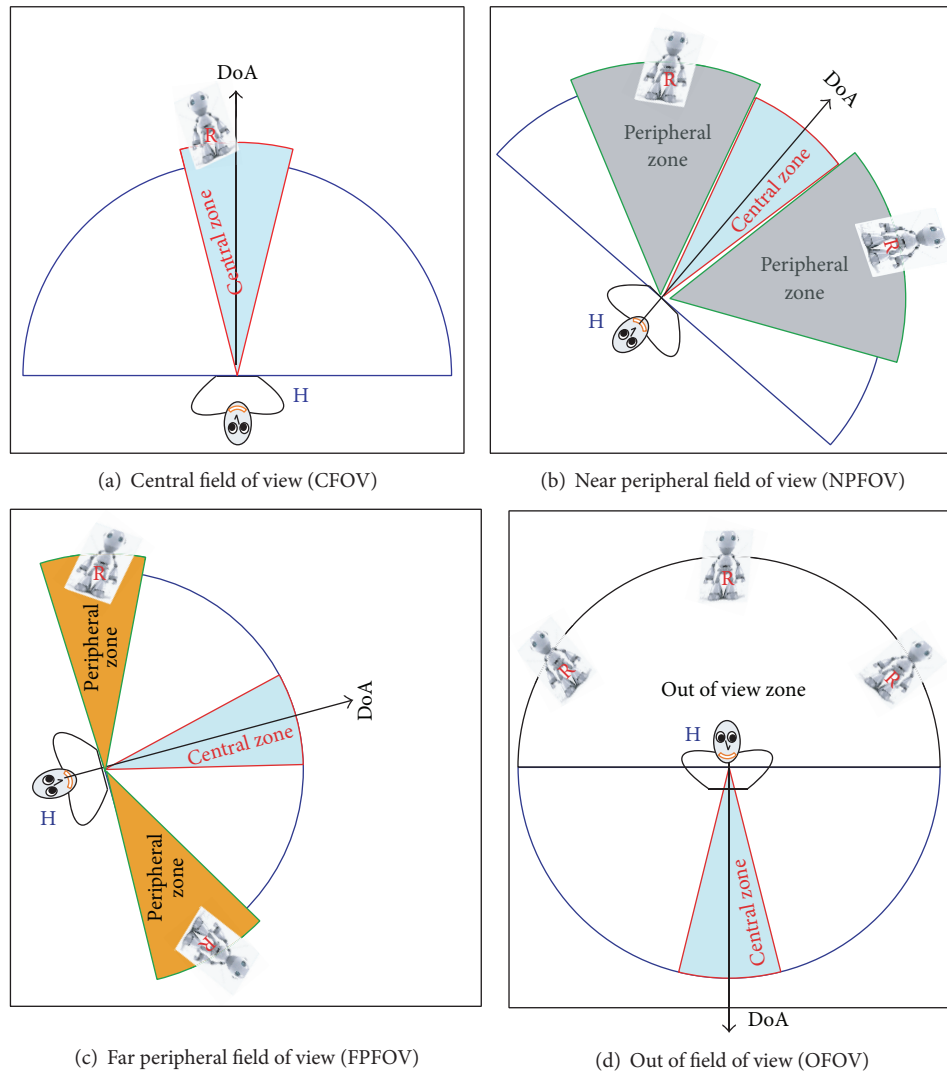

(d) Out of field of view (OFOV)

FIGURE 2: Four viewing situations of the human with respect to the robot's position.

each other initially and the human is engaged in his/her task (i.e., "watching paintings"). During observation of the paintings, the human perceive the robot in his/her different viewing situations due to head and eye movements. Human's field of view is wide and is divided into central vision and peripheral vision. We define the positional relation between the human and the robot by the robot's position in the human's field of view, that is, where the robot is seen in the human's field of view. We consider four positional relations: the robot is seen in the central field of view (CFOV), in the near peripheral field of view (NPFOV), in the far peripheral field of view (FPFOV), and in the out of field of view (OFOV). Figure 2 conceptually illustrates these situations. In any case, the robot and the target human are not facing each other initially except in CFOV.

2.2. Effective Vision Zone for Situations. Due to the placement of human eyes, the binocular field of view covers roughly about 180° horizontally [21]. In each situation, the looking direction (i.e., DoA) of the human will be different.

Based on our previous experiments [22] and the literature on human's field of view, we differentiate the following four types of viewing situations.

 (i) *Central field of view (CFOV):* exists at the center of the field of view of human. This zone is set to a 20° cone-shaped area (10° left and 10° right with respect to the line of regard).

 (ii) *Near peripheral field of view (NPFOV):* exists adjacent to the center of gaze of the human. This zone is thus set to a 120° fan-shaped area (60° to left hand side from CFOV zone and 60° to right side from CFOV zone) in front of a person.

(iii) *Far peripheral field of view (FPFOV):* exists at the edges of the field of view of human. This zone is set to 40° cone-shaped areas (20° to left hand side

from NPFOV zone and 20° to right side from NPFOV zone) in front of a person's head.

(iv) *Out of field of view (OFOV):* exists at the opposite side of the human's frontal direction. Thus, this zone is set to 180° semicircular shaped area.

3. Related Work

The capability of robots that can establish eye contact proactively with the target human attention is in a rudimentary stage still now. In proactive approach, capturing attention and ensuing attention capture are the two important prerequisites for robots. In human-human communication studies, there has not been significant work about how humans capture others' attention to initiate an interaction process beyond the primary facts that the human stop a certain distance [23], start the interaction with a greeting [7, 24], and arrange themselves in a spatial formation [25].

There have recently been a number of studies on people's responses to eye contact with robots in conversational tasks. These models used to produce the robot's gaze behaviour are typically either not based on human gaze behaviour or not reactive to the human partner's actions. In our paper, one basic motivation to focus on humanoid robots is to emulate natural human interaction. Having a naturalistic embodiment is often cited as necessary for meaningful social interaction [26–28] and the appearance of humanoid robot is similar to the human. A more extreme extension of this philosophy, presented by [29], claims that any truly social intelligence must have an embodiment that is structurally and functionally similar to the human. Although the argument for functional similarity between social robots and humans is well accepted, Kozima and others are suggesting that the physical instantiation of that functionality must also be as human-like as possible. Several other robots also have some social capability. These robots were designed as zoomorphic [30, 31] or functional embodiments [32, 33]. These types of robots were specially designed for children or disabled people to provide entertainment, creating companion as a pet, or provide a particular service. Due to the lacks of human-like embodiment, it may be difficult to evaluate the effectiveness of interaction in the human-robot collaborative situation. In the next section, we review the recent findings and developments on the capability of humanoid robots that can make eye contact with the human in terms of two approaches: passive and active.

3.1. Passive Approach. In the passive approach, the robot waits for a human to start an interaction. When people are at a public distance, it is too far for them to talk, but they can recognize each other's presence. Sisbot et al. developed a path-planning algorithm that considers people's positions and orientation to avoid disturbances [34, 35]. Pacchierotti et al. studied passing behavior and developed a robot that waits to make room for a passing person [36]. Bennewitz et al. [37] utilized a prediction of people's position, but it was only used for helping a robot avoid people not for allowing interaction with them. However, in these studies the interactions are

mainly achieved by changing body position and orientation and their studies were aimed at encouraging people's participation. Moreover, these works does not consider how robots should behave to attract human's attention for making eye contact.

Several previous HRI studies have addressed greeting behavior to initiate human-robot conversation and eye contact process at a social distance. These robots are designed to utter some greeting terms to initiate the interaction with the human [38–42]. A few studies have attempted to promote people's participation by such means as encouraging behavior using voice [43, 44] and detecting request behavior [45].

Some robots were equipped with the capability to encourage people to make eye contact by some nonverbal cues such as body orientation and gaze [46], approaching direction [47], standing position [48], and following behaviors [49]. These studies assumed that the target person faces the robot and intends to talk to it; however, in actual practice this assumption may not always hold. Robots may wait for a person to initiate an interaction process. These studies assumed that the target person faces the robot and intends to talk with it; however, in reality this assumption may not hold. Robots may wait for a person to initiate an interaction and using voice certainly attracts other people's attention including the target person. Although such a passive attitude can work in some situations, many situations require a robot to employ a more active approach [50–53].

3.2. Active Approach. It is certain that a robot that approaches people and initiates an interaction proactively by making eye contact should be perceived to act more natural than a passive machine. Some robots were equipped with the capability to initiate interaction proactively with humans. Satake et al. [54] present a method that enables a robot to approach humans proactively by predicting the trajectories of people in a public area. The robot then tries to catch the person's attention and initiates a conversation. A robotic system *Robovie-IV* used by Mitsunaga et al. [55] roams in an office environment actively searching for interaction partners. However, the engaging phase is passive, since the robot waits for a detected human to come close and respond to his interaction initiation.

The *IURO* [56, 57] robot is supposed to initiate interactions proactively by both initiating the conversation and approaching the person. This forces it to plan a path to a position in front of a human in a socially acceptable manner. However, they only focused on finding the suitable robot's approach speed and stopping distance. An Approach from a robot is not an easy problem since the robot's approach needs to be acknowledged nonverbally in advance; otherwise, the approached person might not recognize that the robot is approaching him/her or would be surprised by the robot's impolite interruption. Humans do this well with eye gaze [7, 24], but all the previous studies used the body orientation of the target and the robot for nonverbal interaction and their systems fail to recognize people's gaze and body direction, which are the most important parameters to measure whether the people have responded (been attracted) to the robot's intentional signal or not.

Several other robotic systems were developed to establish eye contact [58–62]. These robots are supposed to make eye contact with humans by turning their eyes (cameras) toward the human faces. Yonezawa et al. [63] use a stuffed-toy robot that can activate a favorable feeling by effective use of eye-contact reactions in combination with joint attention. An active gaze mechanism for conversations has been implemented in human-like robots to signal partners about their roles in conversation [64]. Results showed that participants conformed to these intended roles 97% of the time. Das et al. [65] focus on establishing the communication channel with the human through eye contact by considering the level of visual focus of attention in different tasks. All of these studies focused on the gaze crossing function alone for designing eye contact capability of robots and gaze-awareness function was absent in their system.

Several robotic systems incorporate gaze-awareness functions too. For example, Miyauchi et al. [66] design a system that can make eye contact between human and robot considering gaze crossing and gaze-awareness functions. After gaze crossing, it generates facial expression (i.e., smiling) as gaze-awareness function. This robot used a flat screen monitor as the robot's head and displayed 3D computer graphics (CG) images to produce smiling expression. A flat screen is unnatural as a face. A key aspect is the geometry of the face itself; the half-spherical overall shape lets the audience view the face within 180° wide area. Mimicking the geometry of the eyes, which remains the most important part in the face, helps interpret the robot's gaze, thus improving interaction [67]. Yoshikawa et al. [68] used a communication robot to produce the responsive gaze behaviors of the robot. This robot generates a following response and an averting response against the partner's gaze. They showed that the responsive robotic gazing system increases the feeling of people being looked at. However, it is unknown how the robot produces gaze-awareness behavior to the responded people. Moreover, the robotic heads that are used in previous studies were mechanically very complex and as such expensive to design, construct, and maintain. A recent work used a robot *Simon* to produce the awareness function [69]. The *Simon* blinks its ear when hearing an utterance. Although they considered the single person interaction scenario, they did not use ear blinks as a gaze-awareness purpose rather they used ear blinks to create interaction awareness.

4. Our Approach

The following subsections describe the behavioural protocol of the robot and design of the behavioural cues that are used in the experiments.

4.1. Behavioural Model of the Robot. The main objective of our work is to develop a robotic system that makes eye contact proactively with the human in terms of two consecutive phases: capturing human attention and ensuring attention capture. In our current work, we assume that the human and the robot are not facing each other and the human is attending his/her current task. To do so, the robot should
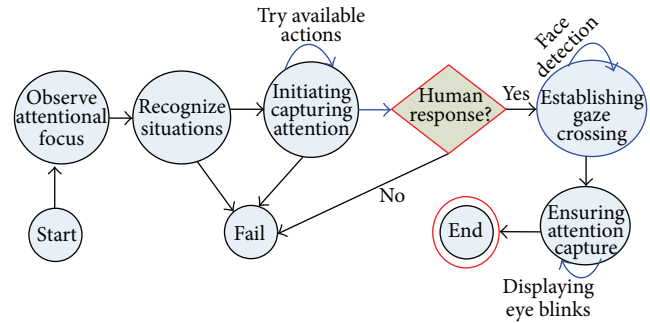

FIGURE 3: A proactive behavioural model of eye contact.

first gain the human's attention by any means. An eye contact event is executed by a finite-state-machine model as shown in Figure 3. In order to initiate the eye contact process, the robot begins to observe the current direction of the human's attention by tracking his/her head and body. After recognizing the viewing situation of the target human (TH), the robot usually turns its head first toward the TH and commences shaking its head and then utters reference terms (if necessary) to capture his/her attention. However, the robot waits about 4 seconds after each attempt for the TH to respond by looking in its direction (silences of more than 4 seconds become embarrassing because they imply a break in the thread of communication [70]).

If the robot is successful in attracting the TH's attention, the two agents will experience gaze crossing. Thus, the robot considers the TH to have responded to its actions if he/she looks at the robot within the expected time frame. Otherwise, it considers the case as a failure and initiates the interaction again. It is able to recognize whether this is so by detecting the front of his/her face in the camera image. After capturing the attention of TH, the robot performs a blinking action to display gaze-awareness as an ensuring attention capture behavior.

4.2. Design of the Behavioral Cues. In the experiments, four cues were used: two physical motion cues and one voice cue are designed to capture the target human's attentions, and an eye motion cue is designed as gaze-awareness to ensure the making of eye contact with the target (attracted) human.

4.2.1. Cues for Attracting Attention. It has been recognized that all cues are not equally effective in drawing people's attention [71]. The success of a particular action to attract human attention of a robot depends on several factors including the existing situation (i.e., direction of attention) as well as the nature of task that he/she is currently engaged in. Turning head toward the target person is the most fundamental action of the robot to whom it would like to communicate [9]. The target person should be aware of the fact that the robot would like to communicate with him/her because it turns its head. Our concerns are as follows: is it always possible to create such awareness of the target person by the head turning action alone, especially, when the robot and the target human are not facing each other or the target human is intensely engaged in

(a)   (b)
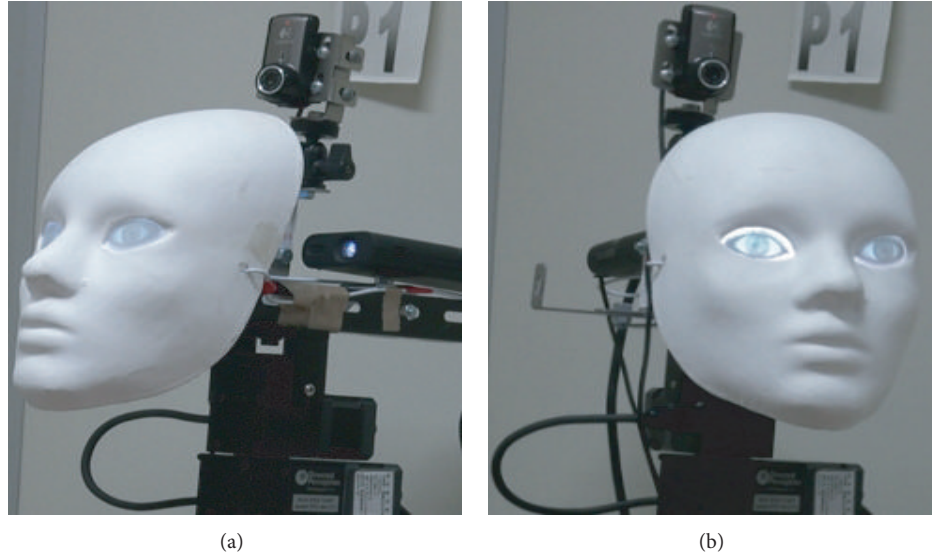
FIGURE 4: Attention attraction cues: (a) initial position and (b) after turning the head toward the human.

his/her current task? We hypothesize that the head turning action is effective only if the target human already looked at the robot or the robot exists in his/her central field of view.

Simple head turning or eye movements may be enough when the robot captures the target human in its central view but not effective in all cases [72]. The robot may need to use stronger action in some situations. For example, such a single turning action alone is not always enough when the robot is captured only in far peripheral field of view of the human. In that case, head shaking action may be an effective cue to attract human attention because object motion is especially likely to draw attention [73]. However, it may be apparent that visual stimuli offered by the robot's nonverbal behaviors cannot affect a person if he/she is in a position where he/she cannot see the robot. That means, when the robot detects the target human in its out of vision field or when robot is facing back of the human, it is difficult to attract his/her attention by any physical movement because that movement cannot be observed by the human's eyes due to the divergence of attentional focus. In this situation, the use of touch or voice should be considered as a last resort. The detailed description of the cues is given as follows.

  (i) *Head turning (HT):* the robot turns its head toward the target participant from its initial position by detecting his/her body position. In order to ensure the smooth movements, we adjusted the pan speed of the pan-tilt unit at 120°/second by performing several experimental trials. Figure 4(b) shows the HT cue after turning the robot's head toward the target participant from its initial position (Figure 4(a)).

 (ii) *Head shaking (HS):* the robot shakes its head back and forth ±30° from its current position. This means that the robot turns its head 30° left and 30° right. It tries up to 3 times in that fashion. The head shaking speed is adjusted at 240°/second.

(iii) *Uttering reference terms (RT):* the robot utters the recorded voice as the reference terms, such as "*excuse me,*" for attracting human attention.

4.2.2. Cues for Ensuring Attention Capture. If the target person is attracted by the robot behaviors, he/she will turn toward the robot that ensures face-to-face orientation and will detect his/her face. As previously discussed, gaze-awareness is also another important function for making successful eye contact.

Although there may be various ways to create such awareness function, we use eye blinking action for the robot because it is one of the most important cues for forming person's impressions [74]. The way one person looks at another and how often they blink seems to have a big impact on what kinds of impression they make on others. For example, it has been pointed out that people's impressions of others are affected by the duration they are looked at [75] and the rate of blinking [76]. These facts imply that it is important for a communication robot to use its eyes, which has encouraged many researchers to study how to use gaze direction and blinking in realistic, informative, and communicative ways [77, 78]. Although actual rates vary by individual, the average is around 10 blinks per minute in a laboratory setting [79]. Yoshikawa et al. [15] showed that the larger the number of blinking performed by the on-screen agent compared to the number of blinks performed by the subject, the more strongly the subject experiences the feeling of being looked at. Thus, we prepared the robot to perform blinking action at a higher rate than the human's average blinking rate. The behaviours of this cue are described in the following.

Eye Blinking. After detecting the face of the target participant, the robot starts blinking its eyes about 3 times. We set the robot to perform eye blinking after detecting the face of
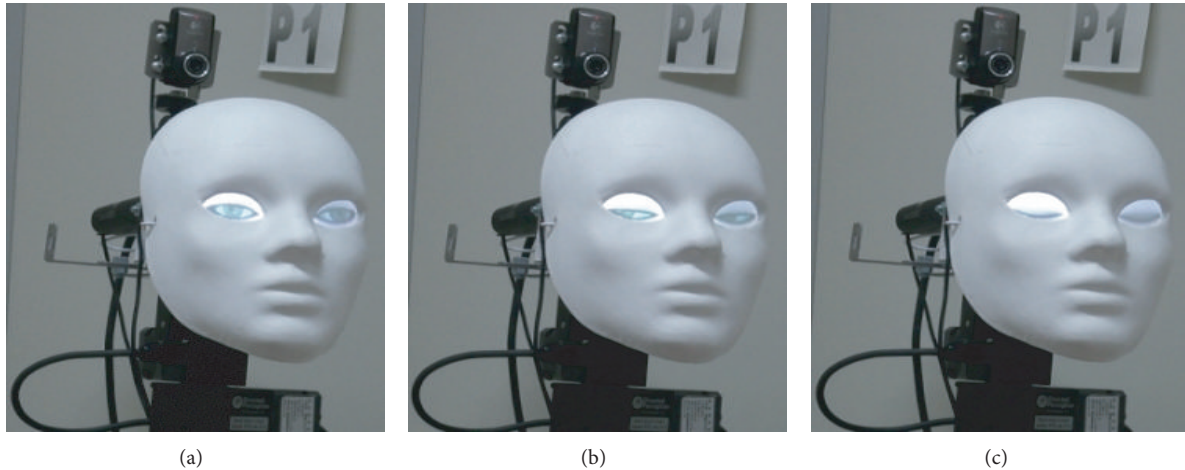
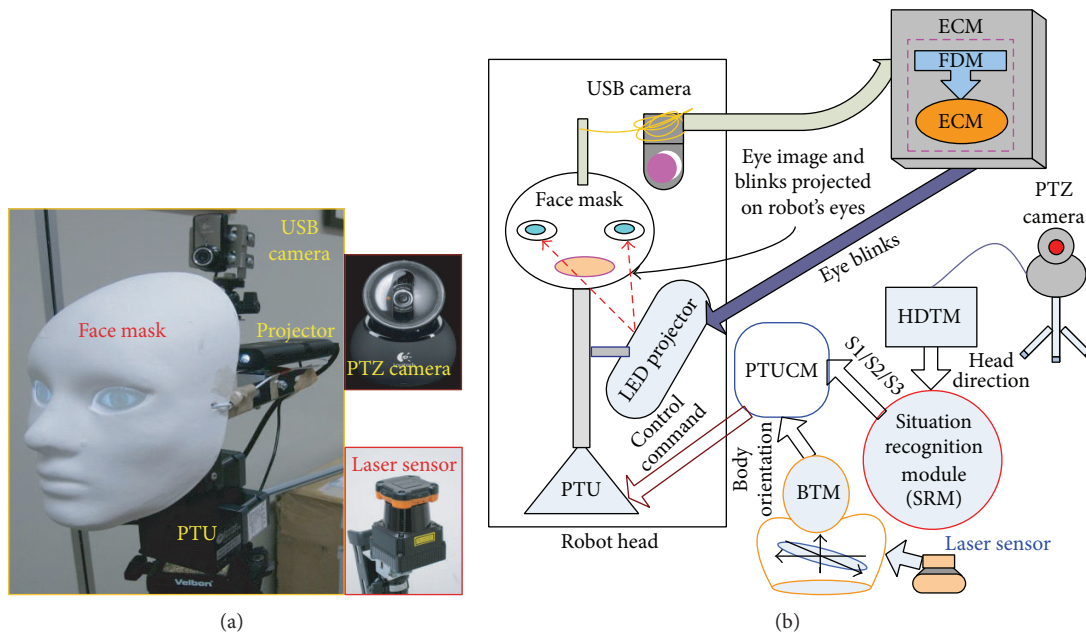FIGURE 5: Eye blinks actions: (a) partially closed eyes, (b) partially opened eyes, and (c) closed eyes.

FIGURE 6: System configuration: (a) prototype of the robotic head and (b) system consists of five modules: HDTM, SRM, BTM, ECM, and PTUCM.

the participant at a rate of 1 blink/second. Eye blinks are produced by the rapid closing and opening of the eyelids of the CG images and are displayed through the LED projector on the robot's eyes. Figures 5(a)–5(c) show some snapshots of a blinking action.

5. System Architecture

We have developed a robotic head for human-robot interaction experiments. In the following sections, we discuss the architecture of our robotic systems in terms of hardware and software configurations.

5.1. Hardware Configuration. Figure 6 shows an overview of our robotic head. The head consists of a spherical 3D mask, an LED projector (3M pocket projector, MPro150), laser range sensor (URG-04LX by Hokuyo Electric Machinery), a USB camera (Logicool Inc., Qcam), and a pan-tilt unit (Directed Perception Inc., PTU-D46). The 3D mask and projector are mounted on the pan-tilt unit. The USB camera is wired on the top of the mask to detect frontal face of human and the laser range sensor is placed on the participant's shoulder level so that the contour of participant's shoulder can be observed.

In order to provide a communication channel between the hardware components of the system, there is a standard RS-232 serial port connection between the general purpose
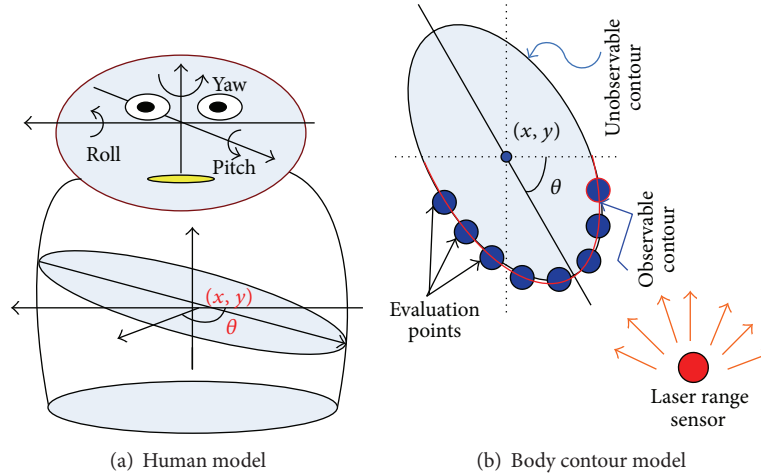
(a) Human model

(b) Body contour model

Figure 7: Human body model.

PC (Windows XP) and the pan-tilt unit. The LED projector projects CG generated eyes on the mask as in [80]. Thus, the head can show nonverbal behaviors by its head movements and eye movements including blinking. A PTZ camera (Logicool Inc., Qcam Orbit AF) is installed to track a human head and laser sensor track the human body. In the current implementation, the PTZ camera and laser sensor are put on a tripod placed at an appropriate position for observing the human body as well as head.

5.2. Software Configuration. The proposed system has five main software modules: the head detection and tracking module (HDTM), the body tracking module (BTM), the situation recognition module (SRM), the eye-contact module (ECM), and the pan-tilt unit control module (PTUCM). The last module controls the head movement and provides attention attraction signals based on the output of the second and third module, respectively.

5.2.1. Body Tracing Module (BTM). A human body can be modeled as an ellipse [81] (Figure 7(a)). We assume the coordinate system is represented with their X and Y axes aligned on the ground plane. Then, the human body model is consequently represented with center coordinates of ellipse $[x, y]$ and rotation of ellipse (θ). These parameters are estimated in each frame by the particle filter framework [82]. We assume that the laser range sensor is placed on the participant's shoulder level so that the contour of his/her shoulder can be observed. When the distance data which captured by the laser range sensor is mapped on the 2D image plane, the contour of participant's shoulder is partially observed, as shown in Figure 7(b).

The likelihood of each sample is evaluated as the maximum distance between evaluation points and the nearest distance data using

$$\pi = e^{-(d_{\max}^2/\sigma_d)}, \tag{1}$$

where π is the likelihood score based on the laser image and $d_{\max}$ is the maximum distance between evaluation points and the nearest distance data. At each time instance, once the distance image is generated from the laser image, each distance d_n is easily obtained. σ_d is the variance derived from d_n. Evaluation procedures are repeated for each sample. Conceptual images of evaluation process are shown in Figure 8(a). We employ several points on the observable contour as the evaluation points to evaluate hypotheses in the particle filter framework. These points are changes that depend on the relational position from the laser range sensor and the orientation of the model. Selection of evaluation points can be performed by calculating the inner product of normal vectors on the contour and its position vector from laser range sensor.

A typical example of the result of the BTM is shown in Figure 8(b). The BTM gives the body positions (x, y) of the human, distance between the human and laser sensor (D), and body orientation (θ). The results of the BTM (body orientation) are sent to the SRM to recognize OFOV situation and the robot adjusts its head orientation based on the position of the human.

5.2.2. Head Detection and Tracking Module (HDTM). To detect, track, and compute the direction of human head in real time (30 frame/sec), we use FaceAPI [83] by Seeing Machines, Inc. It can measure 3D head position (x, y, z) and direction (yaw (α), pitch (β), and roll (γ)) within 3° errors. One USB camera is placed in front of the human to track his/her face up to ±90°. A snapshot of HDTM results is shown in Figure 9(a). The head coordinate frame is a right-handed 3D reference frame and the origin is the midpoint of the line that joins the center of the eyes sclera spheres. When viewed from in front of the face, the x-axis points horizontally to the left toward the person's right eye, the y-axis points vertically upward, and the z-axis points away from the viewer, towards the back of the head. The results of the HDTM are sent to the SRM to classify the current attentional direction (NPFOV and FPFOV) of the target person.
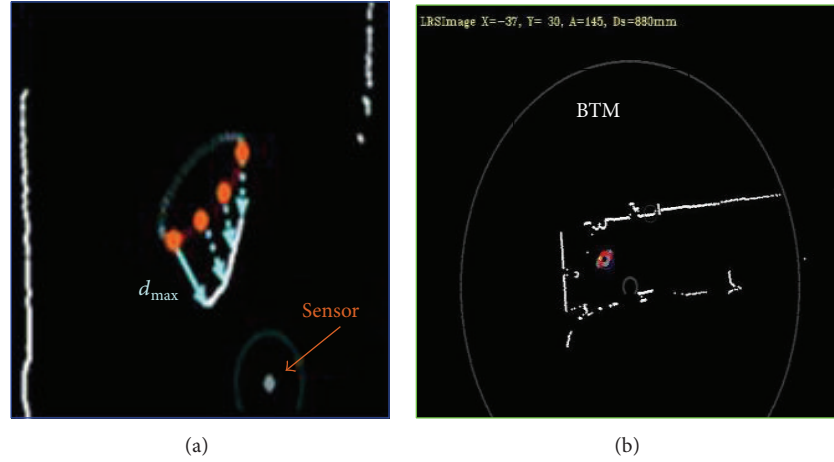
(a)

(b)

FIGURE 8: (a) Likelihood evaluation based on maximum distance. (b) Tracking results of BTM in terms of body position and orientation.
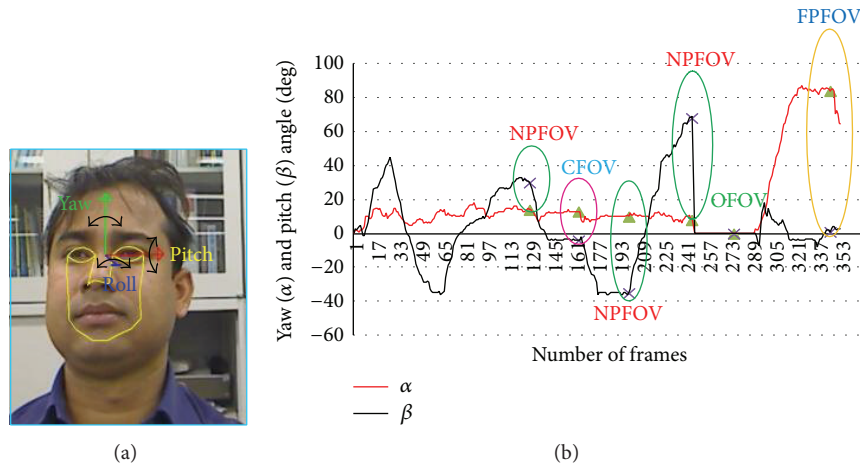

(a)

(b)

FIGURE 9: (a) Result of the HDTM. (b) Recognition of situations.

5.2.3. Situation Recognition Module (SRM). If the partner is looking at a particular direction, we guess that his/her attention is to some object located in that direction. Therefore, to recognize the existing situation (where the human is currently looking), we observe the head as well as body direction estimated by HDTM and BTM, respectively. By extrapolating from the person's head/body information, the SRM determines which situation (NPFOV, FPFOV, or OFOV) exists between the robot and the human. By examining these situations, we found that the human head/body orientations vary from one situation to another. The HDTM tracks within ±90° (right/left) only; therefore, while the human attends to OFOV situation, the system loses his/her head information; in that case, the robot recognizes the current situation based on the body information (laser sensor can track up to 270°). From the results of tracking modules, the system recognizes the four viewing situations of the target participant in terms of yaw (α), pitch (β) movements of head, and/or body direction (θ), respectively, using a set of predefined rules. In each rule, we set the values for yaw, pitch, and body directions by observing several experimental trials.

(i) *Central field of view (CFOV):* recognizes if the current head direction is within $-10° \leq \alpha \leq +10°$ and $-10° \leq \beta \leq +10°$ and remains 30 frames in the same direction.

(ii) *Near peripheral field of view (NPFOV):* recognizes if the current head direction is within $-10° > \alpha \geq +70°$ or $+10° \leq \alpha \leq 70°$ and $-10° \leq \beta \leq +10°$ and remains 30 frames in the same direction.

(iii) *Far peripheral field of view (FPFOV):* recognizes if current head direction is within $-0° > \alpha \geq +90°$ or $+70° \leq \alpha \leq +90°$ and $-10° \leq \beta \leq +10°$ and remains 30 frames in the same direction.

(iv) *Out of field of view (OFOV):* recognizes if the human is looking in the opposite direction with respect to robot's direction. That means, the robot cannot capture the human face/head and current head direction within $\alpha = \beta = 0$ or body direction within $90° < \theta \leq +270°$ or $-90° > \theta \geq -270°$ and remains 30 frames in the same direction.

Figure 10: Results of FDM.

Figure 9(b) represents the results of SRM to recognize four situations (i.e., CFOV, NPFOV, FPFOV, and OFOV, resp.) in terms of head information.

5.2.4. Eye-Contact Module (ECM). The ECM mainly consists of two submodules: FDM (dace detection module) and EBM (eye blinking module). The robot continuously checks the target person, whether his/her face is directed to the robot or not. In any situation, the robot considers that the human has responded against the robots' actions if he/she looks at the robot within expected times. That means we can consider that target person and robot are in face-to-face orientation. In that case, the FDM uses the image of the forehead camera to detect his/her frontal face (Figure 10). We use the face detector, which consists of cascaded classifiers based on AdaBoost and Haar-like features [84]. A sequence of images is captured in speed of 15 frames per second, so the face detector can be applied to whole image for detecting face region. After face detection, the FDM sends the results to the EBM. The EBM produces eye blinks to let the person know that it is aware of his/her gaze. Since the eyes are CG images, the robot can easily blink the eyes in response to the human's gazing at it. The results of EBM have been described in Section 4.2.2.

5.2.5. Pan-Tilt Unit Control Module (PTUCM). In our proactive approach, the robot needs to perform several actions (such as head turning, head shaking, and uttering reference terms) to capture the human attention. All actions are performed by the pan-tilt unit with proper control signal coming from the several modules. Figure 11 shows the PTU-D46 devices used for current design. Several properties of the robotic head are summarized in Table 1.

6. Effects of Capturing Attention Behaviours

The purpose of this experiment is to evaluate the effectiveness of our proposed robotic framework for capturing attention of the target participant while he/she was oriented toward a different viewing direction. In these situations, we assumed that the target participant is being engaged in such a work which did not absorb much attention to perform.

6.1. Participants. A total of 48 subjects (39 males and 9 females) participated in the experiment. The average age of participants was 27.9 years (SD = 4.91). They were all

Figure 11: Pan-tilt unit (PTU-D46).

Table 1: The properties of the robot head.

Items	Characteristics
Head turn (horizontal)	From −159° (left) to +159° (right)
Head turn (vertical)	From −47° (down) to +31° (top)
Eye turn	From −90° (left) to +90° (right)
Eye blinks rate	1/seconds
Rotational speed	300°/second
Head tracking (error < 3°)	From −90° (left) to +90° (right)
Body tracking (error < 6°)	Up to 270°
Tracking distance	Up to 3 meters
Number of people tracking	02

graduate students at Saitama University, Japan. They were randomly assigned to one of the four conditions. There were 10 males and 2 females in the CFOV condition, 9 males and 3 females in the NPFOV condition, 11 males and 1 female in the FPFOV condition, and 8 males and 4 females in the OFOV condition. Each participant experienced all three actions (i.e., head turning, head shaking, and reference terms) of the robot one after another in four sessions. Each session lasted approximately 120 seconds. We deliberately concealed the primary purpose of our experiment. There was no remuneration for participants.

6.2. Experimental Design. As a low attention-absorption task we considered a scenario: "watching paintings." To prompt participants to look in various directions, we hung seven paintings (P1–P7) on the wall at the same height (just above the eye level of the participants). These paintings were placed in such a way that, when observed from a participant's sitting position, they covered their whole field of view (close to 180°). To produce the stimuli, we prepared two robotic heads with the same appearance. The mere existence of such robots in an environment may prompt participants to be attracted to them because of their human-face-like appearance, even if they do not perform any actions [85]. One was a static robot (SR), which was stationary at all times. The other was a moving robot (MR). Initially MR is static and is looking in a direction not toward the human face. Two robots were placed in the participant's left and right monocular fields of view. Participants' head direction would change while watching these paintings. The roles of the left and right robotic heads
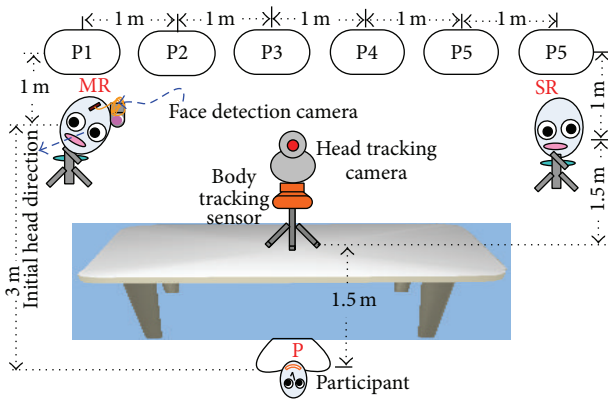
FIGURE 12: Schematic settings of the experiment.

(b) Actions.

(i) Method 1 (M1): the robot always applies head turning action whatever the situation is to attract the participant's attention.

(ii) Method 2 (M2): in order to capture the participant's attention, the robot always applies head shaking action.

(iii) Method 3 (M3): this is our proposed robot. The robot usually turns its head first toward the target human and commences shaking its head and then utters reference terms (if necessary) to capture his/her attention. The design of each behavior and working principle of this robot has been described in Section 4 in detail.

6.5. Hypotheses. The success of the robot in capturing the human attention depends on the existing viewing situation as well as the action played by the robot. Although turning the head of the robot toward the participant to whom it would like to communicate is the most fundamental action, it might be difficult to capture someone's attention by this action alone when the robot and the participant are not facing each other. Thus, a weak action may be enough in some situations, but other situations demand stronger action. We expected that the following hypotheses would be verified by the experiment.

(H1) For the CFOV and NPFOV situations, HT action toward the participant is enough to attract his/her attention.

(H2) Simple HT action is not effective at all times to capture the participant's attention when the robot exists in his/her FPFOV situation. Stronger action (s) (i.e., head shaking) are needed in the FPFOV.

(H3) For OFOV situation, any kind of head motion cannot capture the participant's attention. The robot needs to use voice/sound actions to capture his/her attention.

(H4) The proposed method (M3) will outperform the other two methods for overall evaluation.

We have assumed that the person is looking around the paintings without any particular attention target. Although existence of robots may attract his/her attention to some extent, some robot movements and head motion in the current implementation may help attract his/her attention more. Although this might be apparent, this experiment may confirm this hypothesis.

(H5) Participants will look at the moving robot (MR) significantly more than at the static robot (SR).

6.6. Measures. By observing the experimental videos, we measure the following item.

(i) Success ratio: refers to the ratio between the number of cases where participants looked at the robot in response to its action (N_L) and the total number of cases (N_A).

(SR or MR) were exchanged randomly so that the number of participants experiencing each case could be almost the same. A USB camera and a laser sensor were positioned in front of the participant to track his/her head as well as body. Two video cameras were placed in appropriate positions to capture all interactions. Figure 12 shows the experiment set up.

6.3. Experimental Procedure. Our intention was to let the participants evaluate the various behaviors of the robot as it attempted to acquire their attention when they were not initially looking in its direction. For this purpose, a single participant was asked to sit down on chair and asked to look around at the paintings. Since the positions of the robots were fixed, the participant detected them in his/her various fields of view as he/she moved his/her head and body around. We let the participant watch the paintings. The robot tracks the participant and, hence, the MR did not perform any action during the first 60 s of the interaction.

During observation of paintings, MR shows all actions (during last 60 s) of the participant one after another in each viewing condition to capture his/her attention. However, MR waits about 4 s for human response after giving each signal. If the participant looks at MR within 4 s, the robot considers that the human has been attracted. We videotaped all sessions to analyze human behaviors. Figure 13 shows some experimental scenes while interacting with the robots.

6.4. Experimental Conditions. The robot tried to attract the target participant's attention while he/she was looking at different paintings so that it could obtain data for two types of viewing situations. In order to capture the participant's attention the robot shows all actions in each viewing situation. Thus, we adopted four viewing situations and three action conditions. They were defined as follows.

(a) Viewing Situations.

By our observation, we see that the robot recognized the situation as CFOV when the participant was looking at the painting P1, as NPFOV when the participant was looking at the picture P2/P3, as FPFOV when he/she was looking at the picture P4/P5. However, it recognized the situation as OFOV while looking at the picture P6.
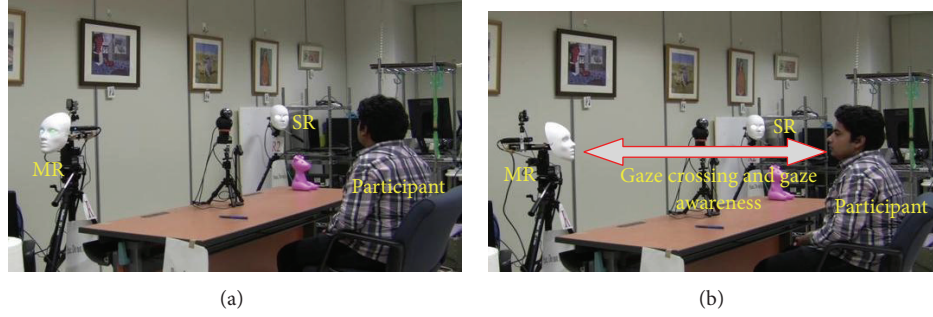
(a)

(b)

FIGURE 13: (a) The participant is watching paintings. (b) The participant has attracted and looked at MR.

TABLE 2: Summary of participants' responses against each action in different viewing conditions. S. L. means the significant level and *** means statistically very significant differences.

Viewing situations	Actions			S. L.		
	M1	M2	M3	M1 versus M2	M2 versus M1	M3 versus M1
	Mean (SD)	Mean (SD)	Mean (SD)	P	P	P
CFOV	0.92 (0.08)	1.0 (0.0)	1.0 (0.0)	0.51	0.0	0.51
NPFOV	0.83 (0.11)	0.92 (0.08)	0.92 (0.08)	0.58	0.0	0.5
FPFOV	0.084 (0.08)	0.83 (0.11)	0.92 (0.08)	0.0001^{***}	0.58	0.0001^{***}
OFOV	0.084 (0.08)	0.16 (0.11)	0.83 (0.11)	0.58	0.0002^{***}	0.0001^{***}

In order to verify the hypothesis H5, we would like use the following measure.

(ii) Number of looks: total number of times that the participants look at the robot.

6.7. Results. The experiment conducted was a 3×4 mixed-model design. For the within-participant factor (*action*), all participants interacted with three actions of the robot (M1, M2, and M3) and for the between-participant factor (*viewing situation*) one group of participants experienced the three actions in one of the four viewing situations (CFOV, NPFOV, FPFOV, and OFOV). We observed a total of 144 (12 (*participants*) $\times$ 3 (*actions*) $\times$ 4 (*situations*)) interactions for all conditions.

Table 2 summarizes mean and standard deviation (SD) of participant's response with respect to the robot's behaviours in each viewing situation. A two-way repeated-measure of ANOVA was conducted for the success ratios. A significant main effect was revealed in the action factor ($F(2, 132) = 25.08$), $P < 0.001$, $\eta^2 = 0.16$) and viewing situation factor ($F(3, 132) = 30.23$, $P < 0.001$, $\eta^2 = 0.28$). The interaction effect between the movement and viewing situation was significant ($F(6, 132) = 8.31$, $P < 0.01$, $\eta^2 = 0.15$). Figure 14 also illustrates these results.

The significant interaction effect between viewing situation and action suggests that the success ratios for different methods are affected by the viewing situation factor. Post hoc
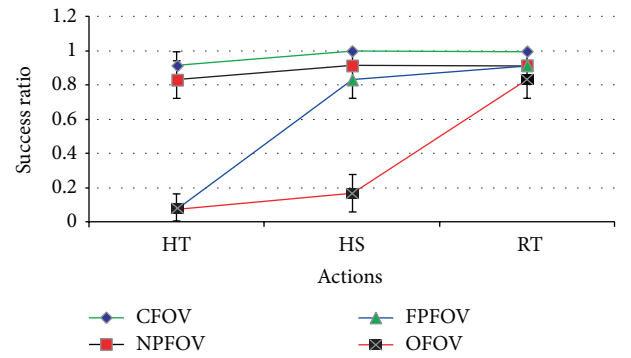

FIGURE 14: Mean values of success ratio of the robot in different actions.

tests for the viewing condition revealed significant differences between pairs (CFOV versus FPFOV: $P < 0.01$, CFOV versus OFOV: $P < 0.01$, NPFOV versus FPFOV: $P < 0.01$, NPFOV versus OFOV: $P < 0.01$) but there was no significant difference between CFOV and NPFOV for M1 action. That means M1 is effective for CFOV and NFPFOV situations. Moreover, multiple comparisons with the Bonferroni method were conducted among three action parameters for each viewing situation condition. For CFOV and NPFOV conditions, no significant differences were found between any action pairs (i.e., M1 and M2, M2 and M3, and M3 and M1). In these conditions, all actions are equally effective to

capture participant's attention toward the robot. Therefore, HT action is sufficient to capture the human attention while he/she perceived the robot in his/her CFOV or NPFOV which verifies hypothesis 1.

Concerning M2, post hoc tests for the viewing condition revealed significant differences between pairs (CFOV versus OFOV: $P < 0.01$, NPFOV versus OFOV: $P < 0.01$, FPFOV versus OFOV: $P < 0.01$) but no significant differences were found for the other pairs (CFOV versus NPFOV, CFOV versus FPFOV, NPFOV versus FPFOV). That means M2 is effective for CFOV, NPFOV, and FFOV situations but is not effective in OFOV situation. Multiple comparisons with the Bonferroni method for FPFOV condition show significant differences between the actions pairs (M1 versus M2: $P < 0.0001$, and M3 versus M1: $P < 0.0001$). No significant difference was found between M2 and M3. This means HS action of the robot achieved a higher success ratio than HT in FPFOV condition. Thus, the robot should use more strong actions in FPFOV viewing condition to gain the participants' attention, which verifies our hypothesis 2.

Concerning M3, post hoc tests for the viewing condition revealed no significant differences between all pairs which means that M3 is effective for all situations in capturing participant's attention. For OFOV condition, significant differences were found between the actions pairs (M2 versus M3: $P = 0.0002$, and M3 versus M1: $P < 0.0001$). No significant difference was between M1 and M2. This means that the RT action of the robot achieved the higher success ratio than HT and HS in the OFOV viewing condition. Thus, it cannot be possible to capture the human attention by any kind of physical action when the robot exists in such a position from where he/she cannot see the robot. In that case, using voice or sound action should be used to capture people's attention. Thus, the hypothesis 3 has verified.

We conducted multiple comparisons with the Bonferroni method that showed significant differences between methods 2 and 3 ($P = 0.01$) and between methods 3 and 1 ($P < 0.001$). Results also revealed that a substantial 92% (44 were selected out of 48 participants) of target participants' attention was captured by the proposed method, while only 48% (23 were selected out of 48 participants) and 73% (35 were selected out of 48 participants) of their attention was captured by methods 1 and 2, respectively. Figure 15 also shows the results. These results mean that the capturing attention performance of the proposed robot is clearly more effective compared to the other two methods, in terms of producing a higher success ratio, when it employs the head turning (for CFOV and NPFOV), head shaking (for FPFOV), and reference terms (for OFOV) actions. Therefore, hypothesis 4 has been verified by the experiment.

As mentioned earlier, each participant experienced four trails corresponding to each viewing situation and each session lasted approximately 120 seconds. We divided the session into two parts: (i) static period (S-period): first 60 seconds while SR and MR were static and (ii) moving period (M-period): last 60 seconds in which MR displays attention capturing actions and static again (i.e., waiting 4 seconds) in this period while SR remained static. We use a total of 48 videotapes [12 (participants) × 4 (situations)] for proposed
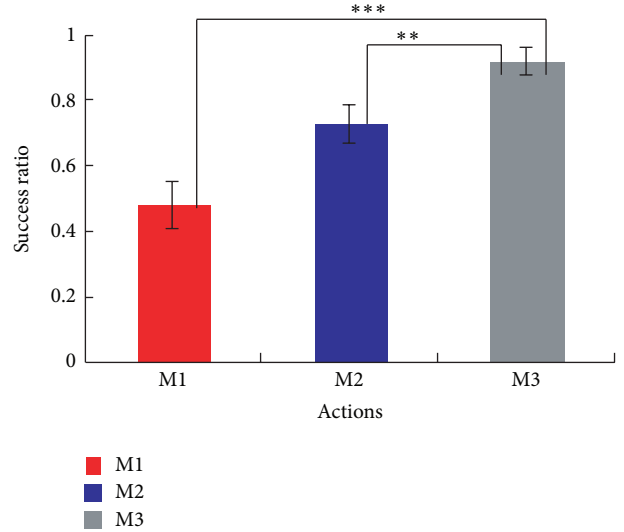

Figure 15: Overall success ratio of capturing attention in three methods. Error bars indicate the standard deviation and *** means statistically significant differences.

Table 3: Summary of participants' looking responses in each robot during static and moving periods, where M denotes means and SD denotes standard deviation.

| Interaction period | Number of looks | | F-measure |
| | MR | SR | |
	M (SD)	M (SD)	
S-period	1.58 (0.51)	1.5 (0.52)	$F(1, 23) = 0.13$
M-period	3.67 (0.49)	1.42 (0.51)	$F(1, 23) = 106.9$

moving robot (i.e., M3) to analyze human behaviors. Table 3 shows the average numbers of participants' gazing behaviors toward each robot in the static and moving periods. Concerning S-period, result shows that there is no significant difference was found in participant's looking response to each robot due to their similar face-like appearances (ANOVA: $P = 0.72$, $\eta^2 = 0.006$). However, the average number of looking toward MR is significantly greater than that toward SR (ANOVA: $P < 0.001$, $\eta^2 = 0.84$) during M-period due to the MR's actions. Although sometimes participants ignore the MR's actions, most of the time moving the head is effective in attracting human attention than the static head. Thus, hypothesis 5 has verified this experiment.

In conclusion, the results indicate that the robot needs stronger actions to attract human attention when the situation becomes tougher. Therefore, the robot should perform appropriate behavior to attract only the target participant's attention toward it in order to establish eye contact. Although, turning robot's head is usually enough for the central field of view and near peripheral field of view conditions, the robot may often need to shake its head in far peripheral field of view cases. Moreover, it is not possible to attract someone's attention using nonverbal actions (except touch) when the robot is present in the participant's OFOV because in that context the nonverbal actions can no longer be perceived

by the human's eyes. In that case, the robot should employ the voice or sound action (such as reference terms in our current design). Thus, to capture the target human's attention, we propose the following: *the robot turns its head toward the target human and then waits for a while for his/her response. If he/she does not respond within expected waiting duration, the robot tries with a stronger action and then utters reference term as a last resort.*

7. Effects of Ensuring Attention Capture Behaviours

For making a successful eye contact event, it is not only important to perform a gaze crossing function (i.e., capturing attention of target human) but also it is important to perform gaze-awareness function. Thus, the purpose of this experiment is to verify the effects of eye blinking action of the robot for ensuring attention capture of the target participant (i.e., gaze-awareness).

7.1. Participants. A total of thirty six graduate students (27 males and 9 females) from Saitama University participated in the experiment. Their ages ranged from 22 to 35 with an average of 26.8 (SD = 3.72). Eight male and four female participants were engineering majors such as electrical engineering and civil engineering, while ten males and two females came from a science background including biological, computer, and information sciences and 9 males and 3 females came from nonengineering fields (i.e., accounting, social science, and English). There was no remuneration paid to the participants. They did not participate in the previous experiment.

7.2. Design and Procedure. The experiment environment and settings were almost the same as the previous ones mentioned, except the behaviors of the proposed robot (M3). Participants are asked to watching the hanging paintings. The robot recognizes his/her viewing situation and applies proposed action corresponding to that situation. It adjusts its head orientation (if needed) to cross the gaze of the attracted participant. After gaze crossing, the robot displays its eye blinking actions to the attracted participant as a gaze-awareness function.

7.3. Experimental Conditions. To verify the effect of eye blinks, we prepared the following two conditions.

(i) *Eye contact robot with no blinks (ECR + NB):* the robot uses attention capture behaviors as proposed in Section 4. If the target participant is looked at, the robot recognizes his/her face, waits about three seconds, and then turns its head toward another direction.

(ii) *Eye contact robot with blinks (ECR + B):* this is our proposed robot. The robot recognizes his/her face while he/she is looking at it. After detecting the face of the target participant, the robot starts blinking its eyes about three times (1 blink per second) and then turns its head toward another direction. A detailed description of the behavior of this robot is given in Section 4.

7.4. Measures. A measurement of this experiment contained the following two items.

7.4.1. Impression of Robots. We asked participants to fill out a questionnaire after interactions with the robots were complete. The measurement was a simple rating on a Likert scale of 1 to 7, where 1 stands for the lowest and 7 for the highest. The questionnaire had the following items.

(i) *Attention attraction:* did you feel that behaviors of the robot captured your attention?

(ii) *The feeling of making eye contact:* did behaviors of the robot create your feeling of making eye contact?

(iii) *Overall evaluation:* how effective is the robot for making eye contact?

7.4.2. Gaze Time. We measure the total time spent by gazing at the robot in each method by observing the experimental videos. This time is measured from the beginning of gaze crossing action of the robot to the end of the participant's looking at it before turning head to another direction.

7.5. Results. The experiment had a within-participant design, and the order of all experimental trials was counterbalanced. Every participant experienced all two conditions. We conducted the repeated measures ANOVA for all measures.

Figure 16 shows the participants response to each question. Concerning attracting participants' attention, ANOVA analysis shows that there is no significant difference between the robot with blinks and without blinks conditions ($F(1, 35) = 0.32$, $P = 0.57$, $\eta^2 = 0.56$). This happens due to the same attention capturing behaviours of robot in two conditions.

In the case of feeling of making eye contact, ANOVA analysis shows a significant difference between the robot with blinks and without blinks ($F(1, 35) = 158.7$, $P < 0.01$, $\eta^2 = 0.69$). This result revealed that the participant's impressions are greatly affected by the eye blinking behaviors of the robot and this behavior produced a better feeling of making eye contact.

Concerning the overall evaluation, a significant main effect was found ($F(1, 35) = 644.7$, $P < 0.0001$, $\eta^2 = 0.9$)) using ANOVA analysis. Participant rated more the robot with blinks condition (ECR + B) (mean score =5.61) than the robot with no blinks condition (ECR + NB) (mean score = 1.72). Figure 17 also illustrates this result. Thus, the results reveal that the proposed system is more preferable than the other method (ECR+NB) to make eye contact with the participants.

We calculate the total time that the participants spent looking at the robot in each method after meeting face-to-face (Figure 17). Results indicate that the participant looks significantly longer in proposed method (2.46 s) than the other method (1.13 s). ANOVA analysis showed a significant difference in gazing time that the participants were spending
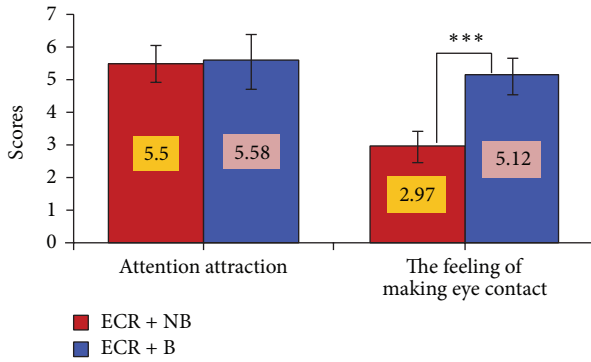
FIGURE 16: Participants' responses on questionnaires items. Error bars indicate the standard deviation and ***means significant differences.
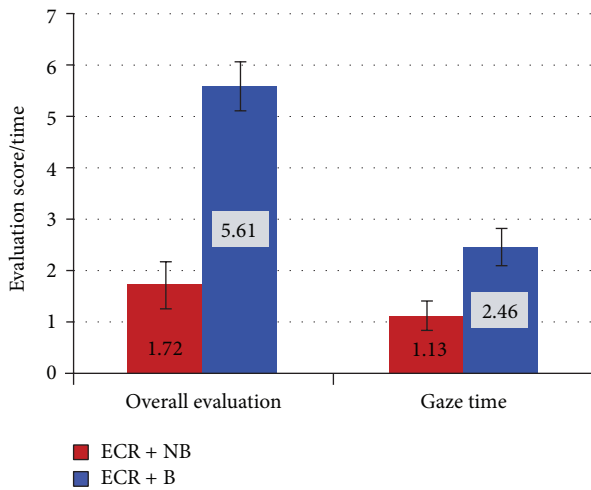

FIGURE 17: Results of overall evaluation (left graph) and total time spent on gazing at two robots (right graph).

to gaze at each robot ($F(1, 35) = 269.2$, $P < 0.0001$, $\eta^2 = 0.78$).

8. Discussion

Eye contact is a primary contributor to the level of intimacy in a conversation along with physical proximity, topic intimacy, amount of smiling, and so on [86]. If one of these dimensions of intimacy is disturbed, compensatory changes will likely occur along the other dimensions. Gaze aversion is defined as the intentional redirection of gaze away from the face of an interlocutor [87]. Person A avoids looking at person B especially if being looked at and/or moves the gaze away from B [88, 89]. In general, people frequently avert their gaze to alleviate feelings of self-consciousness and, while listening, to make speakers more comfortable and to reduce negative perceptions associated with staring [90].

It is certain that both functions, gaze contact and gaze aversion, are important for prolonging conversation. However, gaze contact is important for *initiating a conversation*

and gaze aversion is important for *sustaining a conversation*. Mutlu et al. [9] shows that turning head or gaze should be used as a fundamental nonverbal action/signal when a person would like to capture other person's attention. That means person A should turn his/her face/head/gaze toward person B to capture B's attention. In our work, we focused on one of the gaze behaviors (i.e. gaze contact) that are mandatory for *initiating any conversation* [7].

8.1. Selective Friendly Proactive Eye Contact. In proactive approach, the robot should capture the target human attention first for establishing eye contact. Our purpose is to develop a robot that can make eye contact with a particular human while avoiding attracting other people's attention as much as possible. Thus, the robot should consider the current situation of intended people with whom it would like to start communicating and try to apply an appropriate action to that situation. For this purpose, we propose an eye contact process consisting of capturing attention and ensuring attention capture. In order to initialize an eye contact episode, the robot should start with a weaker action to avoid attracting other people's attention as much as possible except the target person and use stronger actions only when weak action fails to attract his/her attention. This is the basic design concept of our robot.

From the survey of psychology and HRI literatures, we chose turning the head (to look at the person) as the weakest action. We determined to use head shaking if the robot cannot attract the target person's attention and use reference terms if the robot captured the target person in its out of field of view condition. We have confirmed through experiments that our design concept can be useful to realize such robots that can capture a particular person as selectively as possible.

8.2. May Use Blinks as a Gaze-Awareness Modality. Blinking actions strengthen the feeling of being looked at and it can be used to convey an impression more effectively by colorfully understanding human social behavior. Experimental results have also confirmed that eye blinking actions proved helpful in relaying to the target that the robot was aware of his/her gaze. A participant's eyes coupled with the robot's eyes during blinking, and this is why the human participant spent more time gazing at the robot. This behavior may help the human identify quickly attention shift focus of robots. However, without blinking the robot may fail to create the feeling of eye contact being established, due to its lack of a gaze-awareness function. Unfortunately, we did not find any of the previous studies that focused on the eye blinks or iris expression as gaze-awareness function. Thus, the effectiveness of the system compared to the others is unknown.

8.3. Future Challenges. There are still several issues that have not been addressed in the current model. Some of these issues are discussed in the following.

(i) *Generalizability:* we tested our model for a specific scenario where a single participant was engaged in a task that demanded a relatively low level of attention. Therefore, its generalizability is limited. The strategies

required attracting and making eye contact with a target human alone when he/she is situated in a group or otherwise intensely involved in a task may be different. More studies are needed to explore the dynamics of crowds where people are engaged in more attention-demanding tasks.

(ii) *Limited cues:* we limited the robot's behavior to only two physical movements such as head and eye. However, robots may need to use other bodily actions depending on the situation. Robots may also need to use voice in some situations, although this may attract the attention of others as well as the target person. Therefore, we need to explore other possible situations and design appropriate actions for those situations. For example, touch action may be used effectively in place of voice in out of field of view condition.

(iii) *Factors affecting attention:* the success of an action of the robot may depend on several factors including distance between the human and the robot, direction of current attentional focus, duration of the action of the robot, speed of the action of the robot, level of attention in a task that the human is currently involved in, mental condition of the target human, and so on. In this work, we only consider the direction of current attentional focus of the target human. Thus, the future work should include the remaining factors as well develop a better eye contact process.

(iv) *Embodiment:* we verified our design by using a robotic head only which consists of a head mask. Full body embodiment and human-like appearances of robot should affect humans interacting patterns with robots. Therefore, future works consider the usability and performance of current approach in the humanoid robotic platforms.

(v) *Cultural or personal differences:* humans' impression about robot's behaviors may depend on the cultural and personal factors. Robots need to adapt their ways of attention control to their partner's characteristics. Future work should look at how designed behaviors could be extended to robots that work in different cultural contexts, use different languages, and interact with people with different demographic and personality attributes.

(vi) *Technical challenges:* due to the state of vision processing systems, today's social robots offer very limited interactivity in generating behavior and constructing interaction. The system presented in this work recognizes, tracks, and understands responses of a single participant. Therefore, the robot should interpret more people responses and respond appropriately to them to cope with the real world situation. Moreover, building real-time interactivity into social robots will require combining speech and nonverbal behavior recognition and generation and cognitive representations of the world that adapt to new input from users and the environment.

(vii) *Reactive approach:* the proposed model works in a proactive manner. However, to develop a better eye contact model this model should be combined with the reactive approach also. In the reactive approach, the robot responds appropriately and makes eye contact with the human when he/she is looking at the robot for initiating the interaction.

9. Conclusion

The primary focus of our work is to develop a robot that can make eye contact with a particular person by nonverbal means. For this purpose, we have proposed a proactive approach of eye contact that consists of two phases including capturing attention and ensuring attention capture. Throughout this work, we have argued that these steps can be used successfully in a robotic platform that provides social and cognitive benefits in the HRI research community. Although there may be various nonverbal behaviors, we incorporated head turning, head shaking, and eye blinking in respective phases. We have shown that our method can function to establish eye contact event with the target human in a situation where he/she is not initially looking toward the robot (in particular, we have considered three such situations, namely, CFOV, NPFOV, FPFOV, and OFOV) and is involved in a task that does not demand much attention. Making eye contact proactively is an important social phenomena and a prerequisite in several social functions such as engagement, initiating conversation, shared attention, and so on. The robot may use proactive approach for making eye contact in several contexts (i.e., information providing services, providing route direction, salesperson, tutoring services, and so on) that demand such kinds of social functions. Our future work will connect real-world applications with the proposed model of proactive eye contact.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

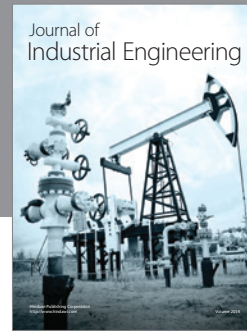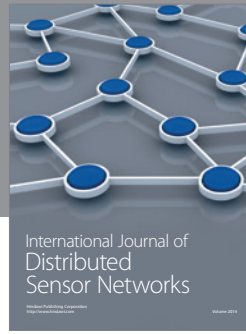## References

[1] M. Argyle, *Bodily Communication*, Routledge, London, UK, 1988.

[2] T. Striano, V. M. Reid, and S. Hoehl, "Neural mechanisms of joint attention in infancy," *European Journal of Neuroscience*, vol. 23, no. 10, pp. 2819–2823, 2006.

[3] T. Farroni, E. M. Mansfield, C. Lai, and M. H. Johnson, "Infants perceiving and acting on the eyes: tests of an evolutionary hypothesis," *Journal of Experimental Child Psychology*, vol. 85, no. 3, pp. 199–212, 2003.

[4] C. Trevarthen and K. J. Aitken, "Infant intersubjectivity: research, theory, and clinical applications," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 42, no. 1, pp. 3–48, 2001.

[5] C. L. Kleinke, "Gaze and eye contact: a research review," *Psychological Bulletin*, vol. 100, no. 1, pp. 78–100, 1986.

[6] C. Fullwood and G. Doherty-Sneddon, "Effect of gazing at the camera during a video link on recall," *Applied Ergonomics*, vol. 37, no. 2, pp. 167–175, 2006.

[7] E. Goffman, *Behaviour in Public Places: Notes on the Social Organization of Gatherings*, The Free Press, New York, NY, USA, 1st edition, 1963.

[8] A. M. Sabelli, T. Kanda, and N. Hagita, "A conversational robot in an elderly care center: an ethnographic study," in *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI '11)*, pp. 37–44, Lausanne, Switzerland, March 2011.

[9] B. Mutlu, "The design of gaze behavior for embodied social interfaces," in *Proceedings of the 28th Annual CHI Conference on Human Factors in Computing Systems*, pp. 2661–2664, Florence, Italy, April 2008.

[10] S. E. Brennan and E. A. Hulteen, "Interaction and feedback in a spoken language system: a theoretical framework," *Knowledge-Based Systems*, vol. 8, no. 2-3, pp. 143–151, 1995.

[11] C. Peters, "Direction of attention perception for conversation initiation in virtual environments," in *Intelligent Virtual Agents*, vol. 3661 of *Lecture Notes in Computer Science*, pp. 215–228, Springer, London, UK, 2005.

[12] S. R. H. Langten, R. J. Watt, and V. Bruce, "Do the eyes have it? Cues to the direction of social attention," *Trends in Cognitive Sciences*, vol. 4, no. 2, pp. 50–59, 2000.

[13] J. E. Hanna and S. E. Brennan, "Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation," *Journal of Memory and Language*, vol. 57, no. 4, pp. 596–615, 2007.

[14] M. Cranach, "The role of orienting behavior in human interaction, behavior and environment," in *The Use of Space by Animals and Men*, H. Esser, Ed., pp. 217–237, Plenum Press, New York, NY, USA, 1971.

[15] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "The effects of responsive eye movement and blinking behavior in a communication robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 4564–4569, Beijing, China, October 2006.

[16] M. Argyle and M. Cook, *Gaze and Mutual Gaze*, Cambridge University Press, Oxford, UK, 1976.

[17] M. Pilu, "On the use of attention clues for an autonomous wearable camera," HP Technical Report HPL-2002-195R1, 2003.

[18] J. M. Loomis, J. W. Kelly, M. Pusch, J. N. Bailenson, and A. C. Beall, "Psychophysics of perceiving eye-gaze and head direction with peripheral vision: implications for the dynamics of eye-gaze behavior," *Perception*, vol. 37, no. 9, pp. 1443–1457, 2008.

[19] N. J. Emery, "The eyes have it: the neuroethology, function and evolution of social gaze," *Neuroscience and Biobehavioral Reviews*, vol. 24, no. 6, pp. 581–604, 2000.

[20] I. Perrett and J. Emery, "Understanding the intentions of others from visual signals: neurophysiological evidence," *Journal of Current Psychology of Cognition*, vol. 13, no. 5, pp. 683–694, 1994.

[21] C. Ware, *Information Visualization: Perception for Design*, Morgan Kaufmann, San Francisco, Calif, USA, 2004.

[22] M. M. Hoque, D. Das, T. Onuki, Y. Kobayashi, and Y. Kuno, "Robotic system controlling target human's attention," in *Intelligent Computing Theories and Applications: Proceedings of the 8th International Conference, ICIC 2012, Huangshan, China, July 25–29, 2012*, vol. 7390 of *Lecture Notes in Computer Science*, pp. 534–544, Springer, Berlin, Germany, 2012.

[23] T. Hall, *The Hidden Dimension: Man's Use of Space in Public and Private*, The Bodley Head, 1966.

[24] A. Kendon, "Features of the structural analysis of human communicational behavior," in *Aspects of Nonverbal Communication*, pp. 29–43, Swets & Zeitlinger, Lisse, The Netherlands, 1980.

[25] A. Kendon, *Spatial Organization in Social Encounters: The F-Formation System*, Cambridge University Press, New York, NY, USA, 1990.

[26] C. Breazeal, *Designing Social Robots*, The MIT Press, Cambridge, Mass, USA, 2002.

[27] H. Kozima and H. Yano, "In search of otogenetic prerequisites for embodied social intelligent," in *Proceedings of the International Conference on Cognitive Science: Workshop on Emergence and Development of Embodied Cognition*, pp. 30–34, Beijing, China, August 2001.

[28] B. Scassellati, *Foundations for a theory of mind for a humanoid robot [Ph.D. thesis]*, Department of Electrical Engineering and Computer Science, MIT, Cambridge, Mass, USA, 2001.

[29] H. Kozima and H. Yano, "A robot that leans to communicate with human caregivers," in *Proceedings of the 1st International Workshop on Epigenetic Robotics*, Lund, Sweden, September 2001.

[30] M. Fujita and H. Kitano, "Development of an autonomous quadruped robot for robot entertainment," *Autonomous Robots*, vol. 5, no. 1, pp. 7–18, 1998.

[31] T. Tashima, S. Saito, T. Kudo, M. Osumi, and T. Shibata, "Interactive pet robot with an emotion model," *Advanced Robotics*, vol. 13, no. 3, pp. 225–226, 1998.

[32] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: challenges and results," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 271–281, 2003.

[33] K. Severinson-Eklundh, A. Green, and H. Hüttenrauch, "Social and collaborative aspects of interaction with a service robot," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 223–234, 2003.

[34] E. A. Sisbot, R. Alami, T. Simeon et al., "Navigation in the presence of humans," in *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots*, pp. 181–188, Tsukuba, Japan, December 2005.

[35] E. A. Sisbot, A. Clodic, L. F. Marin U, M. Fontmarty, L. Brèthes, and R. Alami, "Implementing a human-aware robot system," in *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '06)*, pp. 727–732, Hatfield, UK, September 2006.

[36] E. Pacchierotti, H. I. Christensen, and P. Jensfelt, "Design of an office-guide robot for social interaction studies," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 4965–4970, Beijing, China, October 2006.

[37] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, "Learning motion patterns of people for compliant robot motion," *International Journal of Robotics Research*, vol. 24, no. 1, pp. 31–48, 2005.

[38] R. Gockley, A. Bruce, J. Forlizzi et al., "Designing robots for long-term social interaction," in *Proceedings of the IEEE IRS/RSJ International Conference on Intelligent Robots and Systems (IROS '05)*, pp. 1338–1343, Beijing, China, August 2005.

[39] K. Hayashi, D. Sakamoto, T. Kanda et al., "Humanoid robots as a passive-social medium: a field experiment at a train station," in *Proceedings of the ACM/IEEE Conference on Human-Robot*

Interaction—Robot as Team Member (HRI '07), pp. 137–144, Arlington, Va, USA, March 2007.

[40] S. Li, B. Wrede, and G. Sagerer, "A dialog system for comparative user studies on robot verbal behavior," in *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '06)*, pp. 129–134, Hatfield, UK, September 2006.

[41] M. P. Michalowski, S. Sabanovic, and R. Simmons, "A spatial model of engagement for a social robot," in *Proceedings of the 9th IEEE International Workshop on Advanced Motion Control*, pp. 762–767, Istanbul, Turkey, March 2006.

[42] M. Yamamoto and T. Watanabe, "Timing control effects of utterance to communicative actions on embodied interaction with a robot," in *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN '04)*, pp. 467–472, Okayama, Japan, September 2004.

[43] N. Bergström, T. Kanda, T. Miyashita, H. Ishiguro, and N. Hagita, "Modeling of natural human-robot encounters," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '08)*, pp. 2623–2629, Nice, France, September 2008.

[44] B. Schauerte, J. Richarz, and G. A. Fink, "Saliency-based identification and recognition of pointed-at objects," in *Proceedings of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '10)*, pp. 4638–4643, Taipei, Taiwan, October 2010.

[45] K. Yamazaki, M. Kawashima, Y. Kuno et al., "Prior-to-request and request behaviors within elderly day care: implications for developing service robots for use in multiparty settings," in *Proceedings of the 10th European Conference on Computer Supported Cooperative Work (ECSCW '07)*, pp. 61–78, Limerick, Ireland, September 2007.

[46] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki, "Reconfiguring spatial formation arrangement by robot body orientation," in *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI '10)*, pp. 285–292, Osaka, Japan, March 2010.

[47] K. Dautenhahn, M. Walters, S. Woods et al., "How may I serve you? A robot companion approaching a seated person in a helping context," in *Proceedings of the ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI '06)*, pp. 172–179, Salt Lake City, Utah, USA, March 2006.

[48] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "A model of proximity control for information-presenting robots," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 187–195, 2010.

[49] R. Gockley, J. Forlizzi, and R. Simmons, "Natural person-following behavior for social robots," in *Proceedings of the ACM/IEEE Conference on Human-Robot Interaction—Robot as Team Member (HRI '07)*, pp. 17–24, Arlington, Va, USA, March 2007.

[50] M. Buss, D. Carton, B. Gonsior et al., "Towards proactive human-robot interaction in human environments," in *Proceedings of the 2nd International Conference on Cognitive Infocommunications (CogInfoCom '11)*, Budapest, Hungary, July 2011.

[51] A. Cesta, G. Cortellessa, M. Vittoria Giuliani, F. Pecora, M. Scopelliti, and L. Tiberio, "Psychological implications of domestic assistive technology for the elderly," *PsychNology Journal*, vol. 5, no. 3, pp. 229–252, 2007.

[52] M. M. Hoque, T. Onuki, E. Tsuburaya et al., "An empirical framework to control human attention by robot," in *Computer Visin—ACCV 2010 Workshops, Part I*, vol. 6468 of *Lecture Notes in Computer Science*, pp. 430–439, Springer, London, UK, 2011.

[53] C. Shi, M. Shimada, T. Kanda, H. Ishiguro, and N. Hagita, "Spatial formation model for initiating conversation," in *Proceedings of the International Conference on Robotics: Science and Systems*, Los Angeles, Calif, USA, June 2011.

[54] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans? Strategies for social robots to initiate interaction," in *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI'09)*, pp. 109–116, La Jolla, Calif, USA, March 2009.

[55] N. Mitsunaga, T. Miyashita, H. Ishiguro, K. Kogure, and N. Hagita, "Robovie-IV: a communication robot interacting with people daily in an office," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '06)*, pp. 5066–5072, Beijing, China, October 2006.

[56] P. Althaus, H. Ishiguro, T. Kanda, T. Miyashita, and H. I. Christensen, "Navigation for human-robot interaction tasks," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1894–1900, Barcelona, Spain, April-May 2004.

[57] E. A. Sisbot, K. F. Marin-Urias, R. Alami, and T. Siméon, "A human aware mobile robot motion planner," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 874–883, 2007.

[58] A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M. Williamson, "The cog project: building a humanoid robot," in *Computation for Metaphors, Analogy, and Agents*, vol. 1562 of *Lecture Notes in Computer Science*, pp. 52–87, Springer, Berlin, Germany, 1999.

[59] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, "Development and evaluation of interactive humanoid robots," *Proceedings of the IEEE*, vol. 92, no. 11, pp. 1839–1850, 2004.

[60] Y. Matsusaka, T. Tojo, S. Kubota et al., "Multi-person conversation via multimodal interface—a robot who communicate with multi-user," in *Proceedings of the European Conference on Speech Communication and Technology*, pp. 1723–1726, Budapest, Hungary, 2012.

[61] B. Mutlu, T. Kanda, J. Forlizzi, J. Hodgins, and H. Ishiguro, "Conversational gaze mechanisms for humanlike robots," *ACM Transactions on Interactive Intelligent Systems*, vol. 1, no. 2, article 12, 2012.

[62] M. Shiomi, T. Kanda, N. Miralles et al., "Face-to-face interactive humanoid robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '04)*, pp. 1340–1346, Sendai, Japan, September-October 2004.

[63] T. Yonezawa, H. Yamazoe, A. Utsumi, and S. Abe, "Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking," in *Proceedings of the 9th International Conference on Multimodal Interfaces (ICMI '07)*, pp. 140–145, Nagoya, Japan, November 2007.

[64] B. Mutlu, T. Kanda, J. Forlizzi, J. Hodgins, and H. Ishiguro, "Conversational gaze mechanisms for humanlike robots," *ACM Transactions on Interactive Intelligent Systems*, vol. 1, no. 2, article 12, 2012.

[65] D. Das, Y. Kobayashi, and Y. Kuno, "Attracting attention and establishing a communication channel based on the level of visual focus of attention," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2194–2201, Tokyo, Japan, November 2013.

[66] D. Miyauchi, A. Nakamura, and Y. Kuno, "Bidirectional eye contact for human-robot communication," *IEICE Transactions on Information and Systems*, vol. 88, no. 11, pp. 2509–2516, 2005.

[67] D. Maurer, *Infant's Perception of Facedness*, Ablex, Norwood, NJ, USA, 1985.

[68] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "Responsive robot gaze to interaction partner," in *Proceedings of the International Conference on Robotics: Science and Systems*, pp. 457–462, Philadelphia, Pa, USA, 2006.

[69] C.-M. Huang and A. L. Thomaz, "Effects of responding to, initiating and ensuring joint attention in human-robot interaction," in *Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '11)*, pp. 65–71, Atlanta, Ga, USA, August 2011.

[70] B. Miller, "Response time in man-computer conversational transactions," in *Proceedings of the American Federation of Information Processing Societies, Fall Joint Computer Conference (AFIPS '68)*, pp. 267–277, San Francisco, Calif, USA, December 1968.

[71] B. Most and J. Simons, "Attention capture, orienting, and awareness," in *Attraction, Distraction and Action; Multiple Perspectives on Attentional Capture*, C. Folk and B. Gibson, Eds., vol. 133 of *Advances in Psychology*, pp. 151–173, Elsevier, New York, NY, USA, 2001.

[72] M. Finke, K. L. Koay, K. Dautenhahn, C. L. Nehaniv, M. L. Walters, and J. Saunders, "Hey, I'm over here—how can a robot attract people's attention?" in *Proceedings of the 14th IEEE Workshop on Robot and Human Interactive Communication (RO-MAN '05)*, pp. 7–12, August 2005.

[73] W. James, *The Principle of Psychology*, Dover, New York, NY, USA, 1950.

[74] K. Takashima, Y. Omori, Y. Yoshimoto, Y. Itoh, Y. Kitamura, and F. Kishino, "Effects of avatar's blinking animation on person impressions," in *Proceedings of the International Conference on Graphics Interface*, pp. 169–176, Toronto, Canada, May 2008.

[75] M. Cook and J. M. C. Smith, "The role of gaze in impression formation," *The British Journal of Social and Clinical Psychology*, vol. 14, no. 1, pp. 19–25, 1975.

[76] R. V. Exline, "Multichannel transmission of nonverbal behavior and the perception of powerful men: the presidential debates of 1976," in *Power, Dominance, and Nonverbal Behavior*, Springer Series in Social Psychology, pp. 183–206, Springer, New York, NY, USA, 1985.

[77] 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 International Conference on Human Factors in Computing Systems (SIGCHI '02)*, pp. 41–48, Minneapolis, Minn, USA, April 2002.

[78] L. Itti, N. Dhavale, and F. Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," in *6th Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation*, vol. 64 of *Proceedings of SPIE*, pp. 64–78, San Diego, Calif,USA, August 2003.

[79] M. J. Doughty, "Further assessment of gender and blink pattern-related differences in the spontaneous eyeblink activity in primary gaze in young adult humans," *Journal of Optometry and Vision Science*, vol. 79, no. 7, pp. 439–447, 2002.

[80] J. de Greeff, F. Delaunay, and T. Belpaeme, "Towards retro-projected robot faces: an alternative to mechatronic and android faces," in *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive (RO-MAN '09)*, pp. 306–311, Toyama, Japan, October 2009.

[81] Y. Kobayashi and Y. Kuno, "People tracking using integrated sensors for human robot interaction," in *Proceedings of the IEEE-ICIT International Conference on Industrial Technology (ICIT '10)*, pp. 1617–1622, Viña del Mar, Chile, March 2010.

[82] M. Isard and A. Blake, "Condensation-conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[83] FaceAPI, "Face Tracking for OEM Product Development," Seeing Machines Int., 2010, http://www.faceapi.com.

[84] G. Bradsky, A. Kaehler, and V. Pisarevsky, "Learning based computer vision with Intel's open computer vision library," *Intel Technology Journal*, vol. 9, no. 1, pp. 119–130, 2005.

[85] P. E. Downing, C. M. Dodds, and D. Bray, "Why does the gaze of others direct visual attention?" *Visual Cognition*, vol. 11, no. 1, pp. 71–79, 2004.

[86] M. Argyle and J. Dean, "Eye-contact, distance and affiliation," *Sociometry*, vol. 28, pp. 289–304, 1965.

[87] S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu, "Conversational gaze aversion for humanlike robots," in *Proceedings of the 9th ACM/IEEE International Conference on Human Robot Interaction*, pp. 25–32, Bielefeld, Germany, March 2014.

[88] M. Cranach and J. Ellgring, *The Perception of Looking Behaviour*, Academic Press, London, UK, 1973.

[89] N. J. Emery, "The eyes have it: the neuroethology, function and evolution of social gaze," *Neuroscience and Biobehavioral Reviews*, vol. 24, no. 6, pp. 581–604, 2000.

[90] A. Abele, "Functions of gaze in social interaction: communication and monitoring," *Journal of Nonverbal Behavior*, vol. 10, no. 2, pp. 83–101, 1986.